Chapter 1
On regularization of the convergence path for the implicit solution of contact problems

Giorgio Zavarise[1], Laura De Lorenzis[1], Robert L. Taylor[2]
[1]Università del Salento - Lecce, Italy, [2]University of California at Berkeley - USA.

There is no good science without good friendships. To Peter, for his 60th birthday, and for the long years of friendship.

Following up to a previous investigation, this paper proposes a strategy to deal with frictionless contact problems involving large penetrations, in the context of the node-to-segment formulation and of the penalty method. The rationale is based on two main considerations: the first one is that, within an iteration scheme, the use of consistent linearization is only convenient when the field of the unknowns is sufficiently close to the solution point; the second one is that, if the order of magnitude of the maximum contact pressure can be estimated a priori, this information can be exploited to approach the solution in a faster and more reliable way. The proposed strategy is based on a check of the nodal contact pressure, to select the technique that has to be used to perform each iteration. If the contact pressure is smaller than a predefined limit, the problem is solved in the standard way, using consistent linearization and Newton's method. When the contact pressure exceeds the limit, a modified method is used. This one is based on the enforcement of a contact pressure limit and on the use of a simplified secant stiffness, where the geometric stiffness term is disregarded. The strategy has to be integrated with a specific "safeguard algorithm" to guarantee convergence to the correct solution also in cases where the maximum contact pressure has been underestimated. Two alternative procedures for this purpose are proposed.

1.1 Introduction

The benefits of consistent linearization for the solution of any type of non-linear computational problem are well known. However, consistent linearization guarantees a quadratic rate of asymptotic convergence, provided that **the field of the unknowns is close to the solution value**. It is also true that, the bigger the loading or time step used in the analysis, the greater is the distance of the starting point from the solution one. This aspect is usually disregarded in classical solution schemes, and standard Newton procedures are commonly advocated from the first iteration.

Contact algorithms are usually activated a posteriori, i.e. they first let the bodies penetrate each other, then they detect the penetration as a violation of the impenetrability constraint conditions. Regardless of the method adopted for the enforcement of the contact constraint, the current state which violates the impenetrability condition is used to compute the contact contribution to the virtual work. In the implicit scheme this equation set is linearized to solve the non-linear problem with a Newton type method. The strategy is usually effective for small penetrations and smooth evolution of the contact forces. In this case, convergence is usually achieved. However, the first few iterations are needed simply to stabilize the solution before a quadratic rate of convergence takes place. More importantly, if large penetrations occur between the contacting bodies due to large loading or time steps, the direct application of Newton's method usually produces significant difficulties during the first few iterations. In such cases, the contact pressures resulting from the penetrations can be very large. This affects both the tangent stiffness and the residual vector, and generally produces a large local distortion of the mesh. Quite often a Newton method cannot recover to a smooth deformation state, and catastrophic divergence ensues. Therefore, it is currently necessary to use smaller loading or time increments which limit the amount of penetration.

Following up to a previous investigation [5], this paper proposes a strategy both to perform large steps in the presence of large penetrations and to increase the convergence rate in case of normal penetrations. The rationale of the strategy is based on two main considerations. The first one is that the use of consistent linearization is only convenient when the field of the unknowns is sufficiently close to the solution point. The second one is that, if the order of magnitude of the maximum contact pressure can be estimated a priori (as usually happens in engineering problems), then this information can be exploited to approach the solution in a faster and more reliable way. In the proposed strategy, the solution is split into two phases and a different method is used for each one. Phase one takes place during the first iterations of each time or loading step. Within this phase, it is straightforward to demonstrate that the full Newton strategy with consistent linearization is often useless, as the contact pressures can be many orders of magnitude larger than the real ones. Phase two takes place when, due to the iterations performed during phase one, the contact penetrations have been significantly reduced. Hence the problem has been driven close to the solution point, then a Newton strategy with consistent linearization guarantees the best possible convergence rate. A smooth transition occurs between the two solution phases.

For more details about the proposed procedure, its background, and additional examples, see [4].

1.2 Structure of the consistent tangent stiffness

For a better understanding of what happens during the first iterations using consistent linearization, we start by analyzing the characteristics of the consistent tangent stiffness matrix and of the residual vector for a typical contact problem. The penalty method is considered herein, and the formulation is developed in the framework of the node-to-segment algorithm. However, the proposed approach could be easily extended to other algorithms, such as mortar type methods (see e.g. [1]). For simplicity, the methodology will be presented with reference to 2D problems. For more details, see also [4-6].

The matrix form for the residual and for the consistent tangent stiffness can be established as, respectively,

$$\mathbf{R} = -A_c \varepsilon g_N \mathbf{N}_\mathrm{S} \tag{1.1}$$

and

$$\mathbf{K}_\mathrm{T} = \mathbf{K}_\mathrm{M} + \mathbf{K}_\mathrm{G}^{(\mathrm{a})} + \mathbf{K}_\mathrm{G}^{(\mathrm{b})} \tag{1.2}$$

where

$$\mathbf{K}_\mathrm{M} = A_c \varepsilon \mathbf{N}_\mathrm{S} \mathbf{N}_\mathrm{S}^\mathrm{T} \qquad \mathbf{K}_\mathrm{G}^{(\mathrm{a})} = -\frac{A_c \varepsilon g_N}{l}\left(\mathbf{N}_0 \mathbf{T}_\mathrm{S}^\mathrm{T} + \mathbf{T}_\mathrm{S} \mathbf{N}_0^\mathrm{T}\right) \qquad \mathbf{K}_\mathrm{G}^{(\mathrm{b})} = -\frac{A_c \varepsilon g_N^2}{l^2} \mathbf{N}_0 \mathbf{N}_0^\mathrm{T} \tag{1.3}$$

In the above equations, g_N is the normal penetration, evaluated at the slave node, A_c is the area of competence of the slave node, ε is the penalty parameter, l is the length of the master segment, and the vectors are defined as follows

$$\mathbf{N}_\mathrm{S} = \begin{bmatrix} \mathbf{n} \\ -(1-\xi)\,\mathbf{n} \\ -\xi \mathbf{n} \end{bmatrix} \qquad \mathbf{N}_0 = \begin{bmatrix} \mathbf{0} \\ -\mathbf{n} \\ \mathbf{n} \end{bmatrix} \qquad \mathbf{T}_\mathrm{S} = \begin{bmatrix} \mathbf{t} \\ -(1-\xi)\,\mathbf{t} \\ -\xi \mathbf{t} \end{bmatrix} \tag{1.4}$$

where $\mathbf{n}$ and $\mathbf{t}$ are the unit normal and tangent vectors to the master segment, and ξ is the normalized projection of the slave node onto the master segment. Component $\mathbf{K}_\mathrm{M}$ stems from the contribution related to $\triangle g_N \delta g_N$ in the contact contribution to the virtual work, and can be considered the core part of the stiffness. This part is clearly independent of the amount of penetration. It is very important to note that all terms of this matrix have bounded values. Hence, no matter what is the current value of the penetration during the iteration process, the terms of $\mathbf{K}_\mathrm{M}$ are bounded to well-defined limits, and may only become large as a result of a large penalty parameter multiplying the matrix.

Components $\mathbf{K}_\mathrm{G}^{(\mathrm{a})}$ and $\mathbf{K}_\mathrm{G}^{(\mathrm{b})}$ stem from the term related to $\triangle \delta g_N$ in the contact contribution to the virtual work. It can be shown that $\mathbf{K}_\mathrm{G}$ can be interpreted as the geometric component of the contact stiffness. More in detail, it is related to the change of orientation of the master segment and it vanishes in case of a constant (fixed) normal vector. Once again, the terms in the matrices (see eqs. 1.3b and 1.3c) have bounded values. However, the factors multiplying the matrices depend strongly on the amount of penetration. In particular, the $\mathbf{K}_\mathrm{G}^{(\mathrm{b})}$ component is proportional to the square of g_N. In case of large penetrations it is evident that this contribution to the stiffness can be several orders of magnitude different from its value close to the solution point.

With the reduction of the penetration, the first term of the stiffness becomes more and more important with respect to the second one. Despite the presence of the second term is necessary for the quadratic rate of convergence, the dominating stiffness term when convergence is achieved is the first one, at least for usual values of the penalty parameter. It is then clear that, in case of unrealistic contact forces due to large penetrations, the geometric stiffness term is both useless, because the geometry is simply too far from the final one, and dangerous, because it strongly affects the local properties of the stiffness matrix.

1.3 Large penetration basic algorithm

1.3.1 Strategy outline

In the proposed strategy the solution is split into two phases, and a different method is used for each one. Phase one takes place during the first iterations of each time or loading step. Within this phase, the full Newton strategy with consistent linearization is often useless, as the problem is very far from the solution point and the contact pressures can be many orders of magnitude larger than the real ones. Moreover, due to a rotation of the master segment, the geometric term of the consistent tangent stiffness matrix may feature large negative diagonal terms, with consequent catastrophic effects on the convergence performance. Phase two takes place when, due to the iterations performed during phase one, the contact penetrations have been significantly reduced. The problem has thus been driven close to the solution point, hence a Newton strategy with consistent linearization guarantees the best possible convergence rate. A smooth transition occurs between the two solution phases.

The objective of this work is thus to set up a strategy performing better than a consistent full Newton linearization for the aforementioned phase one. What is needed is a criterion to limit the contact pressures and to construct a modified contact contribution to the stiffness matrix and to the residual vector. With this method, local contact instabilities can be controlled, and a fast solution path to an almost converged point can be constructed.

The basic idea of the proposed large penetration (LP) strategy consists in avoiding the introduction into the system of large, physically meaningless contact forces which originate from unconstrained large penetrations [5]. For this purpose, it is sufficient to estimate a priori the order of magnitude of the maximum contact pressure that the contacting bodies may experience. Such an estimate represents an information which is easily available to the engineers. It can be shown that estimates of the correct maximum contact pressure within one order of magnitude, or even more, are sufficient for the present purposes. Therefore the availability of the estimate is not a limiting assumption. The estimated value is then used to set a bound for the contact pressures. The employment of the upper bound, combined with some modifications of the standard Newton procedure, permits the execution of steps of unusual size. When using the proposed strategy, the limit on the step size is generally due to the large distortion of the continuum elements, and not to the contact.

1.3.2 Modified stiffness and residual during phase one

To overcome the difficulties cited earlier, two modifications are made to the standard Newton procedure during phase one of the solution. The first modification consists in disregarding the geometric stiffness term during this phase. The resulting expression of the tangent stiffness matrix is then

$$\mathbf{K}_{\mathrm{T,mod}} = \mathbf{K}_{\mathrm{M}} = A_c \varepsilon \mathbf{N}_{\mathrm{S}} \mathbf{N}_{\mathrm{S}}^{\mathrm{T}} \tag{1.5}$$

This term is the starting point to build the phase one stiffness matrix. In order to do this, we have to focus first on the second modification. This is related to the contact force which goes into the residual. For large penetrations, the contact forces computed as $A_c \varepsilon g_N$ are grossly in error. The redistribution of these forces to the nodes by the vector $\mathbf{N}_{\mathrm{S}}$, as per eq. (1.1), is the second element of instability of the solution. To limit this force or, equivalently, the corresponding pressure, we propose to modify the linear relationship between contact pressure and normal penetration by using a cut-off with a maximum value independent of the penetration, p_{ME} (see Figure 1.1). In this way the introduction of unrealistic forces into the system is prevented. Note that p_{ME} is an estimated value of the maximum contact pressure arising between the contacting surfaces at convergence. Hence, regardless from the amount of penetration, during phase one the contact pressure computed as $p_N = \varepsilon g_N$ is replaced with the cutoff pressure, p_{ME}.

It has to be remarked that the cut-off alone is not sufficient to perform large steps. In fact, if consistent linearization is used we get a zero derivative when the cut-off limit is enforced. Hence no contact stiffness is associated to the residual, even if penetration persists. The contact forces are then applied without any contact resistance. Once again this will have dangerous effects, because the resulting displacements in most cases lead to release, and then a new instability often takes place with part of the contacting surfaces switching from an open to a closed status during one iteration and viceversa for the next one.

A very good performance has been achieved for phase one by introducing a secant stiffness. In most cases the secant stiffness is able to keep the gap closed and rapidly relax the contact conditions to approach values of the penetration for which consistent linearization can then be employed. The secant stiffness is related to the amount of penetration and to the maximum contact force by a variable penalty, computed at each iteration as

$$\varepsilon_s = \frac{p_{ME}}{g_N} \tag{1.6}$$

This is then used directly in eq. (1.5) in place of ε, see also Figure 1.1. It is worth noting that, due to the constant pressure limit, the secant stiffness depends only on the penetration and increases with a reduction of the penetration, until it reaches the standard penalty value. Subsequently the contact solution procedure shifts smoothly from phase one to phase two, where standard consistent linearization is performed.

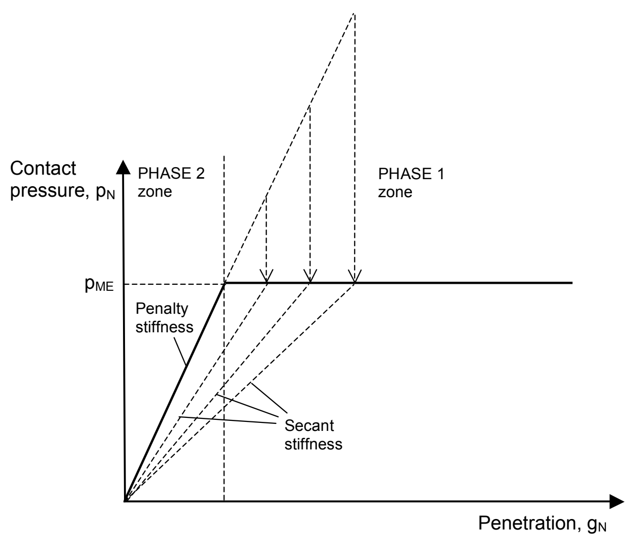


Fig. 1.1: Use of the secant stiffness.

1.3.3 Limitations of the strategy

For the next discussion, it is useful to introduce the following definition

$$r = \frac{p_{ME}}{p_{MR}} \tag{1.7}$$

where p_{MR} is the value of the maximum contact pressure at convergence. The parameter r is therefore the ratio of the estimated to the real maximum contact pressure. The r value is > 1 or < 1 if the maximum contact pressure is, respectively, overestimated or underestimated.

If the correct maximum contact pressure is overestimated by a reasonable amount, the LP algorithm performs very efficiently. If the r ratio is excessively large, convergence may be no longer achieved. This clearly results from a too early shift into phase two, which implies the introduction into the system of still large, unrealistic contact forces like for the standard consistent linearization. If the r ratio is large enough, the contact pressure limit may even exceed the product of the maximum initial (unchecked) penetration by the penalty parameter. In such a case the LP algorithm is not even activated, hence standard consistent linearization is always performed.

For any value of $r > 1$, provided that convergence is achieved, the converged solution is the correct one. "Correct solution" is here intended as the solution whose degree of approximation is related to the chosen value of the penalty parameter, ε. Hence, there is no influence of the r ratio on the accuracy of the converged solution, which only depends on the value of ε. The correctness of the converged solution is proved by the absence of contact elements which are still in phase one at convergence.

As r gets closer to 1, or even $r = 1$ (i.e. if the correct value of the maximum contact pressure is known a priori), the number of iterations to convergence increases. The reason is that, as the limit pressure decreases, phase two is reached at a very late stage, hence most iterations are performed with a sub-quadratic rate of convergence.

For $r < 1$, i.e. if the maximum contact pressure is underestimated, it is obviously not possible to obtain the correct solution, as the correct distribution of contact pressures includes values which are larger than the imposed limit. In such cases, typically convergence is achieved, however there is a number of contact elements that have not reached phase two at convergence. In other words, the problem converges to an incorrect solution, i.e. to a solution whose degree of approximation is worse than that related to the chosen penalty parameter. Therefore, it is necessary to add one more feature to the LP strategy, to be used when $r < 1$.

1.4 Large penetration enhanced algorithm

1.4.1 Solution of the problem for $r < 1$

Two alternative procedures to solve the problem described in Section 1.3.3 have been tested in this study and are described herein. The need to activate one of these procedures for a given contact element is identified by monitoring the evolution of the contact status between subsequent iterations. An easy criterion is a check of the variation of the normal penetration, e.g. through the following quantity

$$R_g = \begin{cases} \left| 1 - \frac{g_N^{(i)}}{g_N^{(i-1)}} \right| & \text{if } g_N^{(i)} < g_N^{(i-1)} \\ \left| 1 - \frac{g_N^{(i-1)}}{g_N^{(i)}} \right| & \text{if } g_N^{(i)} > g_N^{(i-1)} \end{cases} \tag{1.8}$$

where $g_N^{(i-1)}$ and $g_N^{(i)}$ are the values of the penetration at the iterations number $(i-1)$ and i, respectively. From the definition, it follows that $0 \le R_g \le 1$. If R_g is close to the unity, the penetration is rapidly changing within the iterations and no special procedure is needed. Conversely, a small value of R_g (close to zero) indicates that the penetration is undergoing very little change between subsequent iterations. During phase one this occurs when, due to the underestimation of p_{MR}, the contact pressures are forced to remain below the too strict limit imposed by the LP algorithm. In such a case, a special procedure needs to be activated to guarantee convergence to the correct solution. Two procedures for this purpose are illustrated as follows.

The first proposed procedure consists in increasing the pressure limit within the iterations. The resulting algorithm is indicated as LP-IP. The evolution of the contact status has to be monitored to decide whether and when the increment can take place. Whenever R_g is less than a specified threshold, the contact pressure limit for that contact element is updated. The update strategy consists in adding the initial value of the pressure limit to its current value, i.e.

$$p_{ME}^{(u)} = p_{ME}^{(u-1)} + \alpha p_{ME}^{(0)} \qquad u \ge 1 \tag{1.9}$$

where $p_{ME}^{(0)}$ is the initial value of the estimated limit pressure, $p_{ME}^{(u)}$ and $p_{ME}^{(u-1)}$ are the limit pressures after the uth and (u-1)th update, respectively, and α is a weighting factor, which based on our experience can be set to 1. As a result, the current pressure limit is equal to one, two, three, etc. times its initial value as R_g is subsequently attained. Obviously, the more largely the correct maximum pressure is underestimated, the more updates need to be performed for the pressure limit. This reflects on the number of iterations needed to converge to the correct solution. Different update strategies may also be used. In any case, the increment of the pressure limit should be done very carefully, because it presents similar aspects to the increase of contact stiffness within augmentations. Since the attainment of a normal pressure below the cut-off limit coincides with the introduction of the geometric term into the tangent stiffness, a too rapid increase of the cut-off may result in a too early shift into phase two.

In the second procedure, whenever R_g reaches the specified threshold for a contact element, augmentation is performed for that element, i.e. the current contact force is introduced into the system as an "external force". In symbols

$$p_N^{(a)} = p_N^{(a-1)} + \beta \varepsilon_s g_N \qquad a \geq 1 \tag{1.10}$$

where $p_N^{(a)}$ and $p_N^{(a-1)}$ are the normal pressures after the ath and (a-1)th augmentation, respectively, and β is an acceleration factor. For contact problems, using a β factor larger than one has been proved unsuccessful, hence also in this case a good choice could be $\beta = 1$. Whenever R_g reaches again the threshold in subsequent iterations, the value of the augmented force is updated. The resulting algorithm is indicated as LP-AU in the following. With respect to the LP-IP method, the LP-AU one has the additional advantage to reduce the penetration error inherent to the penalty method, due to the introduction of an augmentation scheme. The more largely the correct maximum pressure is underestimated, the more updates are performed for the augmented forces, the smaller is the norm of the normal penetration at convergence. Therefore, this method improves the quality of the solution, in terms of enforcement of the impenetrability condition. Correspondingly, the number of iterations to convergence also slightly increases.

In the implemented procedures, the computation of R_g according to eq. (1.8) is made for each active contact element. However, as soon as R_g for a given contact element is less than the specified threshold, the pressure limit is increased for all active contact elements (LP-IP procedure) or augmentation is performed for all contact elements (LP-AU).

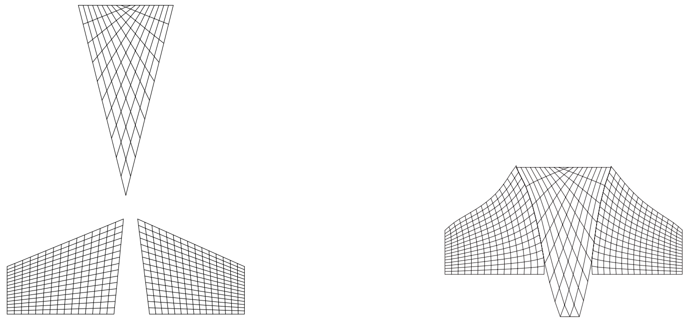
1.5 Example

The example deals with a wedge indented between two deformable bodies, see Figure 1.2a. The lower bodies, made of the softer material ($E = 2500$, $\nu = 0.25$), are restrained on the bottom and on the external sides. The wedge, made of the stiffer material ($E = 25000$, $\nu = 0.25$), is subjected to an imposed vertical displacement on its top surface. The lateral surfaces of the cantilever are the slave ones, and the inclined surfaces of the indentor are the master ones. The contact penalty parameter is $\varepsilon = 10^5$. The continuum is discretized with 4-node large deformation elastic plane-strain elements. All computations have been performed with the finite element program FEAP [2, 3].

Using standard consistent linearization (CL), the maximum vertical displacement of the wedge for which convergence is achieved is 3.0. Conversely, with the LP algorithm, convergence is reached for imposed displacements up to 7.5. The final, converged geometry for the latter value of imposed displacement is shown in Figure 1.2b. The maximum value of the contact pressure is equal to 390.

Figure 1.3 illustrates the deformed shape after the first four iterations, when CL is used. It is shown that the geometry evolves soon towards a highly distorted configuration, until catastrophic divergence occurs. Conversely, Figure 1.4 shows the analogous deformed shapes when the LP algorithm is used. It is evident that the LP strategy prevents unrealistic values of the contact pressure to be attained. As a result, the geometry rapidly approaches the deformed configuration at convergence.

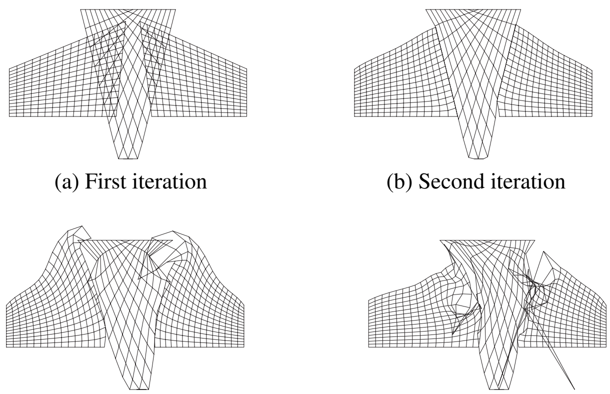
In Figure 1.5, the residual norm and the number of contact elements where the limit on the contact pressure is enforced are plotted versus the iteration number. The limitation on the contact pressure takes place initially for the whole active contact area (30 elements). At each subsequent iteration, the contact pressure becomes less than the cut-off value for an increasing number of slave nodes. Hence, such nodes enter phase two and for them CL is activated. For the remaining contact elements, the pressure limit is enforced and the modified tangent stiffness and residual are constructed. At this stage, the residual norm decreases but the rate of convergence is less than quadratic. After the first 8 iterations, all contact elements have entered phase two. From this iteration, CL is used for all contact elements and a quadratic rate of convergence is achieved. With 3 more iterations, convergence takes place. The results shown in Figures 1.4 and 1.5 have been obtained for $r = 2.5$, i.e. with the estimated maximum contact pressure equal to two and one half times the correct value. However, results are qualitatively similar if other estimates are made, provided that $r > 1$, i.e. that the correct maximum contact pressure is overestimated.



(a) Initial geometry

(b) Geometry at convergence

Fig. 1.2: Example geometry.



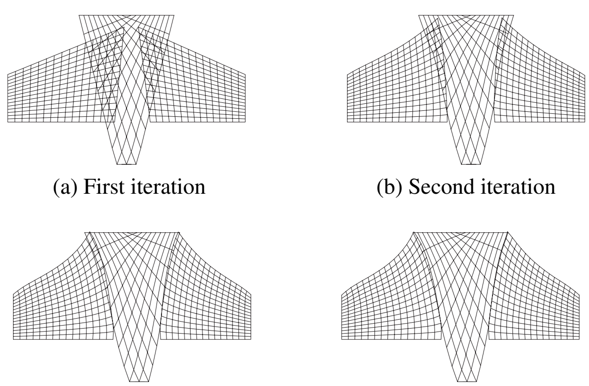
(a) First iteration

(b) Second iteration

(c) Third iteration

(d) Fourth iteration

Fig. 1.3: Example: deformed shape after the first four iterations - CL.



(a) First iteration

(b) Second iteration

(c) Third iteration

(d) Fourth iteration

Fig. 1.4: Example: deformed shape after the first four iterations - LP with $r = 2.5$.

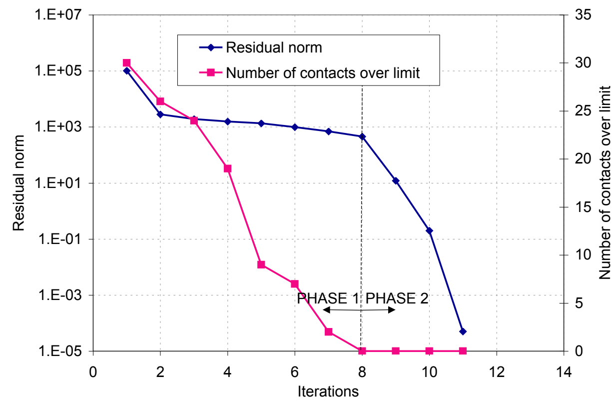


Fig. 1.5: Example: residual norm and number of contacts over limit - LP with $r = 2.5$.

1.6 Conclusions

As a follow-up to a previous investigation [5], this paper has proposed a strategy to deal with contact problems involving large penetrations, in the context of the NTS formulation and of the penalty method.

The strategy has shown a very good capability to deal with contact problems where the standard application of Newton's method does not achieve convergence. The employment of an upper bound for the maximum contact pressure, coupled with the use of a secant contact stiffness contribution, permits to enforce gradually the violated impenetrability condition within the iterations.

The proposed strategy performs efficiently if the estimated maximum contact pressure is reasonably in excess of the correct value. To deal with cases where the correct maximum contact pressure is underestimated, two alternative procedures have been devised. These procedures integrate the basic LP strategy enabling an automatic increase of the contact pressure over the initial limit. While both procedures have been shown to perform satisfactorily, the LP-AU one has the additional advantage to improve the quality of the solution, in terms of the degree of approximation in the enforcement of the impenetrability condition with the penalty method.

Finally, the proposed strategy can not only be used to achieve convergence in cases where standard CL would produce catastrophic divergence, but also to accelerate convergence for moderate values of initial penetrations. In these cases, when using a standard CL approach the first few iterations are typically needed to stabilize the solution before a quadratic rate of convergence is obtained. Using the LP strategy usually shortens this stabilization phase, thereby reducing the number of iterations and ultimately enhancing the computational efficiency.

References

[1] K.A. Fischer, P. Wriggers, “Frictionless 2D contact formulations for finite deformations based on the mortar method”, Computational Mechanics, 36: 226-244, 2005.

[2] R.L. Taylor, “FEAP - A Finite Element Analysis Program, User Manual”, University of California, Berkeley, http://www.ce.berkeley.edu/projects/feap.

[3] R.L. Taylor, G. Zavarise, “FEAP - A Finite Element Analysis Program, Contact Manual”, University of California, Berkeley, http://www.ce.berkeley.edu/projects/feap.

[4] G. Zavarise, L. De Lorenzis, R.L. Taylor, “A non-consistent start-up procedure for contact problems with large load-steps”, submitted.

[5] G. Zavarise, R.L. Taylor, “A force control method for contact problems with large penetrations”, Proceedings of the V International Conference on Computational Plasticity, D.R.J. Owen, E. Oñate, E. Hinton Eds., Pineridge Press, 280-285, Barcelona, Spain, 1997.

[6] G. Zavarise, P. Wriggers, U. Nackenhorst, A Guide for Engineers to Computational Contact Mechanics, The TCN Series on Simulation Based Engineering and Sciences, Vol. 1, Consorzio TCN, Trento, Italy, 2006. ISBN–10 88-95176-00-6, ISBN-13 978-88-95176-00-0.