Structural Engineering and Mechanics, Vol. 24, No. 2 (2006) 155-180
DOI: http://dx.doi.org/10.12989/sem.2006.24.2.155

Optimal shape design of contact systems

F. F. Mahmoud†, A. G. El-Shafei‡ and M. M. Al-Saeed‡†

Department of Mechanical Engineering, Zagazig University, Zagazig, 44511, Egypt

(Received December 8, 2004, Accepted May 16, 2006)

Abstract. Many applications in mechanical design involve elastic bodies coming into contact under the action of the applied load. The distribution of the contact pressure throughout the contact interface plays an important role in the performance of the contact system. In many applications, it is desirable to minimize the maximum contact pressure or to have an approximately uniform contact pressure distribution. Such requirements can be attained through a proper design of the initial surfaces of the contacting bodies. This problem involves a combination of two disciplines, contact mechanics and shape optimization. Therefore, the objective of the present paper is to develop an integrated procedure capable of evaluating the optimal shape of contacting bodies. The adaptive incremental convex programming method is adopted to solve the contact problem, while the augmented Lagrange multiplier method is used to control the shape optimization procedure. Further, to accommodate the manufacturing requirements, surface parameterization is considered. The proposed procedure is applied to a couple of problems, with different geometry and boundary conditions, to demonstrate the efficiency and versatility of the proposed procedure.

Keywords: contact mechanics; shape optimization; mathematical programming; surface parameterization; finite element method.

1. Introduction

Contact of elastic bodies is encountered in many engineering applications. When two or more elastic bodies are pressed together, the contact stresses throughout the contact interface will build up. Contact deformations and stresses depend as much on geometric profiles of the contacting bodies as on the externally applied load (Dundurs 1975). The resulting localized high contact stresses have severe effects on the tribological aspects of the contacting surfaces and consequently on the performance of the mechanical system. Therefore, it is desirable to reduce the peak of contact stress or to have an almost uniform contact pressure distribution at the contact interface. To obtain the desired contact stress distribution, the initial contour or shape of the contacting bodies should be properly designed. Consequently, an optimal shape design problem of the contact profiles will be considered. This problem involves the combination of two disciplines, contact mechanics and shape optimization.

Generally, contact problems of elastic bodies have been investigated analytically (Johnson 1985).

† Professor, Corresponding author, E-mail: faheem@aucegypt.edu
‡ Associate Professor
‡† Research Assistant

The analytical methods gave an insight into the problem. However, they required considerable mathematical effort, and are limited to simple geometry and boundary conditions in addition to suffering the lack of generality. For these shortcomings and limitations, computational models have been developed to solve the contact problems. Computational models are classified into two categories, primary and dual methods. The primary methods, based on the strong formulation, exploit either incremental or iterative procedure in several alternative ways to satisfy contact constraints (Francavilla and Zeinkiewicz 1975, Mahmoud *et al.* 1986). On the other hand, the dual methods, based on a weak variational formulation, make use the mathematical programming methods to solve the inequality constrained problems. The contact conditions throughout the contact interface are represented by a set of unilateral inequality constraints (Campos *et al.* 1982, Zhong and Sun 1989). Several types of contact problems, with different levels of complexity, could be effectively analyzed (Mahmoud *et al.* 1998, El-Shafei and Mahmoud 1999, Barboza *et al.* 2002, El-Shafei 2004).

An extensive survey of contact pressure optimization problems is presented by Hilding *et al.* (1999). Haug and Kwak (1978) have developed a model to minimize the peak contact pressure subject to a set of constraints representing the geometrical bounds of modification of the surface profiles. The simplex-modified Wolf's method, with additional slack and artificial variables, is adopted. The model is based on the generation of the flexibility matrices, which required cumbersome calculations. The interior penalty method is used by Mahmoud *et al.* (1989) to design optimal shape of contact surface subject to a set of geometrical bounds and limit stress constraints. The resulting optimal profile suffered from the non-regularity, which are not easily consistent with the manufacturing requirements. Based on Herskovits's interior point technique, contact shape optimization problem of hyper-elastic plane solids is investigated by Herskovits *et al.* (1998). Shape optimization of contact problems with Coulomb's friction is investigated by Beremlijski *et al.* (2002). Utilizing the meshfree method (Kim *et al.* 2003), continuum-based shape design sensitivity analysis and optimization methods are developed. Contact shape optimization of infinitesimal elasto-plasticity with frictional contact is analyzed by Kim *et al.* (2000). This analysis is extended to consider contact problems of hyper-elastic multibody (Kim *et al.* 2001). Shape optimization of a nonlinear shell structure with frictionless contact is also tackled by Kyung *et al.* (2003). Further, contact optimization problems including the wear process is investigated by Paczelt and Mroz (2004).

The objective of the present paper is to develop an integrated design procedure capable of determining the optimal shape design of elastic bodies in frictionless contact. For each design configuration, the adaptive incremental convex programming method is adopted to solve its corresponding contact problem and consequently obtaining the contact stress distribution. The augmented Lagrange multiplier method is used to investigate the shape optimization problem. Furthermore, to obtain a regular surface profile, satisfying the manufacturing requirements, a surface parameterization technique based on a cubic spline function is adopted.

2. Statement of the optimal shape design of contact system

Due to localized high contact stresses at any contact system as shown in Fig. 1, the contacting bodies may be failed. Contact stresses are highly sensitive to the shape of the contact profile. Variation of the contact profile could enable to substantially influence the value of the maximum

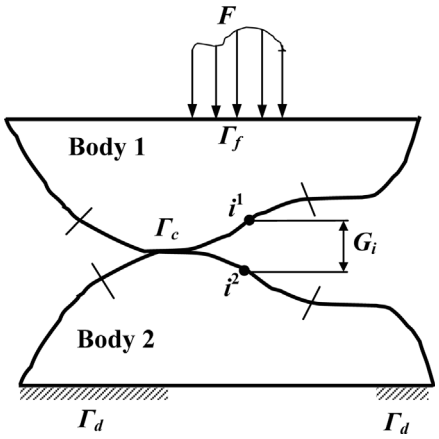

Fig. 1 Contact of two deformable bodies

contact stress or the contact pressure distribution. Therefore, shape optimization is of a significant importance on the design of contact systems. Choice of the objective function and the associated set of the design variables affects the quality of the outcome of the shape optimization process. The objective function should depend on and compatible to the nature of the considered problem. For example, in the design of valves, pistons and gear pumps, the main objective is to prevent fluid leakage and to reduce the wear of the sliding surfaces. Providing uniform contact pressure distribution throughout the contact regions can attain those required conditions. On the other hand, in the impact-hammer, deep drawing and stamping design problems, it is desirable to increase the fatigue life. This can be achieved by reducing the peak contact stress.

Now, consider two elastic bodies, shown in Fig. 1, pressed together under an external static load, such that they come into frictionless contact. The boundary Γ of each body is assumed to consist of three disjoint parts, Γ_f, Γ_d and Γ_c. Γ_f and Γ_d are the portions of the boundary on which the traction and displacement are prescribed respectively. Γ_c is a portion of the boundary that contains the adjacent surfaces which may come into contact upon the application of loads. It is obvious to mention that the boundary Γ_c may contain either nonconformal and conformal contacts regions.

Generally, the objective function to be minimized is an implicit function of the design vector $\tilde{x}$ but explicitly depends on the state vector $\tilde{u}$. Hence, the optimal shape design of a contact system can be formulated as follows:

$$\begin{aligned} &\text{Minimize} \quad f(\tilde{u}(\tilde{x})) \\ &\text{subject to} \quad \begin{cases} \sigma^e(\tilde{u}(\tilde{x})) \le \overline{\sigma} & (\text{in } \Omega) \\ \tilde{x}_l \le \tilde{x} \le \tilde{x}_u & (\text{on } \Gamma_c) \end{cases} \end{aligned} \tag{1}$$

where f is the objective function to be minimized, $\tilde{x}_l$ and $\tilde{x}_u$ are the lower and upper vectors bounds of the design variables, σ^e is the equivalent stress that can be calculated according to any adopted yield criterion, $\overline{\sigma}$ is the allowable stress limit, and Ω is the connected domain of the contacting bodies. In addition, $\tilde{u}(\tilde{x})$ is the state vector that obtained from the solution of the contact problem.

The contact problem is a class of variational inequality problems. The contact area, and consequently the kinematic boundary conditions along that area, is not known a priori. Further, The contact states depend basically on the capacity of applied loads, geometry and relative material compliance of the contacting bodies. Therefore, during the monotonically application of loads, some few boundary conditions throughout the contact interface may be relaxed and other would be added. Accordingly, such type of problems is a nonlinear one having inequality type of constraints.

Let the space domain of the contacting bodies be discretized into a finite element grid. For the discretized linear elastic system subjected to an external force vector $\tilde{F}$ and having a displacement vector $\tilde{u}$, the variational inequality model represents the static equilibrium principle of the contact system can be expressed as:

$$\text{Minimize} \quad \pi(\tilde{u}) = \frac{1}{2}\tilde{u}^T K\tilde{u} - \tilde{u}^T\tilde{F} \qquad (\text{in } \Omega)$$

$$\text{subject to} \quad \begin{cases} u_{i^1}^n - u_{i^2}^n \le G_i \equiv C_i^T\tilde{u} \le G_i \\ \sigma_i^n \le 0 \\ \sigma_i^t = 0 \end{cases} , \quad i = 1, 2, \ldots, NC \ (\text{on } \Gamma_c) \qquad (2)$$

where $\pi(\tilde{u})$ is the objective function representing the total potential energy of the system, K is the overall stiffness matrix and NC is the number of contact pairs, active constraints, throughout the contact interface Γ_c. For a contact pair i, i^1 and i^2 are the contact nodes belonging to the first and second body, respectively, $u_{i^1}^n$ and $u_{i^2}^n$ are the normal displacements of nodes i^1 and i^2 respectively, G_i is the gap, σ_i^n and σ_i^t are the normal and tangential stress components respectively, and C_i is an operator matrix on the state vector $\tilde{u}$, which represents the displacement of the system. It is obvious to mention that the state vector is a function of the vector of design variables $\tilde{x}$.

The first inequality constraint of (2), represents the kinematic constraints where the two contacting surfaces can not be interpenetrated, while the second represents the kinetic one, which state that no tensile stresses along the contact interface, and the last condition states that the contact is frictionless.

The foregoing presentation of the problem indicates that the optimal shape design of contact system consists of two simultaneous models. The first model presented by (1), concerns mainly with the shape optimization problem. This model is a nonlinear mathematical model, since both the objective and constraints functions are nonlinear. The second model defined by (2) concerns with the solution of the contact problem for any configuration proposed by the first model. This model is also a nonlinear convex one, since both the quadratic objective function and the linear contact constraints are convex. To compute the objective and constraints functions of the first model, corresponding to a specific design vector, the vector of the state variables should firstly be determined from the second model. In other words, for each vector of the design variables, selected through the optimization process, the corresponding vector of the state variables is firstly determined by solving the corresponding contact problem. For a selected configuration, the adaptive incremental convex programming method (Mahmoud *et al.* 1993, Hassan and Mahmoud 2002) is used to solve the contact problem and obtain the contact stress distribution. On the other hand, the augmented Lagrange multiplier method (Fletcher 1987) is used to solve the shape optimization problem.

3. Solution of the shape optimization problem

The main problem in the structural optimization is to determine a set of design variables such that an objective function, subject to a set of constraints, is minimized. Both the objective and constraints are usually nonlinear functions. Solution of this problem is oftenly attempted by using the nonlinear programming techniques; among them the multiplier methods are the most effective ones. The basic idea of the multiplier methods is to transform the original constrained optimization model into a sequence of unconstrained ones. The function of each unconstrained model is constructed, using both the objective and constraint functions of the original model in addition to a set of multipliers. The solutions of the sequentially unconstrained problems converge monotonically towards the solution of the original constrained problem.

The general form of the constrained optimization problem (Belegundu and Arora 1985), is given by:

$$\begin{aligned} &\text{Minimize} \quad f(\tilde{x}), \quad \tilde{x} \in R^n \\ &\text{subject to} \begin{cases} h_i(\tilde{x}) = 0, & i = 1, 2, \ldots, l \\ g_i(\tilde{x}) \le 0, & i = l+1, l+2, \ldots, m \end{cases} \end{aligned} \tag{3}$$

where $f(\tilde{x})$ is the objective function, $\tilde{x}$ is a multi-valued vector of the design variables in a space R^n of n dimensions, $h(\tilde{x})$ and $g(\tilde{x})$ are the equality and inequality constraints functions respectively, and l is the number of equality constraints, while m is the total number of constraints.

The optimization model (3) is quite general and can be used to model various engineering optimization problems (Haug and Arora 1979, Arora 1989, 1990). One of the crucial difficulties in the engineering applications is that both the objective and constraints functions are oftenly nonsmooth and depending implicitly on the design variables. This makes the evaluation of their gradient truly tedious and expensive. The multiplier methods can easily alleviate this difficulty, as they require only the gradient of one augmented function. This gradient can be evaluated efficiently using the adjoint variable method (Arora and Haug 1979) without computing gradients of the individual constraints.

The Lagrangian function equivalent to the original model, defined by (3), can be presented as:

$$L(\tilde{x}, \tilde{\mu}) = f(\tilde{x}) + \sum_{i=1}^{l} \mu_i h_i(\tilde{x}) + \sum_{i=l+1}^{m} \mu_i g_i(\tilde{x}) \tag{4}$$

in which $\tilde{\mu}$ is the vector of Lagrange multipliers. The multiplier methods employ an augmented Lagrangian function in which some penalty terms involving constraints are added to the ordinary Lagrangian form (4). Therefore, the Lagrange multipliers and certain penalty parameter, for each constraint, are used to construct an equivalent function, that can be generally written as:

$$\Phi(\tilde{x}, \tilde{\mu}, \tilde{r}) = f(\tilde{x}) + P(h(\tilde{x}), g(\tilde{x}), \tilde{\mu}, \tilde{r}) \tag{5}$$

in which $P(h(\tilde{x}), g(\tilde{x}), \tilde{\mu}, \tilde{r})$ is a generalized penalty function and $\tilde{r}$ is the vector of penalty parameters. The values of the vectors of Lagrange multiplier and penalty parameter are chosen at the beginning of each unconstrained minimization and then the augmented function $\Phi(\tilde{x}, \tilde{\mu}, \tilde{r})$ is

minimized with respect to $\tilde{x}$. At the end of each minimization step, both the Lagrange multiplier and penalty parameter are updated and the process repeated until the satisfaction of the convergence criteria. The basic motivation for the augmented multiplier methods is to avoid the ill-conditioning associated with the usual penalty function methods (Fletcher 1975). In contrast to the penalty function methods, the vector of penalty parameter $\tilde{r}$ need not go to infinity to achieve convergence of the multiplier methods. As a consequence, the augmented Lagrangian function has a good condition.

The augmented Lagrangian function can be defined in several ways. The most popular and commonly form, includes quadratic penalty terms, taking the following form (Fletcher 1975):

$$\Phi(\tilde{x}, \tilde{\theta}, \tilde{r}) = f(\tilde{x}) + \frac{1}{2}\sum_{i=1}^{l} r_i[(h_i + \theta_i)^2 - \theta_i^2] + \frac{1}{2}\sum_{i=l+1}^{m} r_i[(g_i + \theta_i)_+^2 - \theta_i^2] \tag{6}$$

in which $r_i\theta_i(\equiv \mu_i)$ are the Lagrange multipliers and $(f)_+ = \max(0, f)$. This function has a suitable structure for so, it is simpler to work with the Lagrange multipliers μ_i's instead of θ's. Therefore, the augmented Lagrangian function (6) may be written in terms of μ's as follows:

$$\Phi(\tilde{x}, \tilde{\mu}, \tilde{r}) = \begin{cases} f(\tilde{x}) + \sum_{i=1}^{l}\left(\frac{1}{2}r_i h_i^2 + \mu_i h_i\right) + \sum_{i=l+1}^{m}\left(\frac{1}{2}r_i g_i^2 + \mu_i g_i\right); & \text{if } \left(g_i + \frac{\mu_i}{r_i}\right) \geq 0 \text{ for } i > l \\ f(\tilde{x}) + \sum_{i=1}^{l}\left(\frac{1}{2}r_i h_i^2 + \mu_i h_i\right) - \sum_{i=l+1}^{m}\left(\frac{\mu_i^2}{2r_i}\right); & \text{if } \left(g_i + \frac{\mu_i}{r_i}\right) < 0 \text{ for } i > l \end{cases} \tag{7}$$

It is clearly obvious that the above expression has few penalty terms are added to the ordinary Lagrangian function (4), hence it is called the augmented Lagrangian function.

3.1 Unconstrained minimization and updating procedure

The original multiplier methods require an exact solution of the unconstrained minimization model at each step. This is numerically impossible, as it may require much more iterations, and therefore, many function evaluations. However, there are various methods that work quite well with an inexact solution of the model at each step. The minimization procedure is usually terminated once the following condition is satisfied:

$$\left\|\nabla \Phi(\tilde{x}^k, \tilde{\mu}^k, \tilde{r}^k)\right\| \leq \varepsilon_k \tag{8}$$

where k is the unconstrained minimization step, ε_k is the error tolerance such that $\varepsilon_k \to 0$ as $k \to \infty$ and ∇ represents the gradient operator with respect to the vector of design variables $\tilde{x}$. One way to increase efficiency of the method is to terminate the unconstrained minimization with a crude approximation to the solution. This enables the vector of Lagrange multipliers $\tilde{\mu}$ to be updated more frequently. This is important for large scale engineering optimization problems where an exact minimization is impossible to reach. To accomplish this, it is better to terminate the unconstrained minimization when the gradient of the Lagrangian function is less than some measure of the unfeasibility of the constraint (Arora *et al.* 1991).

$$\left\|\nabla \Phi(\tilde{x}^k, \tilde{\mu}^k, \tilde{r}^k)\right\| \leq \xi\left\|g(\tilde{x}^k)\right\| \tag{9}$$

where ξ is some fixed positive parameter and the norm of the right hand side is only includes the violated constraints subset. Further it is suggested that the parameter ξ may be reduced in certain situations (Coope and Fletcher 1980). Therefore, $\xi_k \to 0$ and $\varepsilon_k \to 0$ as $k \to \infty$. In addition to the above termination criteria, a limit on the number of unconstrained minimization iterations may be imposed. Computational experience indicates that a considerable saving is realized by accepting an inexact unconstrained minimization.

A simple procedure for updating the Lagrange multiplier is presented in Powell (1969) such that for the *i*th equality constraint

$$\theta_i^{k+1} = \theta_i^k + h_i(\tilde{x}^k), \quad i = 1, 2, \ldots, l \tag{10}$$

where θ_i^k define the Lagrange multipliers as $\lambda_i^k = r_i \theta_i^k$ and k denotes the unconstrained minimization stage. It is important to note that the last equation does not require the gradient of the individual constraints, and consequently it is more suitable for large-scale engineering problems. On the other hand, when inequality constraints are existing, the Lagrange multipliers can be updated using the constraint function as:

$$\theta_i^{k+1} = \theta_i^k + \max(g_i(\tilde{x}^k), -\theta_i^k), \qquad i = l+1, l+2, \ldots, m \tag{11}$$

It has been shown that the updating procedures defined by (10) and (11) are merely the steepest ascent methods for an equivalent dual of the original optimization model (3). Convergence properties of the multiplier methods using various updating procedures have been analyzed and more sophisticated updating procedures to obtain faster convergence rate have been addressed in Arora *et al.* (1991). However, those procedures require the gradients of individual constraints, making them unsuitable for engineering applications.

3.2 Convergence conditions

The global convergence of the multiplier methods can be analyzed in the framework of the general methods of penalty function (Berisekas 1976). Considering a non-negative parameter $\overline{K}$ that used to enforce the global convergence. This non-negative parameter, represents the maximum constraint violation, can be expressed as:

$$\overline{K} = \max(V_E, V_I) \tag{12}$$

where

$$V_E = \max_{1 \le i \le l} \{|h_i|\} \text{ and } V_I = \max_{l+1 \le i \le m} \{|g_i - \theta_i|\} \tag{13}$$

To enforce the global convergence, $\overline{K}$ is required to be reduced at every iteration of the unconstrained minimization algorithm, and the entire iterative process is stopped if the following criteria are satisfied:

$$\overline{K} \le \varepsilon \text{ and } \|\nabla \Phi(x^k, \theta^k, r^k)\| \le \varepsilon \tag{14}$$

where ε is an error tolerance. If $\overline{K}$ is not reduced at a particular iteration, then the value of the penalty parameter r is increased by any automatic procedure (Haug and Arora 1979). Therefore the role of the penalty parameter is important for obtaining global convergence of the multiplier methods. According to (Berisekas 1976), the detailed algorithm of the augmented Lagrange multiplier procedure is presented in Appendix-A.

3.3 The unconstrained minimization algorithm

In the augmented Lagrange multiplier algorithm presented in Appendix-A, it is required to minimize the unconstrained augmented function $\Phi(\tilde{x}^{k-1}, \tilde{\theta}^{k-1}, \tilde{r}^{k-1})$. The modified conjugate directions method (Arora 1989), is adopted to minimize that function. This method builds up information that approximates the second derivatives of the considered function to be minimized. Now, after receiving the $\tilde{x}^{k-1}$ as the current estimate of the minimum point, as presented in the third step of the algorithm, presented in Appendix-A, the procedure to minimize the unconstrained augmented function is also presented in Appendix-B.

4. Solution of the contact problem as a convex programming model

According to the incremental convex programming method (Mahmoud *et al.* 1993, Hassan and Mahmoud 2002), which is one of the active multipliers methods, the frictionless contact model, defined by (2), is replaced by a sequence of M submodels, each of them is depicted as follows:

$$\begin{aligned} &\underset{\tilde{u}}{\text{Minimize}} \quad \pi_m(\tilde{u}) = \frac{1}{2}\tilde{u}_m^T K \tilde{u}_m - \tilde{u}_m^T \tilde{F}_m \\ &\text{subject to} \quad \tilde{g}(\tilde{u}) = C^T \tilde{u}_m \le \tilde{G}_m \\ &\text{such that} \quad \tilde{F} = \sum_{m=1}^{M} \tilde{F}_m \text{ and } \tilde{G} = \sum_{m=1}^{M} \tilde{G}_m \end{aligned} \tag{15}$$

where m is the submodel number and M is the total number of models.

Since both the objective function $\pi(\tilde{u})$ and the constraints $\tilde{g}(\tilde{u})$ are convex, any local minimum is also a global one (Haug and Arora 1979). Let $(\tilde{u}_m^*, \tilde{\lambda}_m^*)$ be the global minimum point of the convex submodel (15), the global minimum point $(\tilde{u}^*, \tilde{\lambda}^*)$ of the original model (2) can be depicted as

$$\tilde{u}^* = \sum_{m=1}^{M} \tilde{u}_m^* \text{ and } \tilde{\lambda}^* = \sum_{m=1}^{M} \tilde{\lambda}_m^* \tag{16}$$

where $\tilde{\lambda}$ is the non-negative multiplier vector, which physically represents the contact forces vector at the contact region. The global minimum point of each submodel should satisfy the following Kuhn-Tucker necessary and sufficient conditions:

$$K\tilde{u}_m^* + C\tilde{\lambda}_m^* = \tilde{F}, \qquad C_m^T \tilde{u}_m^T \le \tilde{G}_m, \qquad \tilde{\lambda}_m^{*T}(C_m^T \tilde{u}_m^* - \tilde{G}_m) = 0 \text{ and } \tilde{\lambda}_m^* \ge 0 \tag{17}$$

which represents the dual form of the problem.

4.1 Solution procedure

The algorithm of solution of the contact problem presented above is based on an adaptive incremental procedure. The maximum number of increments, and consequently the number of submodels, does not exceed the total number of constraints. Through each submodel, only one new contact event, either contact or separation, is detected. Therefore, the algorithm determines the active and inactive constraints, corresponding to each contact or separation event in an adaptive incremental scheme. In addition, the Lagrange multipliers of the active constraints, which physically represent the contact forces, are calculated. Therefore, at no penalty, the contact events versus load history can be obtained.

Before going to the detailed algorithm of solution, three points should be emphasized;

(i) The constraints of each submodel m may be divided into two sets. The first one consists of all inactive constraints and is designated by Φ_m, and the other one contains only all active constraints and is designated by Ψ_m.

(ii) To identify the components of the submodel number m, vectors $\tilde{F}_m$ and $\tilde{G}_m$ are chosen in such a manner that either one of the inactive constraints of the submodel $m-1$, would change to be an active one, or one of the active constraints would switch to be an inactive one.

(iii) The identification of the inactive constraint, of the submodel $m-1$, candidate to change to be an active one in the submodel m depends on the determination of a violation vector;

$$\tilde{\xi}_m = C_m^T \tilde{u}_m - \tilde{G}_{m-1} \tag{18}$$

Similarly, the identification of the active constraint, of the submodel $m-1$, candidate to switch to be an inactive one in the submodel m depends on the determination of a violation vector;

$$\tilde{\eta}_m = \tilde{\lambda}_m + \sum_{k=1}^{m-1} \tilde{\lambda}_k^* \tag{19}$$

The role of both violation vectors $\tilde{\xi}_m$ and $\tilde{\eta}_m$ will be explained in the following algorithm:

4.2 The algorithm of solution

Assume the solution of the submodel $m-1$ is $(\tilde{u}_{m-1}^*, \tilde{\lambda}_{m-1}^*)$, the gap vector is $\tilde{G}_{m-1}$ and the residual load for the next submodel is $\tilde{F}_m$. The algorithm of solution to obtain the contact status for the mth submodel is depicted as follows:

1. Based on the both inactive and active constraints sets Φ_{m-1} and Ψ_{m-1} respectively, construct the operator matrix C_m.
2. According to (17), construct the equilibrium equations such that

$$\begin{pmatrix} K & C \\ C^T & 0 \end{pmatrix}_m \begin{Bmatrix} \tilde{u} \\ \tilde{\lambda} \end{Bmatrix}_m = \begin{Bmatrix} \tilde{F} \\ \tilde{G} \end{Bmatrix}_m \tag{20}$$

Once the equilibrium Eq. (20) is solved, both the incremental state and Lagrange multipliers vectors $\tilde{u}_m$ and $\tilde{\lambda}_m$ respectively are determined.

3. Detecting a new contact pair:
For the inactive constraints set Φ_{m-1}, and according to (18), the ith component of the violation vector ξ_m^i can be stated as

$$\xi_m^i = C_m^T \Delta\tilde{u}_m - G_{m-1}^i, \quad i = 1, 2, \ldots, I \tag{21}$$

where I is the number of the inactive constraints.
We can notice from (21) that if ξ_m^i is negative or positive, the inactive constraint number i is still inactive or violated, respectively. Among the set of constrains associated with positive violation values, detect a constraint, designated as $C1$, that having the maximum violation value, such that

$$C1 \in \phi_{m-1}: \quad \xi_m^{C1} = \underset{i=1}{\overset{I}{\text{Max}}} \xi_m^i, \quad \forall \xi_m^i > 0 \tag{22}$$

The inactive constraint $C1$ is candidate to be an active one. The scale factor required to establish this new active constraint is:

$$\alpha_{C1} = \frac{G_{m-1}^{C1}}{C_m^T \tilde{u}_m} \tag{23}$$

4. Detecting a new separation pair:
For the active constraints set Ψ_{m-1}, and according to (19), the jth component of the violation vector η_m^j can be stated as

$$\eta_m^j = \lambda_m^j + \sum_{k=1}^{m-1} \lambda_k^{*j}, \quad \forall \lambda_m^j < 0 \tag{24}$$

where J is the number of active constraints.
We can notice from (24) that if η_m^j is positive or negative, the active constraint number j is still active or violated, respectively. Among the set of constraints associated with negative violation values, we can detect the constraint, designated as $C2$, that having the maximum absolute violation value, such that

$$C2 \in \psi_{m-1}: \quad \eta_m^{C2} = \underset{j=1}{\overset{J}{\text{Max}}} \left|\eta_m^j\right|, \quad \forall \eta_m^j < 0 \tag{25}$$

The active constraint $C2$ is candidate to be an inactive one. The scale factor required to establish this new active constraint is:

$$\alpha_{C2} = \frac{\left|\sum_{k-1}^{m-1} \lambda_k^{*C2}\right|}{\left|\lambda_m^{C2}\right|} \tag{26}$$

5. Choose the scale factor of the mth submodel, such that only contact or separation event will occur

$$\alpha_C = \text{Min}\ (\alpha_{C1}\ ,\ \alpha_{C2}) \tag{27}$$

6. Now, based on the new event corresponding to the minimum scale factor, update the cardinality of the two constraints sets. If α_{C1} is the minimum scale factor, new contact event is encountered, the cardinality of the inactive constraints set Φ is reduced by one, where one of its constraints turns into active and joins the active constraints set Ψ. On the other hand, when the α_{C2} is the minimum, new separation event is encountered, the cardinality of the active constraints set Ψ is reduced by one, where one of its constraints turns into inactive and joins the inactive constraints set Φ.
7. Update the displacement, contact force and gap vectors for the mth submodel, such that

$$\tilde{u}_m^* \leftarrow \alpha_C \tilde{u}_m, \quad \tilde{\lambda}_m^* \leftarrow \alpha_C \tilde{\lambda}_m \quad \text{and} \quad \tilde{G}_m = \tilde{G}_{m-1} - C_m^T \alpha_C \tilde{u}_m$$

8. According to (16), update the global solution of the problem, such that

$$\tilde{u}^* = \sum_{k=1}^{m} u_k^* \quad \text{and} \quad \tilde{\lambda}^* = \sum_{k=1}^{m} \tilde{\lambda}_k^* \tag{29}$$

9. Calculate the residual force vector to be applied for the next model

$$\tilde{F}_{m+1} = \tilde{F}_m (1 - \alpha_C) \tag{30}$$

10. $m = m + 1$, initiate the next model

Repeat steps 2 to 10 until the intensity of the load vector $\tilde{P}_{m+1}$ becomes zero.

5. Shape optimization of elastic bodies in contact

The foregoing formulations of the shape optimization and contact problems can be joined together and implemented into an integrated procedure capable of determining the optimal shape of the contacting bodies. This procedure consists of two simultaneous models. The first one is a nonlinear mathematical model concerned with the shape optimization problem. In this model, both the objective and constraints functions are nonlinear and implicitly depending on the design variables. To compute the objective and constraints functions of that model, the state variables should be firstly determined. The state variables are calculated by solving the model, concerns with the contact problem. In this model, the quadratic objective function and the linear contact constraints constitute together a convex programming problem.

In the shape optimization problem of contact systems, three important aspects should be considered. The first aspect is the choice of the objective function to be minimized. The choice of the objective function critically governs the outcome of the shape optimization process, where the objective function depends on the nature of the considered problem. In some cases, it is desirable to have a uniform contact pressure distribution throughout the contact regions, while in other situations, reducing the peak contact stress is highly desirable. The second aspect; it is important to know how the response of the contact system is sensitive for a slight changing of the shape of the contacting bodies. The optimization procedure requires the sensitivity factor in order to compute the proper search direction. Hence, shape sensitivity analysis should be considered to provide such type of information. The coordinates of the nodal points represent geometry of the contacting bodies. The change of the shape is represented by a finite variation of the design variables. Therefore, the

third aspect is concerned with the parameterization of the shape. The parameterization of the shape has two folds, the first is to relate the change of the design variables and nodal locations, while the second is to obtain a reliable surface profile satisfying the manufacturing requirements. Those three aspects will be discussed in the following sections.

5.1 Formulation of the objective functions

As mentioned before, choice of the objective function depends on the nature of the considered problem. Results of the optimization process are too sensitive to the choice of the objective function. There are several choices for the objective function to be used. Two types of the objective function are considered, the first is the minimization of the peak contact stress, while the second is to achieve an almost uniform contact pressure over the actual contact area.

When the contact area is very small compared to the gross dimensions of the contacting bodies, the contact pressure in the contact region is several times larger than elsewhere. In addition, within this contact area, a peak contact stress is encountered. Therefore, it is desirable to minimize the peak contact pressure, consequently, the objective function to be minimized will take the form:

$$f(\tilde{u}(\tilde{x})) = \text{Max}(p_i), \quad i = 1, 2, \ldots, NC \tag{31}$$

in which p_i is the normal contact stress at the ith contact node at the contact interface and NC is the number of candidate contact pairs currently in contact, i.e., the number of active constraints. There is a direct correlation between contact pressures and nodal contact forces. In fact contact pressure and nodal contact forces may be used interchangeably (Haug and Kwak 1978). The obtained results demonstrate that no significant difference is found while substituting the contact pressure by the nodal contact forces. Thus when using the maximum nodal contact force to be minimized, the objective function will take the following form:

$$f(\tilde{u}(\tilde{x})) = \text{Max}(\lambda_i), \quad i = 1, 2, \ldots, NC \tag{32}$$

where λ_i is the nodal normal contact force at the ith contact node at the contact interface.

On the other hand, it is oftenly desirable to have a uniform pressure distribution over the actual area of contact. Uniform contact pressure ensures uniform wear of the components at the mating interface. In practice, to have a uniform pressure distribution over the actual area of contact, it is suggested to minimize the standard deviation of the contact pressure distribution along contact boundaries. Thus, the objective function may be formulated as follows:

$$f(\tilde{u}(\tilde{x})) = \frac{1}{\Gamma_c}\sqrt{\int_{\Gamma_c}(p(\tilde{x}) - p_{av})^2 dx} \tag{33}$$

with

$$p_{av} = \frac{1}{\Gamma_c}\int_{\Gamma_c} p(\tilde{x})dx \tag{34}$$

where $p(\tilde{x})$ is the final contact pressure distribution over the contact interface obtained by fitting a polynomial of the pressures at the discrete contact nodes, p_{av} is the mean contact pressure and Γ_c is the contact area. If Γ_c represents the potential contact area, then the effect of the optimization

process is to spread the actual final contact profile, i.e., increase the area of contact and simultaneously make the pressure distribution uniform. On the other hand, if Γ_c represents the actual contact area, then the optimization procedure needs no effort to increase the final area of contact.

Again, the objective function defined by (33) may be written in terms of the contact forces as:

$$f(\tilde{u}(\tilde{x})) = \frac{1}{NC}\sum_{i=1}^{NC}(\lambda_i - \lambda_{av})^2 \tag{35}$$

and

$$\lambda_{av} = \frac{1}{NC}\sum_{i}^{NC}\lambda_i \tag{36}$$

We can easily noticed that the problem has a non-smooth nature, since the objective function, as stated by (31) or (33), and the strength constraint given by (1) are both implicit with respect to the design variables. However the problem is a non-smooth, but for the sake of simplicity, the derivatives of the objective and constraints functions are calculated by the forward divided difference scheme.

5.2 Sensitivity analysis of the problem

An important aspect in the shape optimization is the computation of the derivatives of the objective and constraints functions, which are nonsmooth functions. Therefore, computation of derivatives requires special care in contact problems. The purpose of shape sensitivity analysis is to evaluate how and how much the behavior of a system changes due to slight changes in the shape of the contacting bodies. The optimization procedure vitally requires this information in order to select and compute the search direction that decreases the objective function. There are two approaches to carryout the sensitivity analysis (Haug and Kwak 1978), the numerical computation by divided differences or the analytical computation by implicit differentiation of the governing system of equations. In the proposed model, the divided differences scheme is used to obtain such derivatives, since there are no formulae of both objective and constraints functions.

In the contact shape optimization problem the nodal contact force is a non-differentiable function of design variables. Therefore, the first derivative of a contact force with respect to the design variables x_i, using forward divided difference scheme, is given by:

$$\frac{d\Lambda}{dx_i} = \frac{\Lambda(x_1, x_2, \ldots\ldots, x_i + b_i, \ldots, x_n) - \Lambda(\tilde{x})}{b_i} \tag{37}$$

in which $\Lambda(x_1, x_2, \ldots\ldots, x_i + b_i, \ldots, x_n)$ represents the maximum contact force due to the new vector of design variables and b_i is the slight change of the design variable x_i.

5.3 Surface parameterization

In the optimal shape design of contact systems, the coordinates of the nodal points, along the contact interface, are considered as design variables. As the number of the design variables increases, the search process to find the optimal solution, will increase and consumes more computations, which in turn includes more oscillation and less convergence rate. The purpose of surface parameterization is to provide an approach for decreasing the number of design variables required for

the optimization process. The suggested new design variables are known as the control points or control nodes. The number of control points depends on the complexity of the surface profile and is usually less than the number of contacting nodes. The number of control points should be enough to derive the coordinates of any point belongs to the contacting surfaces using a convenient interpolation scheme. Therefore, the adopted interpolation algorithm represents an intermediate phase between the contact and optimization modules. The objective of the interpolation scheme is to receive the values of the design variables and interpolates them to generate the new shape of the contact interface. Further, this new shape should satisfy the manufacturing requirements.

General polynomial interpolation can be used as interpolation scheme. However, with higher-order polynomials there are cases where these functions can lead to erroneous results. An alternative approach is to apply lower-order polynomials to subsets of nodes. Such connecting polynomials are called spline functions. These functions have the additional property that the connections between adjacent cubic equations are visually smooth. Further, on the surface it would seem that the third-order approximation of the spline would be in lower order than the higher-order polynomial (Chapra and Canale 1989). Therefore, the cubic spline interpolation function would be used.

The main idea of the cubic spline is to derive a third-order polynomial for each interval between the control nodes, such as

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i, \quad i = 1, 2, \ldots\ldots, n \tag{38}$$

Thus, for $n + 1$ nodes there are n intervals and, consequently, $4n$ unknown constants have to evaluate. Therefore, $4n$ conditions are required to evaluate these unknowns. An efficient interpolation scheme using cubic spline function, but with $n - 1$ instead of $4n$ equations is used to represent the contact profile. According to Cheney and Kincaid (1985), the first step is based on the observation that because of each pair of control nodes is connected by a cubic function, the second derivative within each interval is a straight line. On this basis, the second derivatives of (38) can be represented by a first-order Lagrange interpolating polynomial

$$f_i''(x) = f''(x_{i-1})\frac{x - x_i}{x_{i-1} - x_i} + f''(x_i)\frac{x - x_{i-1}}{x_i - x_{i-1}} \tag{39}$$

where $f_i''(x)$ is the value of the second derivative at any point x within the ith interval. Thus, this equation is a straight line connecting the second derivative at the first control node $f_i''(x_{i-1})$ with the second derivative at the second one $f_i''(x_i)$.

Eq. (39) can be integrated twice to yield an expression for $f_i(x)$. However, this expression will contain two unknown constants of integration. These constants can be evaluated by using the function-equality conditions, $f_i(x)$ equals to $f_i(x_{i-1})$ at x_{i-1} and also $f_i(x)$ equals to $f_i(x_i)$ at x_i. By performing these evaluations, the following cubic equation results

$$\begin{aligned} f_i(x) = {} & \frac{f''(x_{i-1})}{6(x_i - x_{i-1})}(x_i - x)^3 + \frac{f''(x_i)}{6(x_i - x_{i-1})}(x - x_{i-1})^3 \\ & + \left[\frac{f(x_{i-1})}{x_i - x_{i-1}} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x_i - x) \\ & + \left[\frac{f(x_i)}{x_i - x_{i-1}} - \frac{f''(x_i)(x_i - x_{i-1})}{6}\right](x - x_{i-1}) \end{aligned} \tag{40}$$

This relationship, for the cubic spline for the *i*th interval, is a much more complicated expression than (38). However, it is noticed that it contains only two unknown coefficients, the second derivatives at the beginning and at the end of the interval $f''(x_{i-1})$ and $f''(x_i)$.

The second derivatives can be evaluated by using the condition that the first derivatives at the control nodes must be continuous, i.e.,

$$f'_{i-1}(x_{i-1}) = f'_i(x_i) \tag{41}$$

The first derivatives can be evaluated by differentiating (40). If this is done for both $(i-1)$th and *i*th intervals and the results are set equal according to (41), the following relationship results:

$$\begin{aligned}(x_i - x_{i-1})f''(x_{i-1}) + 2(x_{i+1} - x_{i-1})f''(x_i) + (x_{i+1} - x_i)f''(x_{i+1}) \\ = \frac{6}{(x_{i+1} - x_i)}[f(x_{i+1}) - f(x_i)] + \frac{6}{(x_i - x_{i-1})}[f(x_{i-1}) - f(x_i)]\end{aligned} \tag{42}$$

If (42) is written for all interior control nodes, $n-1$ simultaneous equations result with $n+1$ unknown second derivatives. However, because this is a natural cubic spline, the second derivatives at the end control points are zero and the problem reduces to $n-1$ equations with $n-1$ unknowns. In addition, it is noted that the system of equations will be tridiagonal. Thus, not only the number of equations is reduced but also they are presented in a form that is extremely easy to solve.

5.4 The integrated procedure for shape optimization of contact system

The foregoing formulations of the optimal shape design problem of contact systems indicate that two simultaneous models can present the problem out. The first one concerns mainly with the shape optimization problem, while the second model concerns with the solution of the contact problem for any configuration proposed by the first model. To compute the objective and constraints functions of the first model, corresponding to a specific design vector, the state vector should firstly be determined from the second model. For the selected configuration, the adaptive incremental convex programming model, presented in section 4, is used to solve the contact problem and obtain the contact pressure distribution. On the other hand, the augmented Lagrange multiplier method, outlined in section 3, is adopted to solve the shape optimization problem. In addition, the objective function and the surface parameterization technique are formulated in sections 5.1 and 5.3 respectively.

These formulations and the associated algorithms can be joined together and implemented into an integrated procedure capable of determining the optimal shape of contact systems. The constrained optimization model is considered the milestone of the proposed integrated procedure. Fig. 2 illustrates the flow diagram of the proposed integrated procedure.

Based on the geometry of the contact system under consideration, the initial vector of the design variables is constructed. Also, some initial parameters are initiated to start the constrained optimization algorithm. As mentioned previously, the design variables are manipulated as control nodes, located on the contact profile. Construction and size of the design vector are based on the complexity of the geometry of the contact system. Based on the considered application, both equality and inequality constraints are evaluated. From the control nodes, which representing the design vector, the initial geometry of the contact profile is obtained using the cubic spline

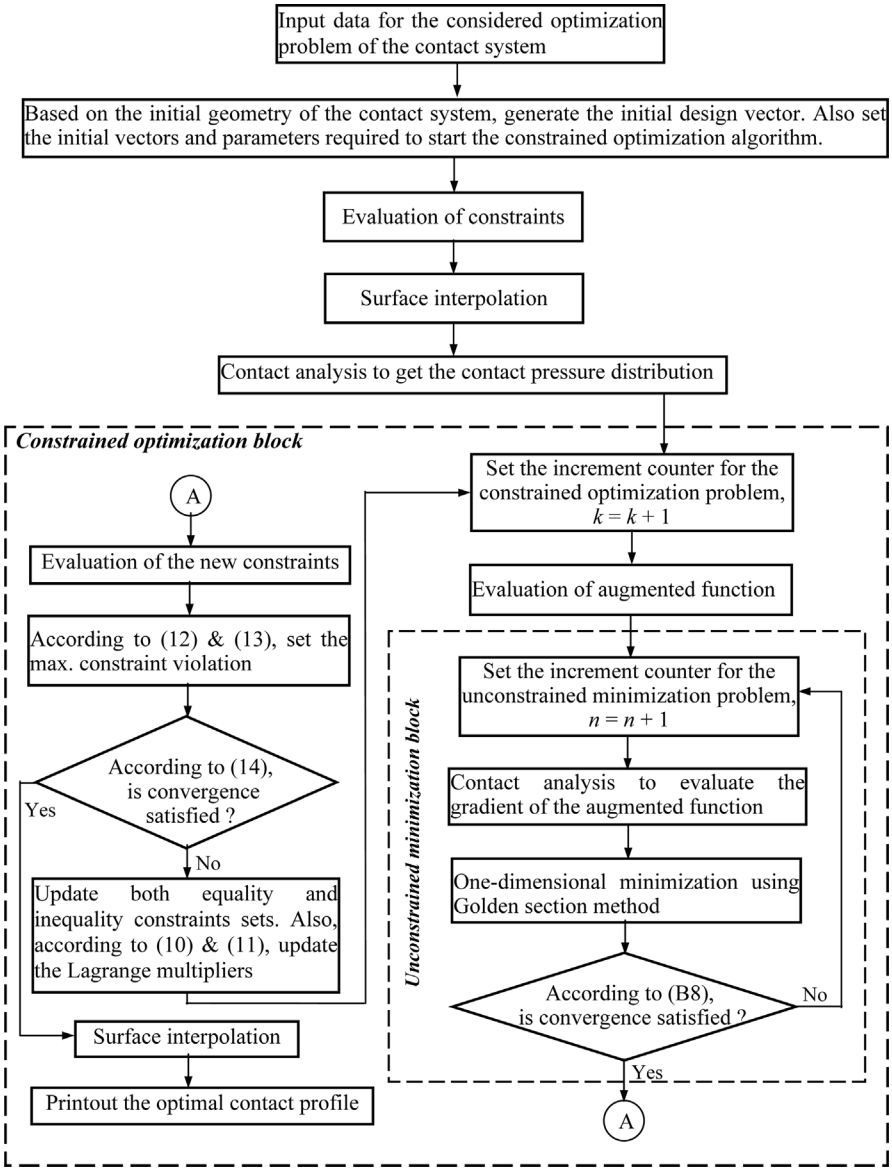

Fig. 2 Flow diagram of the proposed integrated procedure for shape optimization of contact systems

interpolation function. The initial configuration is then solved, using the incremental convex programming model, to get the contact pressure distribution. Once the contact pressure distribution is determined, the constrained optimization model, using the augmented Lagrange multiplier method to minimize the maximum contact pressure, is started.

Starting from the current configuration, the constrained optimization model is beginning by evaluating the augmented Lagrangian function. This function is minimized through the unconstrained minimization block, shown in Fig. 2, to get the best point in the current constrained optimization step k. The maximum step size, which required in the unconstrained minimization

process, is evaluated using the Golden section method. Further, the contact model is needed to evaluate the gradient of the augmented function. To terminate the unconstrained minimization process, the convergence criterion, defined by (B8) in Appendix B, should be satisfied.

Based on the best point that obtained from the unconstrained minimization process, the constrained optimization model is going to evaluate the new constraints. According to the status of the new constraints, the maximum constraint violation is determined by (12) and (13). If the convergence condition, defined by (14), is not satisfied, both equality and inequality constraints sets are modified. Accordingly, the Lagrange multiplier vector should be updated, as presented in (10) and (11), then a new optimization step is initiated. On the other hand, should the convergence be satisfied, the constrained optimization process is terminated with the optimal design vector. This design vector is interpolated again to locate the optimal contact profile. The detailed algorithms for both constrained optimization and unconstrained minimization techniques are presented in Appendices A and B respectively.

6. Results and disscusion

The developed integrated procedure for the optimal shape design problem of contact systems is implemented into a two-dimensional linear finite element model. The model is applied to a couple of different problems to demonstrate the efficiency and versatility of the proposed procedure. The considered problems are different in geometry, and boundary and loading conditions as well as the material characteristics. Furthermore, the contacting bodies are different, since elastic/rigid and elastic/elastic contact cases are studied. In addition, the initial contact profile of the contacting bodies is not restricted to a specified geometry. Therefore, the initial contact profiles having a quadratic or a flat geometry could be considered. Also, in some cases there are no limitations imposed on the final contact profile, while in other applications, the contact surface is bounded by upper and lower boundaries. Therefore, the developed procedure can be applied to any contact systems in the frame of the above assumptions.

In the design of these contact problems, the objective is to determine the contact profile that minimizing the maximum contact pressure. The analyses of these problems are based on the plane stress conditions. Further, the materials of the contacting bodies are assumed to be homogeneous, isotropic, and obey Hooke's law, i.e., linear elastic behavior is only considered.

6.1 An elastic curved plate pressed against a rigid foundation

The shape optimization problem of an elastic curved plate resting on a rigid foundation is considered to get the optimal contact profile that minimizing the peak contact pressure. The material of the plate is having a Young's modulus of 215 GPa, and Poisson's ratio of 0.29. The geometry and loading conditions are shown in Fig. 3. The initial contact profile is a quadratic curve having the relation $X_2 = R - [R^2 - (X_1 - 4)^2]^{1/2}$, where $X_1 \in [0,4]$ and $R = 160.0125$ m. The optimization process is terminated after fifteen iterations where the solution is converged to the desired accuracy.

Fig. 4 illustrates the predicted optimal contact profile. It is clear that the obtained optimal shape is dramatically different from the initial one. In addition, due to the surface parameterization, the obtained profile is almost regular satisfying the manufacturing requirements. Fig. 5 shows the variation of the contact pressure throughout the contact interface for the initial and final optimal

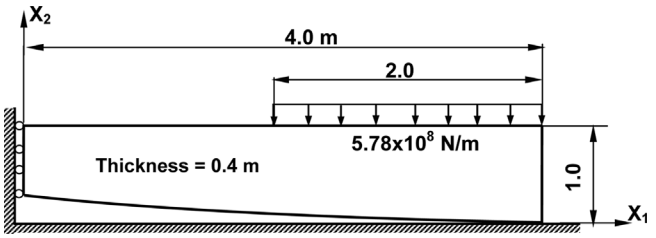

Fig. 3 An elastic curved plate pressed against a rigid foundation

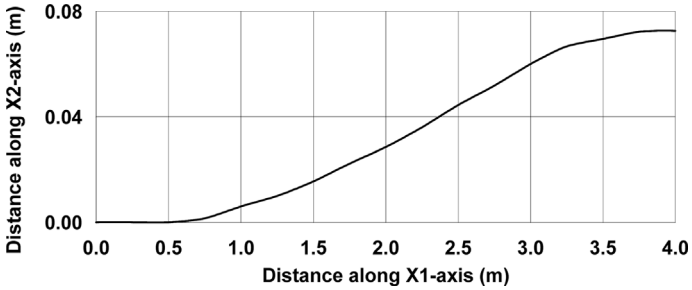

Fig. 4 The optimal contact profile

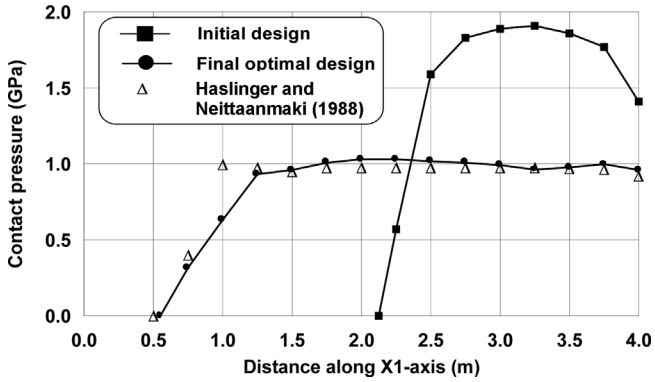

Fig. 5 Variation of the contact pressure for different contact profiles

contact profiles. It is found that the maximum contact pressure is reduced from 1.91 GPa for the initial contact profile to 1.03 GPa for the final optimal contact profile. This reduction is attributed to the increasing of the contact length due to the optimization process. The same problem is solved in Haslinger and Neittaanmaki (1988) and a good agreement with their solution is found.

6.2 Contact of an elastic body with a wavy surface

This example presents the optimal shape design problem of an elastic body with a weavy surface pressed against a rigid support. The potential contact profile of the body is a sinusoidal curve. Due to certain consideration, the contact surface is tolerated to limit dimensions. Therefore, the optimal contact profile is constrained to be located between the upper and lower limit tolerated profiles. The objective of the analysis is to minimize the maximum contact pressure. The material of the body is

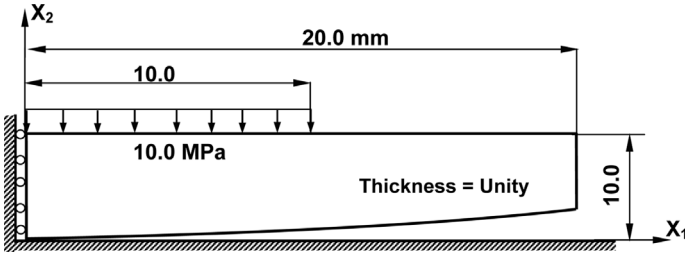

Fig. 6 Contact of an elastic body with a weavy surface and a rigid foundation

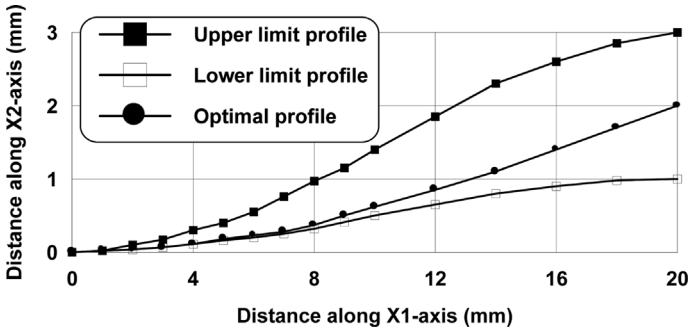

Fig. 7 The optimal contact profile

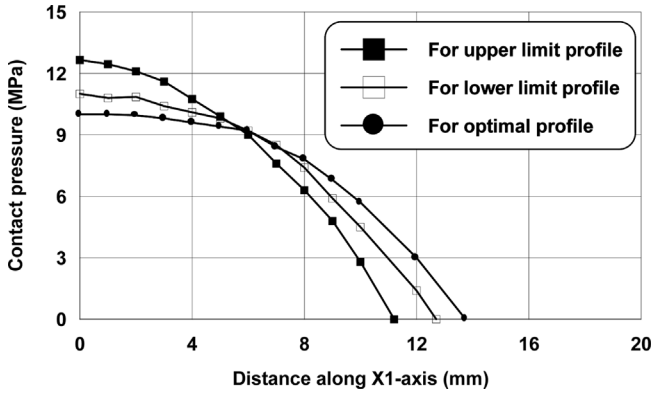

Fig. 8 Variation of the contact pressure for initial and final optimal contact profiles

having a Young's modulus of 15.0 MPa, and Poisson's ratio of 0.3. The geometry and loading conditions are shown in Fig. 6. The optimization process is terminated after eleven iterations where the solution is converged smoothly according to the desired accuracy.

Fig. 7 illustrates the upper and lower limit profiles as well as the optimal profile. It is obviously clear that the obtained optimal contact profile is located between the two limit profiles. Also, it is noticed that the obtained optimal profile is almost regular satisfying the manufacturing and operation conditions. Fig. 8 shows the variation of the contact pressure throughout the contact interface for the initial upper and lower profiles as well as for the final optimal contact profile. It is found that the maximum contact pressure is reduced from 12.6 MPa for the initial upper contact profile to 10.0 MPa for the final optimal contact profile.

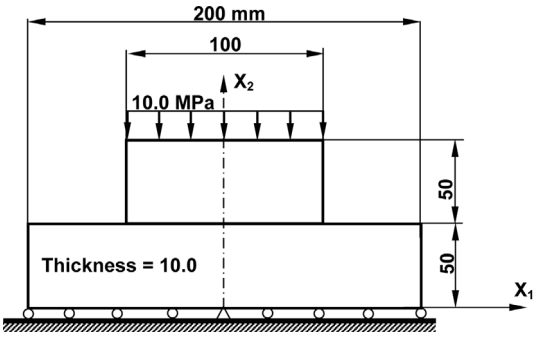

Fig. 9 Contact of two elastic bodies

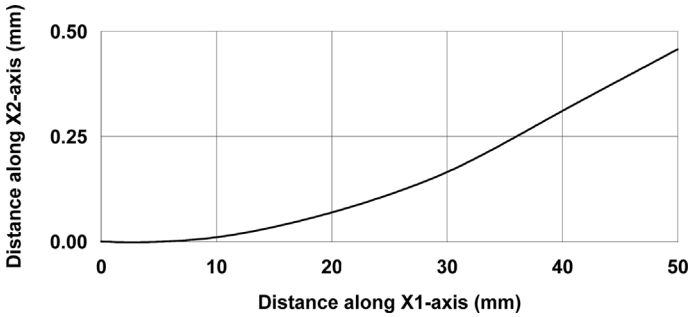

Fig. 10 The optimal contact profile

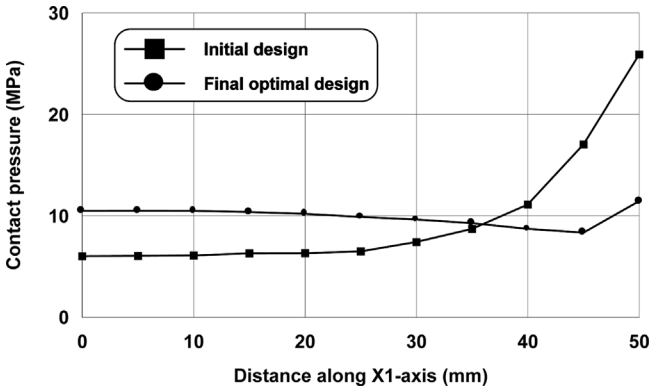

Fig. 11 Variation of the contact pressure for initial and final optimal contact profiles

6.3 Contact of two elastic flat bodies

This example presents the optimal shape design of an elastic flat body pressed against another elastic flat one. The potential contact profile of the system is a flat surface. The objective of the analysis is to minimize the peak contact pressure. The contact interface of the upper body is the optimized profile. The materials of the upper and lower bodies have Young's module of 71.0 GPa and 10.0 GPa, respectively, while the Poisson's ratio is 0.3 for both. The upper surface of the upper body is subjected to uniform traction of 10.0 MPa. The geometry and loading conditions are shown

in Fig. 9. The optimization process is terminated after twenty-four iterations, where the solution is converged according to the desired accuracy.

Fig. 10 displays the optimal contact profile. It is obviously clear that the obtained optimal contact profile is a quadratic smooth one, while the initial profile is a flat surface. Fig. 11 shows the variation of the contact pressure throughout the contact interface for the initial and optimal profiles. In the initial design, a very sharp increase in contact stress is found at the contact boundaries. Through the optimization process, such a concentration of stress can be eliminated. Furthermore, it is clear that the final design has almost uniform distribution of the contact pressure.

6.4 An elastic flat body pressed against a rigid foundation

Consider the design problem of an elastic flat body pressed against a rigid foundation. The initial contact profile is a flat surface. The Young's modulus and Poisson's ratio of the material of the elastic body are 71.0 GPa and 0.3 respectively. The geometry and loading conditions are shown in Fig. 12. The objective of the analysis is to minimize the peak contact pressure. The optimization process is terminated after twenty-two iterations, where the solution is converged according to the desired accuracy.

Fig. 13 illustrates the optimal contact profile. Due to the loading condition, the optimal contact profile is found to be a curved shape with maximum gap at the center. The optimal contact profile for this case is completely different from the preceding three problems. Fig. 14 shows the variation of the contact pressure throughout the contact interface for the initial shape, after twelve iterations and for the optimal profile. In the initial design, the peak contact stress of 19.88 MPa is found. With the application of the developed integrated procedure, the contact pressure distribution is modified

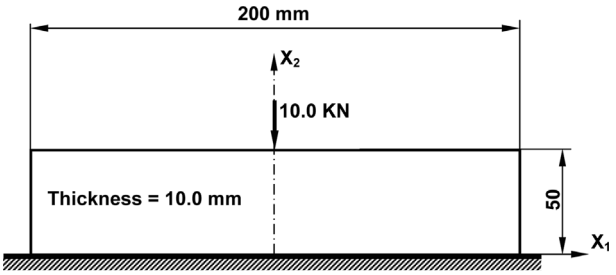

Fig. 12 Contact of an elastic flat body and a rigid foundation

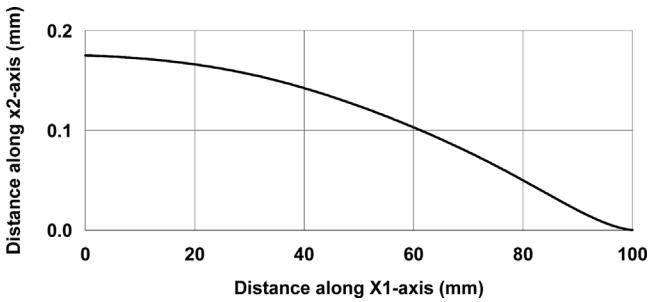

Fig. 13 The optimal contact profile

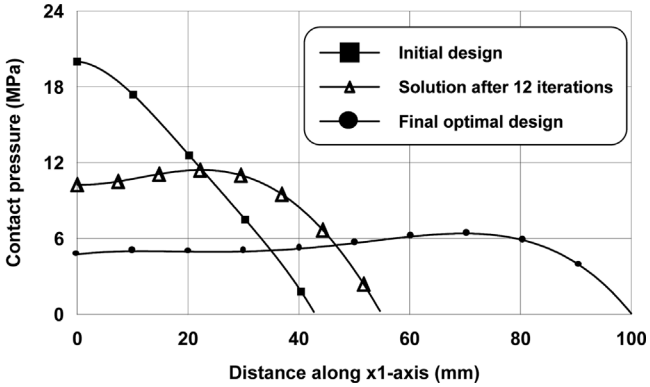

Fig. 14 Variation of contact pressure for different contact profiles

to be almost uniform with a maximum value of 6.29 MPa located near the edges of the contact area. This means that the maximum contact stress, for the initial design, is reduced by 68.5%. This result represents a considerable improvement on the initial design.

7. Conclusions

The distribution of the contact pressure throughout the contact interface plays an important role in the performance of the contact systems. In many applications, it is desirable to have an approximately uniform contact pressure distribution or to minimize the maximum contact stress. Such requirement can be attained through a proper design of the initial contour of the contacting bodies. Therefore, an efficient integrated procedure capable of determining the optimal profiles of contact systems is developed. The proposed procedure consists of two simultaneous models. The first model concerns with the solution of the boundary value problems of non-conformal contact between elastic bodies. The adaptive incremental convex programming method is adopted to solve the contact problem and obtaining the contact status for a given design configuration. The second model investigates the optimal shape design problem. The maximum contact pressure, which is already determined in the first model, is the objective function to be minimized. The augmented Lagrange multiplier method is used to manage and control the shape optimization procedure, in which the contact profile is parameterized using the cubic spline curves. Instead of the coordinates of the nodes of the contact profile, control nodes of a cubic spline curves are used as the design variables. This simulation has several advantages; reducing the dimensions of the optimization problem and consequently reduces the computational time, making a higher convergence rate of the optimization problem and obtaining a regular surface profile satisfying the manufacturing and operational requirements.

The developed procedure is applied to four different problems, with different geometry and boundary conditions, to illustrate the efficiency and versatility of the model. The obtained results are representing a substantial change in the initial contact profile and consequently reducing significantly the peak value of the contact pressure.

References

Arora, J.S. (1989), *Introduction to Optimum Design*, McGraw-Hill, New York.

Arora, J.S. (1990), "Computational design optimization: A review and future directions", *Struct. Safety*, **7**, 131-148.

Arora, J.S. and Haug, E.J. (1979), "Methods of design sensitivity analysis in structural optimization", *AIAA*, **17**, 970-973.

Arora, J.S., Chahande, A.I. and Paeng, J.K. (1991), "Multiplier methods for engineering optimization", *Int. J. Numer. Methods Eng.*, **32**, 1485-1525.

Barboza, J., Fourment, L. and Chenot, J.L. (2002), "Contact algorithm for 3D multi-bodies problems: Application to forming of multi-materials parts and tool deflection", *5th World Congress on Computational Mechanics*, WCCMV, Vienna, Austria, July 7-12.

Belegundu, A.D. and Arora, J.S. (1985), "A study of mathematical programming methods for structural optimization, Part I: Theory, Part II: Numerical aspects", *Int. J. Numer. Methods Eng.*, **21**, 1583-1624.

Beremlijski, P., Haslinger, J., Kocvara, M. and Outrata, J. (2002), "Shape optimization in contact problems with Coulomb friction", *J. Optim.*, **13**(2), 561-587.

Berisekas, D.P. (1976), "On penalty and multiplier methods for constrained minimization", *SIAM J. Control Optim.*, **14**, 216-235.

Campos, L.T., Oden, J.T. and Kikuchi, N. (1982), "A numerical analysis of a class of contact problems with friction in elastostatics", *Comp. Meth. App. Mech. Eng.*, **34**, 821-845.

Chapra, S.C. and Canale, Raymond P. (1989), *Numerical Method for Engineers*, 2nd ed., McGraw-Hill, New York.

Cheney, W. and Kincaid, D. (1985), *Numerical Mathematics and Computing*, 2nd ed., Brooks/Cole, Monterey, CA.

Coope, I.D. and Fletcher, R. (1980), "Some numerical experience with a globally convergent algorithm for nonlinearly constrained optimization", *J. Optim. Theory Appl.*, **32**, 1-16.

Dundurs, J. (1975), "Properties of elastic bodies in contact", *The Mechanics of Contact Between Deformable Bodies*, (Kalker, J.J. and de Pater, A.D., Eds.), Delft Univ. Press.

El-Shafei, A.G. (2004), "Solution of nonlinear frictional contact problems of multibody with application to multileaf springs", *J. Eng. & Appl. Sci., Faculty of Eng., Cairo Univ.*, **51**(2), 309-328.

El-Shafei, A.G. and Mahmoud, F.F. (1999), "Nonlinear analysis of thermoelastic frictional contact problems", *Proc. of the IASTED Int. Conf. Artificial Intelligence and Soft Computing*, Honolulu, Hawaii, USA, August 9-12.

Fletcher, R. (1975), "An ideal penalty function for constrained optimization", *J. Ins. Math. Optim.*, **15**, 319-342.

Fletcher, R. (1987), *Practical Methods of Optimization,* 2nd ed., Wiley, New York.

Francavilla, A. and Zeinkiewicz, O.C. (1975), "A note on numerical computation of elastic contact problems", *Int. J. Numer. Methods Eng.*, **9**, 913-924.

Haslinger, J. and Neittaanmaki, P. (1988), *Finite Element Approximation for Optimal Shape Design: Theory and Applications*, John Wiley & Sons.

Hassan, M.M. and Mahmoud, F.F. (2002), "A generalized adaptive incremental approach for solving inequality problemss of convex nature", *J. Eng. & Appl. Sci., Faculty of Eng., Cairo Univ.*, **49**, 627-645.

Haug, E.J. and Arora, J.S. (1979), *Applied Optimal Design*, John Wiley and Sons, New York.

Haug, E.J. and Kwak, B.M. (1978), "Contact stress minimization by contour design", *Int. J. Numer. Methods Eng.*, **12**, 917-930.

Herskovits, J., Leontiev, A., Dias, G. and Santos, G. (1998), "An interior point algorithm for optimal design of unilateral constrained mechanical systems", in "*Computational Mechanics, New Trends and Applications*", Idelsohn, S., Onate, E. and Dvorkin, E. (Eds.), CIMNE, Barcelona, Spain.

Hilding, D., Klarbring, A. and Petterson, J. (1999), "Optimization of structures in unilateral contact", *J. Appl. Mech. Rev.*, ASME, **52**(4), 139-160.

Johnson, K.L. (1985), *Contact Mechanics*, Cambridge Univ. Press, Cambridge.

Kim, N.H., Choi, K.K. and Botkin, M.E. (2003), "Numerical methods for shape optimization using meshfree method", *Struct. Multidisc. Optim.*, **24**, 418-429.

Kim, N.H., Choi, K.K. and Chen, S.C. (2000), "Shape design sensitivity analysis and optimization of elasto-plasticity with frictional contact", *AIAA*, **38**(9).
Kim, N.H., Park, Y.H. and Choi, K.K. (2001), "Optimization of hyper-elastic structure with multibody contact using continuum-based shape design sensitivity analysis", *Struct. Multidisc. Optim.*, **21**, 196-208.
Kyung, K., Choi, K., Yo, Kiyoung, Kim, N.H. and Botkin, M.E. (2003), "Design sensitivity analysis of nonlinear shell structure with frictionless contact", *Proc. of DETC'03, ASME 2003 Design Engineering Technical Conf.*, Chicago, Illinois, USA, Sep. 2-6.
Mahmoud, F.F., Al-Saffar, A.K. and Hassan, K.A. (1993), "An adaptive incremental approach for the solution of convex programming models", *Math. & Comp. in Simul.*, **35**, 501-508.
Mahmoud, F.F., Ali-Eldin, S.S., Hassan, M.M. and Emam, S.A. (1998), "An incremental mathematical programming model for solving multi-phase frictional contact problems", *Comput. Struct.*, **68**, 567-581.
Mahmoud, F.F., EI-.Sharkawy, A.A. and Hassan, K.M. (1989), "Contour design for contact stress minimization by interior penalty method", *Appl. Math. & Mod.*, **13**, 596-600.
Mahmoud, F.F., Salamon, N.J. and Pawlak, T.P. (1986), "Simulaion of structural elements in receding/advancing contact", *Comput. Struct.*, **22**(4), 629-635.
Paczelt, I. and Mros, Z. (2004), "Contact optimization problems associated with the wear process", XXI ICTAM, Warsaw, Poland, Aug. 15-21.
Powell, M.J.D. (1969), *A Method for Nonlinear Constraints in Minimization Problems*, in R. Fletcher (Ed.), Optimization Academic Press, New York.
Zhong, W.X. and Sun, S.M. (1989), "A parametric quadratic programming approach to elastic contact problems with friction", *Comput. Struct.*, **32**(1), 37-43.

Appendix-A

According to Berisekas (1976), the augmented Lagrange multiplier algorithm is stated as follows:

1. Set the following initial vectors and parameters:
   $k = 0$, initiate the constrained optimization step,
   $\tilde{x}^k$ and $\tilde{\theta}^k$ initiate the vectors of design variables and augmented Lagrange multipliers respectively,
   $\tilde{r}^k$, initiate the vector of penalty parameter,
   $K = \infty$, initiate the maximum constraint violation value, and scalars $\alpha > 1$, $\beta > 1$ and $\varepsilon > 0$,
   where ε is the desired accuracy; α is used to enforce sufficient decrease in the constraint violations, and β is used to increase the penalty parameters.
2. Set $k = k + 1$.
3. Set $\Phi(\tilde{x}^k, \tilde{\theta}^k, \tilde{r}^k)$, according to (6), then minimize it with respect to $\tilde{x}$ and let $\tilde{x}^k$ be the best point obtained in this step.
4. Evaluate $h_i(\tilde{x}^k)$, $i = 1, 2, \ldots, l$ and $g_i(\tilde{x}^k)$, $i = l+1, l+2, \ldots, m$.
5. Set $\overline{K}$, the maximum constraint violation according to (12).
6. Check for convergence criteria according to (14), If these criteria are satisfied, then stop, otherwise, establish the following sets of equality I_E and inequality I_I constraints whose violation did not improve by the factor α:

$$I_E = \{i : |h_i| > K/\alpha,\ i = 1, 2, \ldots, l\} \qquad \text{(Equality constraint set)}$$

$$I_I = \{i : |\max(g_i, -\theta_i)| > K/\alpha,\ \ i = l+1, l+2, \ldots, m\} \qquad \text{(Inequality constraint set)}$$

7. If $\overline{K} < K$, then go to step 8, otherwise
   This condition means that the constraint violation did not improve. Satisfaction of this condition also means that $\overline{K} > K/\alpha$. In this case, increase the penalty parameters by the factor β and reduce the corresponding θ_i by the same factor, such that

$$\tilde{r}^{k+1} = \beta \tilde{r}^k \quad \text{and} \quad \theta_i^{k+1} = \theta_i^k / \beta \quad \text{for all} \quad i \in I_E \cup I_I$$

thus keeping the multipliers unchanged.
Go to step 9.
Note that these calculations are performed only for constraints whose the constraint violation have been improved by the parameter α.

8. Here $\overline{K} < K$, (i.e., the maximum constraint violation is improved) In this case $\overline{K} \le K/\alpha$. According to (10) and (11), update θ_i^k by setting

$$\theta_i^{k+1} = \theta_i^k + h_i(\tilde{x}^k), \qquad i = 1, 2, \ldots, l$$

$$\theta_i^{k+1} = \theta_i^k + \max(g_i(\tilde{x}^k), -\theta_i^k), \qquad i = l+1, l+2, \ldots, m$$

Also, note that these calculations are performed only for constraints whose the constraint violation have been improved.

9. Set $K = \overline{K}$ and go to step 2
The initial choice of $\tilde{\theta}, \tilde{r}$, α and β are very important. Usually initial θ_i are set to zero for all constraints. Considering the rate of convergence, the initial choice of penalty parameters r_i are important as discussed before. Penalty parameters are usually chosen to satisfy $\frac{1}{2} r_i g_i^2(\tilde{x}^0) = |\Delta f(\tilde{x}^0)|$ where $|\Delta f(\tilde{x}^0)|$ is the expected decrease in the objective function at the initial design point x^0. This requires some prior knowledge about the problem. Another way to choose the initial penalty parameters is to impose a condition that the objective function and the generalized penalty function or the norm of their gradients to be equal at the initial point x^0. Different values of the parameters α and β are discussed in (Belegundu and Arora 1985, Berisekas 1976). To prevent excessive iterations, a limit on the maximum number of function and gradient evaluations may be imposed.

Appendix-B

This Appendix presents the unconstrained minimization algorithm according to Arora (1989). After receiving the $\tilde{x}^k$ as the current estimate of the minimum point, as presented in the third step of the algorithm, presented in Appendix-A, the procedure to minimize the unconstrained augmented function is proceeds as follows:

1. Set $n = 0$, initiate the unconstrained minimization step, and set a symmetric positive definite matrix H_{ij}^o
2. Set $\Phi(x_i^o, \tilde{\theta}^k, \tilde{r}^k) = \Phi(x_i^k, \tilde{\theta}^k, \tilde{r}^k)$
3. Set $n = n + 1$ unconstrained minimization counter
4. Compute the search direction vector S^n as follows:

$$S_i^n = -H_{ij}^{n-1} \nabla \Phi^T(x_j^{n-1}, \tilde{\theta}^k, \tilde{r}^k), \; i \text{ and } j = 1, 2, \ldots, ND \tag{B1}$$

where ND is the number of design variables

5. Using the Golden section method (Haug and Arora 1979) to evaluate the scalar parameter η, which represents the maximum step size in the minimization search direction of the augmented function $\Phi(x_i^{n-1}, \tilde{\theta}^k, \tilde{r}^k)$, such that $\Phi(x_i^{n-1} + \eta S_i, \tilde{\theta}^k, \tilde{r}^k)$ is minimum.
6. Compute:

$$x_i^n = x_i^{n-1} + \eta S_i, \qquad i = 1, 2, \ldots, ND \tag{B2}$$

$$H_{ij}^n = H_{ij}^{n-1} + A_{ij} + C_{ij}, \qquad i, j = 1, 2, \ldots, ND \tag{B3}$$

where

$$A_i = \frac{\sigma_i \sigma_i^T}{\sigma_i^T y_i}, \qquad i = 1, 2, \ldots, ND \tag{B4}$$

and $$C_i = \frac{H_{ij}^n y_i y_i^T H_{ij}^n}{y_i^T H_i y_i}, \qquad i, j = 1, 2, \ldots, ND \tag{B5}$$

$$\sigma_i = \eta_i S_i^n \tag{B6}$$

with $$y_i = \nabla \Phi^T(x_i^n, \tilde{\theta}^k, \tilde{r}^k) - \nabla \Phi^T(x_i^{n-1}, \tilde{\theta}^k, \tilde{r}^k) \qquad i = 1, 2, \ldots, ND \tag{B7}$$

7. If $\nabla \Phi(x_i^n, \tilde{\theta}^k, \tilde{r}^k) \le \varepsilon$ or $\left\| x_i^n - x_i^{n-1} \right\| \le \varepsilon, \quad i = 1, 2, \ldots, ND$ (B8)
is sufficiently small, then terminate the process. Else, go to step 3

Finally, it is important to mention that for engineering design problems, it is necessary to impose some bounds or limitations on the design variables during the unconstrained minimization process to prevent impractical designs.