Fast Convergent Solvers for Augmented Lagrangian Equations in Contact Mechanics

J.H. Heegaard, C.B. Hovey, and M.L. Kaplan

Department of Mechanical Engineering
Stanford University
Stanford, CA 94305, USA

Summary

We compare the efficiency of several algorithms commonly used to solve unilateral contact problems. The penalty method is the simplest algorithm to implement, but its poor convergence properties for large values of the penalty parameter require costly continuation procedures. Augmented Lagrangian methods partly alleviate conditioning problems of the penalty method but typically use slowly converging update schemes like Uzawa's algorithm (UA). Finally, the Generalized Newton-Raphson Method (GNRM), which simultaneously updates the primal and dual variables, displays local quadratic rates of convergence. Of all the solvers discussed in this paper, we found that the GNRM applied to the augmented Lagrangian formulation was clearly the most efficient algorithm in terms of number of iterations required for convergence.

Introduction

The interaction between two or more deformable bodies is among the most challenging problems in continuum mechanics. The equations describing equilibrium of the contacting bodies must include additional conditions to account for the impenetrability and possible friction at the interface. These conditions are typically expressed as inequality constraints on the displacement fields of the contacting bodies. The problem is then to solve the balance equations in the presence of these inequality constraints. The objective of this contribution is to compare the convergence properties of three families of algorithms commonly used to solve unilateral contact problems in continuum mechanics, namely the penalty method (PM), the augmented Lagrangian formulation based on the Uzawa algorithm (UA), and the augmented Lagrangian formulations based on the Generalized Newton-Raphson Method (GNRM). Comprehensive reviews of these methods applied to contact problem can be found in [1] and [2].

In the present work we restrict ourselves to frictionless contact between deformable bodies. We further assume all processes to be quasi-static, hence discarding dynamical effects such as impact. However, we allow the contacting bodies to undergo finite displacements and therefore consider geometric nonlinearities in the problem formulation. Impenetrability of the contacting bodies is enforced by inequality constraints on the gap measuring the separation between the interacting bodies.

The augmented Lagrangian algorithm has become the method of choice to solve the resulting inequality constrained problem [3, 4, 5, 6, 7]. A standard Lagrangian is first constructed from the objective function and constraint functions (*e.g.*, energy functional and signed gap distance). It is well known however, that for inequality constraints, the use of standard Lagrange multipliers does not remove the inequality constraints which get merely transfered to the multipliers. To remove the constraints on the multipliers, an augmented Lagrangian is constructed by appending a regularization term to the standard Lagrangian [8, 9, 10].

The equations characterizing the unconstrained saddle point of the augmented Lagrangian are typically solved using an Uzawa type algorithm [11]. Starting with an initial estimate for the Lagrange multipliers, the primal problem is solved up to a given tolerance after which a first order update is performed on the Lagrange multiplier. This procedure is repeated until convergence on both primal and dual variables is achieved. While the UA does not increase the size of the underlying primal minimization problem, its convergence rate is usually slow.

An alternative approach is to linearize the nonlinear problem simultaneously along the primal and dual variables with a GNRM [12]. Although the size of the corresponding saddle point problem is increased by the additional unknown multipliers, its local convergence rate is quadratic on the fully coupled problem.

We start by providing a short description of the geometrically nonlinear contact problem. We then briefly review each of the three solution methods mentioned above. The numerical performance of each algorithm is then discussed in the light of two canonical contact problems. Although no formal proof is presented, these examples consistently indicate that the GNRM applied to the augmented Lagrangian operator requires consistently fewer iterations to convergence than the PM or UA.

Contact Problem Formulation

We consider two deformable bodies Ω^α, $\alpha = 1, 2$ whose boundaries are denoted by Γ^α. The associated deformation maps $\boldsymbol{\varphi}^\alpha$ are used to express a current configuration $\boldsymbol{x}^\alpha = \boldsymbol{x}^\alpha(\boldsymbol{X}^\alpha, t) = \varphi_t^\alpha(\boldsymbol{X}^\alpha)$ of Ω^α in terms of a reference configuration $\boldsymbol{X}^\alpha$. We further assume the existence of a locally convex energy functional $\phi = \sum_\alpha \phi^\alpha$ from which the constitutive models describing the material behavior of Ω^α can be derived (hyperelastic materials.) We define Ω^1 to be the striker and Ω^2 the target body. Impenetrability of the bodies is controlled by a gap function $g_n : \Gamma^\alpha \mapsto \mathbb{R}$ defined as $g_n = \boldsymbol{g} \cdot \hat{\boldsymbol{n}}$ where $\boldsymbol{g}$ is the usual gap vector between pairs of closest points on Γ^α (*i.e.*, points which are also mutual projections of each other), and $\hat{\boldsymbol{n}}$ is an outward normal vector to the striker at the projection point [13, 14]. Any feasible solution to the standard boundary value problem on Ω^α must then satisfy the following inequality constraint on the gap function

$$g_n \geq 0 \qquad \forall \boldsymbol{X}^\alpha \in \Gamma^\alpha \tag{1}$$

The gap distance g_n is related to the conjugate pressure t_n by a multivalued non-differentiable unilateral contact law [15] characterized by three complementary Signorini conditions (equivalent to the Kuhn-Tucker conditions in constrained optimization)

$$g_n \geq 0 \qquad t_n \leq 0 \qquad g_n t_n = 0 \tag{2}$$

respectively expressing that the contacting bodies cannot penetrate each other, cannot pull on each other and are either separated or pressing on each other. These conditions can be conveniently expressed as a concise subdifferential inclusion

$$t_n \in \partial\psi_{\mathbb{R}^-}(g_n) \tag{3}$$

where $\psi_{\mathbb{R}^-}$ is the usual indicator function of the convex cone $\mathbb{R}^-$ [15]. The boundary value problem expresses then as a constrained optimization problem

$$\begin{cases} \min\limits_{\boldsymbol{u}} & \phi := \sum\limits_\alpha \phi^\alpha - \eta \\ \text{s.t.} & g_n \geq 0 \end{cases} \tag{4}$$

where η represent the work done by the natural boundary conditions. In practice a discretized version of the previous problem is solved using the finite element method. The unknown displacement field $\boldsymbol{u}$ is discretized into a finite dimensional vector $\boldsymbol{U}^h$ and the contact conditions are discretized into an

M-dimensional vector $\boldsymbol{g}_n$. The corresponding discrete optimization problem expresses then as

$$\begin{cases} \min\limits_{\boldsymbol{U}^h} & \phi := \sum\limits_{\alpha} \phi^{\alpha} - \eta \\ \text{s.t.} & \boldsymbol{g}_n \in \mathbb{R}^{M+} \end{cases} \tag{5}$$

where $\mathbb{R}^{M+}$ denotes the M-dimensional positive orthant of $\mathbb{R}^M$.

Solution Algorithms

From Eq. 3 we notice that the constrained minimization problem in (5) can be restated as an unconstrained indicated minimization problem

$$\min_{\boldsymbol{U}^h} (\phi + \psi_{\mathbb{R}^{M-}}(\boldsymbol{g}_n)) \tag{6}$$

This unconstrained minimization problem is typically solved using the PM or an augmented Lagrangian based method. The main idea underlying the PM is to approximate the indicator function $\psi_{\mathbb{R}^{M-}}$ by a penalty function leading to the following unconstrained problem

$$\min_{\boldsymbol{U}^h} \left(\phi + \frac{1}{2} r \,\mathbf{proj}^2_{\mathbb{R}^{M-}} (\boldsymbol{g}_n) \right) \tag{7}$$

where r is a penalty parameter which in the limit $r \to +\infty$ leads to the exact indicated problem (6). The main advantages of the PM for contact problems are the lack of additional unknowns in the discrete problem and the extreme simplicity of the corresponding numerical implementation. One important drawback of this method is that the penalty parameter r must take large values to obtain a good approximation of $\psi^{M-}_{\mathbb{R}}$.

Lagrangian based methods are obtained by taking the Fenchel transform of the minimization problem (6) leading to the following constrained saddle point problem

$$\min_{\boldsymbol{U}^h} \max_{\boldsymbol{\lambda} \in \mathbb{R}^{M-}} \Lambda \tag{8}$$

where $\boldsymbol{\lambda}$ is an M-dimensional vector of multiplier (contact forces) and Λ is the standard Lagrangian functional defined by $\Lambda = \phi + \boldsymbol{\lambda} \cdot \boldsymbol{g}_n$. However, this saddle point problem is constrained on the multiplier vector $\boldsymbol{\lambda}$, consistent with the Signorini conditions (2). A fully unconstrained saddle point problem can nonetheless be recovered by using the following augmented Lagrangian functional Λ^r instead of Λ [4, 7]

$$\Lambda^r = \phi + -\frac{1}{2r} \|\boldsymbol{\lambda}\|^2 + \frac{1}{2r} \text{dist}^2(\boldsymbol{\lambda}^r, \mathbb{R}^{M+}) \tag{9}$$

where $\boldsymbol{\lambda}^r = r\,\boldsymbol{g}_n + \boldsymbol{\lambda}$ is an augmented multiplier vector. A key feature of the augmented Lagrangian functional is that the regularization parameter need not be large to exactly enforce the constraints in (6), and therefore leads to well-conditioned problems. The main drawback of the method is the introduction of an additional unknown vector $\boldsymbol{\lambda}$.

The UA is commonly used to solve the augmented Lagrangian saddle point problem. This algorithm makes a clear distinction between the primal variable $\boldsymbol{U}^h$ and the dual multipliers $\boldsymbol{\lambda}$ which are updated independently. The UA starts with an initial guess for the multiplier $\boldsymbol{\lambda}^{\{0\}}$ and then proceeds to minimize Λ^r with respect to the primal variable for a fixed value of the multiplier vector. The multiplier is then updated as the projection of the augmented multiplier on $\mathbb{R}^{M-}$

$$\boldsymbol{\lambda}^{\{k+1\}} = \mathbf{proj}_{\mathbb{R}^{M-}}(\boldsymbol{\lambda}^{r\,\{k\}}) \tag{10}$$

The frequency of the multipliers update defines two variants of the UA [5]: the Simultaneous scheme (the multiplier is updated after each Newton-Raphson iteration of the primal minimization problem) and the Nested schemes (the multiplier is updated only after a certain tolerance on $\nabla_{\boldsymbol{U}^h} \Lambda^r$ has been reached.)

The slow convergence rate typical of UA can be greatly enhanced by designing a global and coupled update strategy for the primal and dual variables. To this end the GNRM is invoked by considering a subgradient of the subdifferential Hessian matrix $\nabla^2 \Lambda^r$ [4, 7] (the gradients of Λ^r are nonsmooth across the gap-contact status line.) Although the Hessian is not definite, it is invertible, at least in a close enough neighborhood of the saddle point [16]. The resulting simultaneous updates on $\boldsymbol{U}^h$ and $\boldsymbol{\lambda}$ lead to quadratic convergence rates on the full problem.

Numerical Examples

The numerical behavior of the three methods previously discussed are assessed by means of two simple canonical contact problems. The first one has only one degree-of-freedom, enabling insights into the behavior of each method without producing undue complications arising from a large system. The second one illustrates how each method performs in the presence of non-trivial changes of the contact status. Results for both problems are obtained using a fully nonlinear symbolic finite element code, `CRUNCH`, written in *Mathematica.*

The first example consists of a continuum plane strain quadrilateral finite element with a single node-on-facet contact element, as shown in Figure 1. The striker node undergoes a prescribed displacement $h = -0.25$ causing a displacement u of the single degree-of-freedom system. From geometric considerations alone, the resulting displacement is found to be $u = -0.5$. We use a Kirchhoff-St.Venant material model with $E = 210$ and $\nu = 0.3$ for both example problems. We assess the convergence prop-

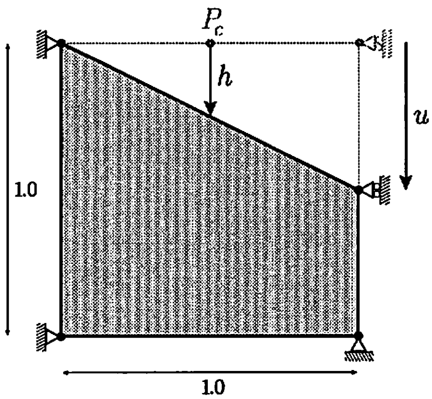

Fig. 1. Single DOF contact problem

erties of PM, UA, and GNRM on the static residual in the L_2-norm. The tolerance used for the PM, simultaneous UA, and GNRM is 10^{-8}. In the nested UA we used a tolerance of 10^{-3} for the displacement iteration loop and 10^{-8} for the dual updates. The initial guess for (u^0, λ^0) was $(0, 0)$ for all cases. Table 1 summarizes the iterations history. In this example the GNRM performed best. For values of

Table 1. Iterations for Example 1

Method	$r = 1$	$r = 10$	$r = 100$	$r = 1000$	$r = 10000$
Penalty	3	4	5	5	cycling
Simultaneous Uzawa	> 500	325	49	cycling	cycling
Nested Uzawa	> 500	> 500	82	160	cycling
		332 updates	48 updates	149 updates	
GNRM	3	3	3	3	divergence

r ranging from 1 to 1000 the method converged to a tolerance of 10^{-8} in only 3 iterations. The PM converged almost as fast, but produced a crude approximation of the correct displacement $u = -0.5$. For example, $r = 100$ resulted in a displacement of $u = -0.0988$ and $r = 1000$ gave $u = -0.387$. The simultaneous and nested UA converged only for values of r between 10 and 1000; moreover, many iterations were needed—exceeding 500 for some cases. For the larger r values, these methods cycled. Both the GNRM and the PM produced quadratic convergence rates. For the nested UA, the convergence on the displacement iteration loop for a given λ was quadratic while the convergence on the dual update was linear at best.

In the second example a symmetric mesh, consisting of twelve continuum finite elements and seven contact elements, was pressed against a rigid foundation. A static load was applied in ten even increments from zero to the values indicated on Figure 2. The contact patterns underwent complex changes as the bar was pressed down, with gap openings near the extremities and gap closings toward the center. As with the one degree-of-freedom example, this problem also converged best with the GNRM.

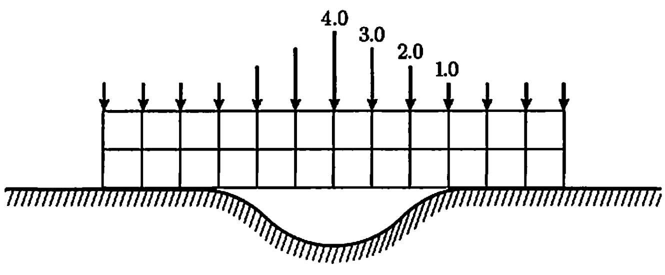

Fig. 2. Simple contact problem with multiple status changes

Results for the iteration history are summarized on Table 2. For all five values of the regularization parameter, GNRM produced quadratic convergence. The solution using the PM converged in about the same number of iterations for values of r between 100 and 10000 but diverged otherwise. The

Table 2. Total Iterations over 10 Time Steps for Example 2

Method	$r = 1$	$r = 10$	$r = 100$	$r = 1000$	$r = 10000$
Penalty	divergence	divergence	87	79	80
Simultaneous Uzawa	divergence	divergence	divergence	divergence	divergence
Nested Uzawa	cycling	733	321	347	120
		338 updates	190 updates	262 updates	42 updates
GNRM	78	80	80	80	80

simultaneous UA always diverged. The nested UA converged for four out of the five r values; however, convergence was slow, taking between 50% and 1000% more total iterations than the GNRM. We further notice that the number of Uzawa updates was generally larger than the total number of GNRM iterations.

Conclusion

The PM and UA are simpler to implement than the GNRM for the augmented Lagrangian formulation. However the PM often fails to convergence (mostly due to cycling in the NR iterations) and only provides approximate solution to the contact problem. On the other hand, the UA and GNRM applied to the augmented Lagrangian functional produce exact enforcement of the constraints. The previous examples indicate however, that the GNRM possesses a much better convergence rate than the UA and requires far less iterations to converge. Although the dimension of the underlying linearized problem is slightly larger in the GNRM, the overall performance of the algorithm by far outperforms the UA in these problems. This trend has been further observed in larger problems involving thousands of DOFs and intricate 3D contact interfaces [17].

Of all the solvers discussed in this paper, we found that the GNRM applied to the augmented Lagrangian formulation was clearly the most efficient and robust algorithm. Further investigations will be necessary to compare GNRM to quasi-Newton schemes applied to UA [18]. The present trends do however suggest that even without counting the primal updates in UA, GNRM requires less iterations than dual updates in UA.

References

1. Wriggers, P. (1995). Finite element algorithms for contact problems. Arch Comput Meth Engng, 2:1–49.
2. Criesfield, M. (1997). Non-linear finite element analysis of solids and structure, vol. 2. Wiley, Chichester, NY.
3. Wriggers, P., Simo, J., and Taylor, R. (1985). Penalty and augmented Lagrangian formulations for contact problems. In Proc. NUMETA '85 Conf.
4. Alart, P., and Curnier, A. (1991). A mixed formulation for frictional contact problems prone to Newton like methods. Comp. Meth. Appl. Mech. Engng., 9:353–375.
5. Simo, J., and Laursen, T. A. (1992). Augmented Lagrangian treatment of contact problems involving friction. Comput. Struct., 42:97–116.
6. Laursen, T. A., and Simo, J. C. (1993). A continuum-based finite element formulation for the implicit solution of multibody, large deformation frictional contact problems. Int. J. Num. Meth. Engng., 36:3451–3485.
7. Heegaard, J. H., and Curnier, A. (1993). An augmented Lagrangian method for discrete large slip contact problems. Int. J. Num. Meth. Engng., 36:569–593.
8. Hestenes, M. (1969). Multiplier and gradient methods. J. Opt. Th. Appl., 4:303–320.
9. Powell, M. J. D. (1969). A method for nonlinear constraints in minimization problems. In Optimization, E. R. Fletcher, Ed. Academic Press, London.
10. Rockafellar, R. (1973). A multiplier method of Hestenes and Powell applied to convex programming. J.Optim. Th. Appl., 12:555–562.
11. Brezzi, F., and Fortin, M. (1991). Mixed and hybrid finite element methods, vol. 15 of Springer series in computational mathematics. Springer-Verlag, New York.
12. Curnier, A., and Alart, P. (1988). A generalized Newton method for contact problems with friction. Méch. Théo. Appl., Suppl.[1]:7:67–82.
13. Laursen, T. A. (1994). The convected description in large deformation frictional contact problems. Int J Solids Structures, 31:669–681.
14. Heegaard, J. H., and Curnier, A. (1996). Geometric properties of 2d and 3d unilateral large slip contact operators. Comp Meth Appl Mech Engng, 131:263–286.
15. Moreau, J. J. (1974). On unilateral constraints, friction and plasticity. In New variational techniques in mathematical physics, CIME 1973, G. Capriz and G. Stampacchia, Eds. Edizioni Cremonese, Roma.
16. Bertsekas, N. (1995). Nonlinear programming. Athena Scientific, Belmont, MA.
17. Heegaard, J., Leyvraz, P., Curnier, A., Rakotomanana, L., and Huiskes, R. (1995). Biomechanics of the human patella during passive knee flexion. J Biomechanics, 28:1265–1279.
18. Laursen, T. A., and Maker, B. N. (1995). An augmented Lagrangian quasi-Newton solver for constrained nonlinear finite element applications. Int J Num Meth Engng, 38:3571–3590.