Proceedings of the 7th GACM Colloquium on Computational Mechanics
for Young Scientists from Academia and Industry
October 11-13, 2017 in Stuttgart, Germany

An exact penalty approach for the finite element solution of frictionless contact problems

Fabian Sewerin[1]* and Panayiotis Papadopoulos[2]

Micro Abstract
Considering a discrete formulation of the frictionless two-body contact problem, we adopt an exact penalty approach in order to enforce the kinematic impenetrability constraints. This approach is based on an augmented discrete force equilibrium and a smooth estimation of the Lagrange multipliers in terms of the nodal displacements. A main feature of the resulting formulation is that an exact enforcement of the impenetrability constraints is achieved for a finite value of the penalty parameter.

[1]Institute of Solid Mechanics, Technical University of Braunschweig, Braunschweig, Germany
[2]Department of Mechanical Engineering, University of California, Berkeley, United States
***Corresponding author**: f.sewerin@tu-braunschweig.de

Introduction

Contact interaction plays a central role in many engineering applications. Frequently, accurate estimates of the contact traction distribution are sought, while the impenetrability constraints ought to be satisfied accurately. These two points are mainly influenced by the evaluation of the contact integral and the choice of constraint enforcement technique. In this work, we focus on the latter and, specifically, present a numerical solution scheme for finite deformation, frictionless contact problems that is based on a so-called differentiable exact penalty formulation of the discrete system. Here, the discretization originates from a standard displacement-based finite element approach and the contact integral is evaluated by invoking surface-on-surface kinematics [12, 16]. The constraint enforcement technique which we adopt can, however, be equally well combined with other discretization techniques or node-on-surface kinematics.

A characteristic in the field of contact mechanics is that constraint enforcement techniques are adopted from methods developed for constrained optimization problems, although, strictly, the governing model formulation cannot always be cast into the form of an optimization problem. However, by virtue of the principle of virtual work [9] and Lagrange's multiplier method, or, equivalently, a weighted residual approach [8], the governing equations may be reformulated in a way that resembles necessary conditions of optimality, thus providing a link to constrained optimization techniques.

In the method of Lagrange multipliers, these necessary conditions of optimality are solved directly [1, Section 4.4] as an augmented system expressed in terms of discrete displacements and multipliers. Compared to an unconstrained formulation, the presence of multipliers entails an increase in system size and the tangent stiffness matrix is characterized by an indefinite structure [14]. Recently, Popp et al. [12, 13] showed that these shortcomings may be circumvented by imposing a biorthogonality condition in the contact integral as this allows for a static condensation of the multipliers within each equilibrium iteration.

In a similar vain, albeit by different means, the main idea with methods from the class of penalty-based constraint enforcement techniques is to estimate the multipliers in terms of the displacements, while omitting a direct constraint enforcement. This class of methods includes the perturbed Lagrangian approach [15] and the quadratic penalty approach, for instance, which achieve an exact constraint enforcement only in the limit of an infinitely large penalty parameter. In practice, large penalties often entail ill-conditioning of the tangent stiffness matrix and, although the ill-conditioning may be mitigated to some extend [1], practical limitations of the accuracy with which constraints can be satisfied remain. In the method of multipliers [14], on the other hand, a sequence of unconstrained equilbrium problems is developed whose limit solution exactly obeys the contact constraints for a finite value of the penalty parameter [7].

Based on the success of the method of multipliers, Fletcher [2–4] and Fletcher and Lill [5] advanced the rationale in the early 1970s that if it is possible to construct converging sequences of displacements and corresponding multipliers, then it might be feasible to continuously estimate the multipliers in terms of the displacements. From a contact mechanical perspective, this leads to the notion of an equilibrium system that is based on a differentiable exact penalty function. Here, a penalty parameter also occurs, albeit, as in the method of multipliers, convergence is ensured for a finite value. While it proves difficult to determine the minimum such value *a priori*, Mukai and Polak [10] and Glad and Polak [6] found that the equilibrium scheme may be complemented by an automatic penalty update step which increases the penalty as soon as it appears that the solution converges to a limit point which violates the contact constraints [11]. From an implementational viewpoint, one drawback of the exact penalty approach is that second derivatives of the contact constraints appear in the equilibrium system.

In the present work, we adopt the differentiable exact penalty formulation due to Glad and Polak [6] and examine the implications of its application to the finite deformation two-body contact problem. Further to rationalizing the exact penalty solution scheme, our focus lies on examining the computational expense it entails, the accuracy with which the contact constraints can be imposed and the conditioning of the tangent stiffness matrix.

1 A differentiable exact penalty formulation

In discrete terms, the Galerkin formulation of the finite deformation two-body contact problem may be cast into the following form

$$\mathbf{f}(\mathbf{d}) + \nabla\mathbf{g}(\mathbf{d})^T\boldsymbol{\lambda} = \mathbf{0} \tag{1}$$

subject to the constraints

$$\mathbf{g}(\mathbf{d}) \geq \mathbf{0}, \tag{2}$$

$$\boldsymbol{\lambda} \leq \mathbf{0}, \tag{3}$$

$$\lambda_i g_i(\mathbf{d}) = 0 \quad \forall i \in \mathcal{S}, \tag{4}$$

where $\mathbf{f}(\mathbf{d})$ represents the vector of internal/external forces, $\mathbf{g}(\mathbf{d}) = (g_i(\mathbf{d}))_{i\in\mathcal{S}}$ is the (weighted) gap vector, $\boldsymbol{\lambda} = (\lambda_i)_{i\in\mathcal{S}}$ denotes the vector of Lagrange multipliers and $\mathcal{S}$ is an index set for labelling the inequality constraints in Eqs. (2) through (4). For any $c > 0$ [6, Lemma 1], these equations are equivalent to the pure equality constraints

$$\mathbf{a}_c(\mathbf{d}, \boldsymbol{\lambda}) = \frac{1}{c}\left(\min(\mathbf{0}, \boldsymbol{\lambda} + c\mathbf{g}(\mathbf{d})) - \boldsymbol{\lambda}\right) = \mathbf{0}, \tag{5}$$

which we consider for the remaining part of this abstract.

Similar to the quadratic penalty approach or the method of multipliers, the exact penalty approach is based on an unconstrained augmented force equilibrium

$$\mathbf{r}_c(\mathbf{d}) = \mathbf{f}(\mathbf{d}) + \nabla(\boldsymbol{\lambda}_c(\mathbf{d})^T\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d}))) = \mathbf{0}, \tag{6}$$

where c represents a penalty parameter and the multipliers $\boldsymbol{\lambda} = \boldsymbol{\lambda}_c(\mathbf{d})$ are estimated in terms of the displacements according to

$$\boldsymbol{\lambda}_c(\mathbf{d}) = \boldsymbol{\phi}(\mathbf{d}) + \frac{c}{2}\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d})). \tag{7}$$

The form of the function $\boldsymbol{\phi}(\mathbf{d})$ which appears herein may be rationalized in different ways. For example, considering equality constrained problems, Fletcher [2] proposed to require that the projection of $\mathbf{f}(\mathbf{d})$ onto the tangent space of the contraint manifold $\mathbf{g}(\mathbf{d}) = \mathbf{0}$ (whose normal is $\nabla\mathbf{g}(\mathbf{d})$) vanishes in the limit as $\mathbf{g}(\mathbf{d}) = \mathbf{0}$. His suggestion for $\boldsymbol{\phi}(\mathbf{d})$ coincides with the best least-squares solution (attributed to Powell) of

$$\mathbf{f}(\mathbf{d}) + \nabla\mathbf{g}(\mathbf{d})^T\mathbf{z} = \operatorname*{Min}_{\mathbf{z}}! \tag{8}$$

For inequality constraints as in contact mechanics, Glad and Polak [6] slightly augmented this rationale to prevent the occurrence of discontinuous derivatives,

$$\boldsymbol{\phi}(\mathbf{d}) = \arg\min_{\mathbf{z}} \left\{ \left\|\mathbf{f}(\mathbf{d})\nabla\mathbf{g}(\mathbf{d})^T\mathbf{z}\right\|^2 + \mathbf{z}^T\mathbf{G}(\mathbf{d})\mathbf{z} \right\}, \tag{9}$$

where $\mathbf{G}(\mathbf{d})$ is a diagonal matrix with entries $(g_i(\mathbf{d})^2)_{i\in\mathcal{S}}$. Eq. (9) possesses the following unique solution

$$\left(\nabla\mathbf{g}(\mathbf{d})\nabla\mathbf{g}(\mathbf{d})^T + \mathbf{G}(\mathbf{d})\right)\boldsymbol{\phi}(\mathbf{d}) = -\nabla\mathbf{g}(\mathbf{d})\mathbf{f}(\mathbf{d}). \tag{10}$$

The basis of the exact penalty approach is the result that there exists a finite penalty $\bar{c}$ such that if $\mathbf{d}$ presents a solution of Eqs. (6), (7) and (10) for some $c \geq \bar{c}$, then the pair $(\mathbf{d}, \boldsymbol{\lambda} = \boldsymbol{\lambda}_c(\mathbf{d}))$ solves the constrained formulation in Eqs. (1) through (4). In order to circumvent the challenge of estimating $\bar{c}$ *a priori*, Mukai and Polak [10] pioneered the idea of increasing the penalty parameter in the course of the equilibrium scheme if the sequence $\mathbf{d}^k$, $k = 0, 1, \ldots$, constructed by the non-linear system solver appears to converge to a solution of Eq. (6) that violates Eq. (5) (or, equivalently, Eqs. (2) through (4)). To this end, $c = c_k$ is chosen such that, in every iteration k, the following condition is obeyed

$$t_{c_k}(\mathbf{d}^k) \leq 0 \tag{11}$$

and the sequence $\{c_k\}$ is non-decreasing. Here, $t_c(\mathbf{d})$ is termed a test function; it is constructed such that $t_c(\mathbf{d}) \leq 0$ and $\mathbf{r}_c(\mathbf{d}) = \mathbf{0}$ imply satisfaction of our original problem in Eqs. (1) through (4) and that, for any $\mathbf{d}'$ and a whole neighbourhood about $\mathbf{d}'$, $t_c(\cdot)$ can be rendered non-positive by choosing a sufficiently large penalty c [6]. We emphasize that the scheme in Eq. (11) is conservative, that is, the penalty chosen in this way may, ultimately, overestimate $\bar{c}$. However, as we briefly note in the following section, a small penalty does not always imply well-conditioning of the tangent stiffness matrix.

In a slight amendment to the original rule of Mukai and Polak [10], we base the test function on the physical rationale that the equilibrium step $-\mathbf{K}_c(\mathbf{d})^{-1}\mathbf{r}_c(\mathbf{d})$, where $\mathbf{K}_c(\mathbf{d})$ represents an (approximate) tangent stiffness matrix, points, to some degree, in the same direction as a Newton step $\mathbf{v}_c(\mathbf{d}) = -\nabla\mathbf{a}_c(\nabla\mathbf{a}_c\nabla\mathbf{a}_c^T)^{-1}\mathbf{a}_c$ towards the constraint surface $\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d})) = \mathbf{0}$,

$$\mathbf{v}_c^T(-\mathbf{K}_c^{-1}\mathbf{r}_c) \geq \mathbf{a}_c^T(\nabla\mathbf{a}_c\nabla\mathbf{a}_c^T)^{-1}\mathbf{a}_c \quad \Leftrightarrow \quad t_c \equiv \mathbf{a}_c^T(\nabla\mathbf{a}_c\nabla\mathbf{a}_c^T)^{-1}\mathbf{a}_c + \mathbf{v}^T\mathbf{K}_c^{-1}\mathbf{r}_c \leq 0. \tag{12}$$

Here, the matrices $(\nabla\mathbf{a}_c\nabla\mathbf{a}_c^T)^{-1}$ and $\mathbf{K}_c(\mathbf{d})^{-1}$ mainly serve dimensional consistency, although we found $\mathbf{K}_c(\mathbf{d})^{-1}$ to act like a pre-conditioner on a steepest-descend step. In practice, we resolved to approximate $\mathbf{K}_c(\mathbf{d})^{-1}$ (see Eq. (13) below) by the inverse of its diagonal as this maintains the key c-dependency of the equilibrium step. After the equilibrium scheme terminated, the penalty is reset to unity, $c_0 = 1$.

As a final point, we briefly turn to the linearization of $\mathbf{r}_c(\mathbf{d})$ for the purpose of a Newton-based equilibrium scheme. Strictly, since $\nabla\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d}))$ is discontinuous at points $(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d}))$ at which

$\phi_i(\mathbf{d}) + cg_i(\mathbf{d})$ vanishes for at least one $i \in \mathcal{S}$ (that is, strict complementarity is violated), the Hessian of the constraints $\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d}))$ remains undefined at these points. Recall that this stiffness discontinuity reflects the potential for constraints to flip-flop in the context of active set searches. As a remedy for the stiffness discontinuity, Glad and Polak [6] resolved to compute separately the tangent stiffness matrices for the cases that all constraints are either active ($\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d})) = \mathbf{g}(\mathbf{d})$ in Eq. (5)) or inactive ($\mathbf{a}_c(\mathbf{d}, \boldsymbol{\phi}(\mathbf{d})) = -\boldsymbol{\phi}(\mathbf{d})/c$) and to approximate the true tangent stiffness matrix by a weighted average of both. Here, the weights are chosen such that if $\mathbf{d}$ approaches a solution at which strict complementarity holds, then the approximate tangent stiffness reproduces the exact one. Specifically, we have

$$\mathbf{K}_c = \nabla \mathbf{f} + \nabla \mathbf{g}^T \mathbf{B} \nabla \boldsymbol{\phi} + (\nabla \boldsymbol{\phi} + c \nabla \mathbf{g})^T \mathbf{B} \nabla \mathbf{g} + \boldsymbol{\phi}^T \mathbf{B} \nabla (\nabla \mathbf{g}) - \frac{1}{c} \nabla \boldsymbol{\phi}^T (\mathbf{I} - \mathbf{B}) \nabla \boldsymbol{\phi}, \qquad (13)$$

where $\mathbf{B}$ is a diagonal matrix of weighting factor $b_i(\mathbf{d})$, $i \in \mathcal{S}$, and arguments have been omitted for brevity. Also, second derivatives of multipliers and the term $c\mathbf{g}^T \mathbf{B} \nabla (\nabla \mathbf{g})$ have been evicted from Eq. (13). In our experience, a Newton-type scheme based on $\mathbf{K}_c(\mathbf{d})$ in Eq. (13) is not noticably impaired by these omissions and recovers a quadratic rate of convergence ultimately.

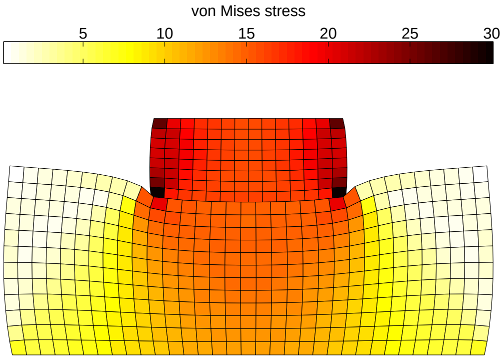
2 An example problem

In order to assess the accuracy and benefits of the exact penalty approach, we consider the indentation problem depicted in Figure 1(a). Here, the top rectangle ($W = 2$, $H = 1$, $E = 100$, $\nu = 0.3$) is vertically displaced by -0.6 units of length and pressed into a soft foundation ($W = 5$, $H = 2$, $E = 50$, $\nu = 0.3$) over the course of four displacement steps. W and H indicate the widths and heights of the rectangles, while E and ν represent, respectively, Young's modulus and Poisson's ratio of a Neo-Hookean material. The rectangles are discretized using 508 standard displacement-based quadrilateral finite elements. Exemplarily, Figure 1(b) shows the convergence in terms of the maximum violation of the active impenetrability constraints and the absolute residual norm for the first displacement step. In the course of the first iteration, the penalty parameter c is automatically increased from unity to 1024, while, in subsequent steps, it is changed from unity to 256, 2.6×10^5 and 64, respectively. As indicated above, these limit penalties are rather conservative; indeed, we found that solutions with accurate constraint enforcement can also be obtained by keeping c constant at unity throughout.

In equilibrium, the 1-norm condition numbers of the unconstrained tangent stiffness and the exact penalty stiffness (Eq. (13)) are comparable. In general, we observed that, for the exact penalty scheme, a small penalty parameter does not guarantee that the tangent stiffness matrix is well-conditioned. This is enhanced by the observation that the active and inactive contributions to the tangent stiffness may largely and, sometimes, adversely affect the conditioning of the cumulative tangent stiffness. In practice, there appears to be an optimal choice of penalty which balances these influences and, for the cases we examined, the automatic update scheme seems to drive the penalty towards this value. Concomitantly, convergence may be accelerated if the penalty renders the tangent stiffness matrix well-conditioned.

Conclusions

In this abstract, we examined the rationale underlying a constraint enforcement technique based on a differentiable exact penalty function in the context of the finite deformation two-body contact problem. This technique is based on a continuous approximation of the multipliers in terms of the discrete displacements and achieves an exact enforcement of the contact constraints for a finite value of the penalty parameter. A main feature is that the penalty parameter is automatically increased if iterates seem to approach an equilibrated solution that violates the contact constraints. From an algorithmic perspective, both the active set search and the automatic penalty update scheme are merged inside a single unconstrained equilibrium scheme. On the minus side, we found that the computationally expense is enhanced by the analytical



k	c_k	$\max_{i \in \mathcal{S}} \|g_i(\mathbf{d}^k)\|$	$\|\mathbf{r}_c(\mathbf{d}^k)\|$
0	1	–	8.51×10^{1}
$1^\star$	128	7.1×10^{-3}	2.08×10^{1}
2	1024	5.4×10^{-5}	2.80×10^{-1}
3	1024	2.6×10^{-6}	8.22×10^{-3}
4	1024	2.0×10^{-8}	5.43×10^{-5}
5	1024	1.2×10^{-12}	2.35×10^{-9}

(a) Deformed configuration

(b) Convergence of the equilibrium scheme

Figure 1. A stiff indenter is pressed into a soft foundation. Figure (a) depicts a contour plot of the von Mises stress (dimensionless) in the deformed configuration, while figure (b) illustrates the convergence of the exact penalty equilibrium scheme for the first displacement step. Here, a $\star$ indicates a change in the active set.

evaluation of second derivatives of the contact constraints (which appear in the augmented equilibrium system) and the solution of additional linear systems to estimate the multipliers and their derivatives.

Based on a simple indentation problem, we illustrated the accuracy of the exact penalty constraint enforcement technique and provided evidence of the well-conditioning of the tangent stiffness matrix.

Acknowledgements

Fabian Sewerin gratefully acknowledges the financial support by the German-American Fulbright Association.

References

[1] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Athena Scientific, 1996.

[2] R. Fletcher. A class of methods for nonlinear programming with termination and convergence properties. In J. Abadie, editor, *Integer and Nonlinear Programming*, chapter 6, pages 157–175. North-Holland Publishing Company, Amsterdam, 1970.

[3] R. Fletcher. A class of methods for nonlinear programming III: Rates of convergence. In F. A. Lootsma, editor, *Numerical Methods for Nonlinear Optimization*, pages 371–382. Academic Press, New York, 1972.

[4] R. Fletcher. An exact penalty function for nonlinear programming with inequalities. *Mathematical Programming*, 5(1):129–150, 1973.

[5] R. Fletcher and S. A. Lill. A class of methods for non-linear programming II: Computational experience. In J. B. Rosen, O. L. Mangasarian, and K. Ritter, editors, *Nonlinear Programming*, pages 67–92. Academic Press, New York, 1971.

[6] T. Glad and E. Polak. A multiplier method with automatic limitation of penalty growth. *Mathematical Programming*, 17(1):140–155, 1979.

[7] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.

[8] R. E. Jones and P. Papadopoulos. A novel three-dimensional contact finite element based on smooth pressure interpolations. *International Journal for Numerical Methods in Engineering*, 51(7):791–811, 2001.

[9] C. Lanczos. *The Variational Principles of Mechanics*. Dover Books on Physics. Dover Publications, Inc., New York, 4 edition, 1970.

[10] H. Mukai and E. Polak. A quadratically convergent primal-dual algorithm with global convergence properties for solving optimization problems with equality constraints. *Mathematical Programming*, 9(1):336–349, 1975.

[11] E. Polak. On the global stabilization of locally convergent algorithms. *Automatica*, 12(4):337–342, 1976.

[12] A. Popp, M. W. Gee, and W. A. Wall. A finite deformation mortar contact formulation using a primal-dual active set strategy. *International Journal for Numerical Methods in Engineering*, 79(11):1354–1391, 2009.

[13] A. Popp, M. Gitterle, M. W. Gee, and W. A. Wall. A dual mortar approach for 3D finite deformation contact with consistent linearization. *International Journal for Numerical Methods in Engineering*, 83(11):1428–1465, 2010.

[14] J. C. Simo and T. A. Laursen. An augmented Lagrangian treatment of contact problems involving friction. *Computers & Structures*, 42(1):97–116, 1992.

[15] J. C. Simo, P. Wriggers, and R. L. Taylor. A perturbed Lagrangian formulation for the finite element solution of contact problems. *Computer Methods in Applied Mechanics and Engineering*, 50(2):163–180, 1985.

[16] B. Yang, T. A. Laursen, and X. Meng. Two dimensional mortar contact methods for large deformation frictional sliding. *International Journal for Numerical Methods in Engineering*, 62(9):1183–1225, 2005.

Frictionless contact problems in two-dimensional space are formulated by complementarity theory, where the system of equations is established by the nonlinear complementary functions and boundary integral equations. This algorithm by BEM is established. The accuracy and effectiveness of the method have been demonstrated by two numerical examples, and the effect of discretization has also been studied in the Hertzian contact problem.Â P. Papadopoulos and J. M. Solberg, â€œA Lagrange multiplier method for the finite element solution of frictionless contact problems,â€ Mathematical and Computer Modelling, vol. 28, no. 4â€"8, pp. 373â€"384, 1998. View at Publisher Â· View at Google Scholar Â· View at Scopus. Numerical solutions for the frictionless and frictional contact problems are compared with the results obtained by using a general-purpose finite element program ANSYS (that uses variational equality formulation). ANSYS results match reasonably well with the solutions of KKT optimality conditions for the frictionless contact problem and the two-phase procedure for the frictional contact problem. The validity of the analytical formulation for frictional contact problems (with one contacting node) is verified. Thevariational equality formulation for frictionless and frictional, contact problems Micro Abstract: Considering a discrete formulation of the frictionless two-body contact problem, we adopt an exact penalty approach in order to enforce the kinematic impenetrability constraints. This approach is based on an augmented discrete force equilibrium and a smooth estimation of the Lagrange multipliers in terms of the nodal displacements. A main feature of the resulting formulation is that an exact enforcement of the impenetrability constraints is achieved for a finite value of the penalty parameter. Show Extended Abstract.