Struct Multidisc Optim 20, 29–36

A heuristic smoothing procedure for avoiding local optima in optimization of structures subject to unilateral constraints

D. Hilding

Abstract Structural optimization problems are often solved by gradient-based optimization algorithms, e.g. sequential quadratic programming or the method of moving asymptotes. If the structure is subject to unilateral constraints, then the gradient may be nonexistent for some designs. It follows that difficulties may arise when such structures are to be optimized using gradient-based optimization algorithms. Unilateral constraints arise, for instance, if the structure may come in frictionless contact with an obstacle. This paper presents a heuristic smoothing procedure (HSP) that lessens the risk that gradient-based optimization algorithms get stuck in (nonglobal) local optima of structural optimization problems including unilateral constraints. In the HSP, a sequence of optimization problems must be solved. All these optimization problems have well-defined gradients and are therefore well-suited for gradient-based optimization algorithms. It is proven that the solutions of this sequence of optimization problems converge to the solution of the original structural optimization problem.

The HSP is illustrated in a few numerical examples. The computational results show that the HSP can be an effective method for avoiding local optima.

Key words unilateral constraints, smoothing procedure, gradient-based algorithms, finite elements, method of moving asymptotes (MMA), trusses

1 Introduction and preliminaries

This paper deals with the problem of optimizing structures subject to unilateral constraints. The assumptions made include linear elastic discrete and finite element approximated structures under a small displacement assumption. The type of unilateral constraints allowed include constraints that arise from frictionless contact supports and members such as elastic ropes, which can support load in tension only. Examples of such structural optimization problems may be found in the paper by Hilding, Klarbring and Petersson (1999) (review of optimization of structures in contact) and in the book by Haslinger and Neittaanmäki (1988) (shape optimization).

The following structural optimization problem is considered:

$$(P) \quad \min_{\mathbf{s}} \; c(\mathbf{u}(\mathbf{s}), \mathbf{p}(\mathbf{s}), \mathbf{s}) \,,$$

$$\text{subject to } \mathbf{s} \in S \,, \tag{1}$$

where c is the cost function, $\mathbf{s} \in I\!R^{n_s}$ is the design variable, $\mathbf{u} \in I\!R^{n_d}$ is the displacement of the structure, and $\mathbf{p} \in I\!R^{n_c}$ is the force due to the unilateral constraints, e.g. the contact force in case of frictionless contact. The design space S is the set of allowed designs and $\mathbf{s} \in S$ expresses the design constraints, e.g. a limit on the maximum weight of the structure or simple upper and lower limits on the components of the design variable.

Under the assumptions of Theorem 1 in Sect. 3.2, the state variables of the structure are continuous (but not in general differentiable) functions of the design. It is assumed that the state problem determining the state variables $\mathbf{u}(\mathbf{s})$ and $\mathbf{p}(\mathbf{s})$ has the following form[1]:

$$\mathbf{K}(\mathbf{s})\mathbf{u}(\mathbf{s}) + \mathbf{C}(\mathbf{s})^T\mathbf{p}(\mathbf{s}) = \mathbf{f}(\mathbf{s}) \,, \tag{2}$$

$$(\mathbf{g}(\mathbf{s}) - \mathbf{C}(\mathbf{s})\mathbf{u}(\mathbf{s}))_i \geq 0 \,, \tag{3}$$

$$p_i(\mathbf{s}) \geq 0 \,, \tag{4}$$

$$p_i(\mathbf{s})(\mathbf{g}(\mathbf{s}) - \mathbf{C}(\mathbf{s})\mathbf{u}(\mathbf{s}))_i = 0 \,, \quad i = 1 \ldots n_c \,, \tag{5}$$

In a mechanical setting (2)–(5) may be given the following interpretation: (2) is force equilibrium combined with a linear constitutive law, where $\mathbf{K}(\mathbf{s})$ is the stiffness ma-

Received February 26, 1999
Revised manuscript received July 15, 1999

D. Hilding

Division of Mechanics, Department of Mechanical Engineering, Linköping University, SE-581 83 Linköping, Sweden
e-mail: `danhi@ikp.liu.se`

[1] Notation: vectors are one column matrices and $(\mathbf{w})_i$ is the i-th component of the vector $\mathbf{w}$

trix and $\mathbf{f}(\mathbf{s})$ is the load vector. The inequalities (3) are the unilateral constraints, (4) ensures that the forces from the unilateral constraints do not change direction, and (5) ensures that a force is zero if its associated unilateral constraint is not fulfilled as an equality. A detailed derivation of the state problem (2)–(5) for an elastic structure in frictionless contact can be found in e.g. the paper by Hilding, Klarbring and Petersson (1999).

Structural optimization problems are often solved using gradient-based optimization algorithms. Two such algorithms are sequential quadratic programming (SQP) (Bazaraa *et al.* 1993) and the method of moving asymptotes (MMA) (Svanberg 1987). Applying a gradient-based algorithm to (P) has a conceptual problem: due to the unilateral constraints, $\mathbf{u}$ and $\mathbf{p}$ are in general not differentiable functions of the design variable. Thus, the gradient may not exist at some points in the design space S. With the motivation that the nondifferentiable points of (P) often are very scarce, one may use a gradient-based algorithm anyway. This has indeed been found to be a valid approach, see Klarbring and Rönnqvist (1995). However, Hilding, Klarbring and Pang (1999) found that *in such an approach the optimization algorithm tended to get stuck in (nonglobal) local optima of* (P). A possible explanation of why there may be many local optima in (P) is offered in Sect. 2.

The objective of the present paper is to present a heuristic, designed to lessen the risk that gradient-based algorithms get stuck in (nonglobal) local optima of (P). The procedure will be denoted the heuristic smoothing procedure (HSP). The HSP is a quite simple technique and may be used to increase the performance of any gradient-based algorithm, when applied to structural optimization problems with unilateral constraints, i.e. (P).

The HSP is similar to the penalty interior point algorithm (PIPA), Hilding, Klarbring and Pang (1999), which seems good at avoiding local optima of some structural optimization problems with unilateral constraints. The idea of the HSP stems from the experience the author had with PIPA. The HSP is also most similar to a procedure described by Facchinei *et al.* (1999). Compared to the latter work the novelty in the present paper lies mainly in the application to structural optimization and the purpose (in the cited work the purpose is not to avoid local optima).

2 Local optima of (P)

In this section it is argued that the introduction of unilateral constraints in the state problem may introduce additional nonglobal local optima in (P). The small example below serves as an illustration.

Consider the following simple model 1D problem ($n_s = n_c = n_d = 1$) of minimizing the tip displacement of a bar, see the upper part of Fig. 1. The cost function is $c(\mathbf{u}, \mathbf{p}, \mathbf{s}) = \mathbf{u}$. For simplicity, let $\mathbf{f} = 1, \mathbf{g} = 0, \mathbf{C} = 0$, select the cross-sectional area of the bar as design variable, i.e. $\mathbf{K}(\mathbf{s}) = \mathbf{s}$, and set the design space $S = \{\text{all } \mathbf{s} \text{ such that } 0.5 \le \mathbf{s} \le 2\}$. The graph of the cost function is displayed in Fig. 2 and shows that the solution to the problem is $\mathbf{s}^* = 2$ and that there are *no nonglobal local optima*.

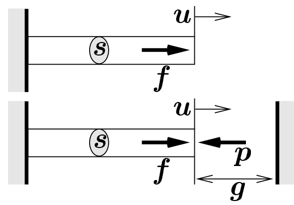

Fig. 1 Example problem structure shown both with and without the unilateral constraint, which models a rigid obstacle

Now introduce a unilateral constraint, modeling a rigid obstacle with which the bar may come into frictionless contact, by setting $\mathbf{g} = 1$ and $\mathbf{C} = 1$, see the lower part of Fig. 1. Again, the graph of the cost function is displayed in Fig. 2 and shows that the solution to the problem is still $\mathbf{s}^* = 2$, but that there are now *many local optima*. Thus, in this case, *introducing unilateral constraints introduces additional nonglobal local optima.*

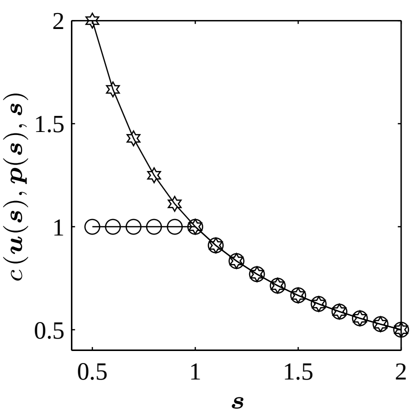

Fig. 2 Example problem cost function for all $\mathbf{s} \in S$: lower curve is with and the upper curve is without unilateral constraint

The above local optima appeared together with jumps in the gradients of $\mathbf{u}$ and $\mathbf{p}$. These jumps may occur when $\mathbf{s}$ is changed in such a way that a unilateral constraint changes from being fulfilled as an equality to being fulfilled as a strict inequality or vice versa. As the jumps and local optima appeared together, it does not seem far-fetched to assume that these jumps are partly responsible for the local optima. If this is the case then a multitude of local optima might be introduced even for moderate n_c, because the amount of possible jumps may grow exponentially in n_c. There is evidence, see Hilding, Klarbring and Pang (1999), that this introduction of nonglobal local optima may occur for realistic instances of (P).

3 The heuristic smoothing procedure

In the previous section it was suggested that the jumps in the gradients of the state variables $\mathbf{u}$ and $\mathbf{p}$, which are due to the unilateral constraints, are responsible for some of the nonglobal local optima in (P). A gradient-based optimization algorithm may get stuck in these local optima, see Hilding, Klarbring and Pang (1999).

The HSP is a heuristic that helps a gradient-based method to avoid nonglobal local optima in (P). In the HSP a sequence of optimization problems are solved for decreasing values of a smoothing parameter $\mu > 0$,

$$(P_\mu) \quad \min_{\mathbf{s}} \quad c(\mathbf{u}_\mu(\mathbf{s}), \mathbf{p}_\mu(\mathbf{s}), \mathbf{s}),$$

$$\text{subject to } \mathbf{s} \in S. \tag{6}$$

Problem (P_μ) differs from (P) in that $\mathbf{u}$ and $\mathbf{p}$ have been replaced with approximations $\mathbf{u}_\mu$ and $\mathbf{p}_\mu$, where the gradient jumps have been smeared out or smoothed. As the jumps of the gradients have been smoothed in (P_μ), it does not seem unlikely (in the light of the argument in Sect. 2) that (P_μ) has fewer nonglobal local optima than (P). If this is the case, then a gradient algorithm is less likely to get stuck in nonglobal local optima in (P_μ) than in (P). Gradient algorithms are well-suited for solving (P_μ), because the gradient is always well-defined; see Theorem 1. This is in contrast to (P), where the gradient may not exist for some designs.

The smoothing parameter μ indicates the level of smoothing, it is assumed that $\mu = 0$ means no smoothing and larger values of μ means more smoothing. It follows that if $\mu = 0$ then (P_μ) is identical to (P). Selecting large μ has the advantage that the gradients are smeared a lot, but the disadvantage that (P_μ) may be a too poor approximation of (P) to render useful solutions. To avoid this problem the following procedure is proposed, which will be referred to as the heuristic smoothing procedure (HSP).

The heuristic smoothing procedure (HSP)

Step 0 (initialization). Select an initial design $\mathbf{s}^0 \in S$, a decreasing sequence of smoothing parameters $\{\mu_k\}$, $\mu_k > 0$, and set $k = 0$.

Step 1 (main step). Apply a gradient optimization algorithm to (P_{μ_k}) starting at $\mathbf{s}^k$. This results in a new iterate $\mathbf{s}^{k+1}$.

Step 2 (repetition). Set $k = k + 1$, and go to Step 1.

In the numerical examples the following simple sequence was used: $\{\mu_k\} = \{10, 1{\times}10^{-3}\}$. In the test problems, it was found that if $\mu \leq 1{\times}10^{-3}$ then the difference between (P) and (P_μ) was negligible.

3.1 Creating $\mathbf{u}_\mu$ and $\mathbf{p}_\mu$

The idea is to replace the state problem (2)–(5) with the following smoothed approximation:

$$\mathbf{K}(\mathbf{s})\mathbf{u}_\mu(\mathbf{s}) + \mathbf{C}(\mathbf{s})^T \mathbf{p}_\mu(\mathbf{s}) = \mathbf{f}(\mathbf{s}), \tag{7}$$

$$\phi_\mu((\mathbf{p}_\mu(\mathbf{s}))_i\,,\ (\mathbf{g}(\mathbf{s}) - \mathbf{C}(\mathbf{s})\mathbf{u}_\mu(\mathbf{s}))_i) = 0\,,$$

$$i = 1, \dots, n_c\,, \tag{8}$$

where μ is the smoothing parameter and the smoothing function is

$$\phi_\mu(a, b) = (a + b) - \sqrt{(a - b)^2 + 4\mu^2}. \tag{9}$$

The function ϕ_μ has the property that

$$\phi_\mu(a, b) = 0 \Leftrightarrow a \geq 0,\ b \geq 0,\ ab = \mu^2, \tag{10}$$

which implies that if $\mu = 0$ then $\mathbf{u}_\mu = \mathbf{u}$ and $\mathbf{p}_\mu = \mathbf{p}$.

The above smoothing function is the same as that of Facchinei *et al.* (1999). For other smoothing functions see e.g. Chen and Mangasarian (1995) and Leung *et al.* (1998).

The above approach is not to be confused with the so-called penalty method, which is sometimes used in the solution of contact problems.

3.2 Some important properties of (P_μ)

The question of what is the relation between (P) and (P_μ), is to some extent answered in the following theorem.

Theorem 1

Suppose that

(i) the set S is nonempty, bounded, and closed,
(ii) the rows of $\mathbf{C}(\mathbf{s})$ are linearly independent for all $\mathbf{s} \in S$,
(iii) $\mathbf{K}(\mathbf{s})$ is positive definite for all $\mathbf{s} \in S$,
(iv) $\mathbf{K}, \mathbf{f}, \mathbf{C}$, and $\mathbf{g}$ are i times continuously differentiable, where $i \geq 1$, on $\mathbb{R}^{n_s}$,
(v) c is continuous on $\mathbb{R}^{n_d} \times \mathbb{R}^{n_c} \times S$,

then it holds that

(a) the state problem (2)–(5) has a unique solution $\mathbf{u}(\mathbf{s})$, $\mathbf{p}(\mathbf{s})$ for all $\mathbf{s} \in S$,
(b) the smoothed state problem (7)–(8) has a unique solution $\mathbf{u}_\mu(\mathbf{s})$, $\mathbf{p}_\mu(\mathbf{s})$ for all μ and all $\mathbf{s} \in S$,

(c) $\mathbf{p}_\mu \to \mathbf{p}$ and $\mathbf{u}_\mu \to \mathbf{u}$ uniformly on S as $\mu \to 0$,
(d) if $\mu > 0$ then $\mathbf{u}_\mu$ and $\mathbf{p}_\mu$ are i times continuously differentiable on some open set containing S,
(e) $\mathbf{u}$ and $\mathbf{p}$ are continuous on S,
(f) both (P_μ), for all $\mu > 0$, and (P) have a least one solution,
(g) let S^*_μ be the solution set of (P_μ) and S^* be the solution set of (P), then

$$\lim_{\mu\to 0} \max_{\mathbf{s}^*_\mu \in S^*_\mu} \min_{\mathbf{s}^* \in S^*} \|\mathbf{s}^* - \mathbf{s}^*_\mu\|_\infty = 0. \tag{11}$$

Proof: See the Appendix. □

Part (g) of the theorem essentially assures that if μ is small, then a solution of (P_μ) is close to a solution of (P). [If $\mu = 0$ then (P_μ) is identical to (P), and thus has the same solutions as (P).] This justifies calling (P_μ) an approximation of (P). Note that, if the assumptions hold and the cost function is differentiable, then (P_μ) is a *constrained differentiable optimization problem*, for which there are many standard gradient-based algorithms, cf. Bazaraa *et al.* (1993).

3.3 Example problem

To see the effect of the HSP, consider again the small example problem in Sect. 2. The following sequence of smoothing parameters is used: $\{\mu_k\} = \{0.3, 0.2, 0.1, 0.05\}$. The graph of the cost function of (P_{μ_k}) for $k = 1, \ldots, 4$ and (for comparison) that of (P) can be found in Fig. 3. It is clearly seen that the gradient jump at $\mathbf{s} = 1$ in (P) is smeared out in (P_{μ_k}).

From Fig. 3 one may conclude that, as opposed to (P), (P_{μ_k}) has no nonglobal local optima. Thus, any gradient-based optimization algorithm is most likely to find the global optima $\mathbf{s}^k = 2$ of (P_{μ_k}), $k = 1, \ldots, 4$. The final iterate of the HSP will therefore be $\mathbf{s}^4 = 2$, which is the global optimum of (P). Using the HSP on this simple problem will therefore lead to that most gradient algorithms will find the global optimum of (P).

If the HSP is not used and the starting point of the gradient algorithm is $0.5 \le \bar{\mathbf{s}} < 1$, then the gradient algorithm will get stuck immediately in the nonglobal local optima $\bar{\mathbf{s}}$. Thus, for this problem, it is beneficial to use the HSP.

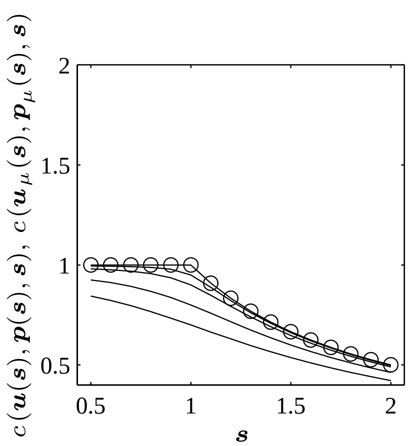

Fig. 3 Example problem (case with unilateral constraint). Curves from below upwards: cost function of (P_{μ_k}) for $k = 1, \ldots, 4$, followed by that of (P)

4 Solving the state problem

To calculate $\mathbf{u}_\mu(\mathbf{s})$ and $\mathbf{p}_\mu(\mathbf{s})$, (7)–(8) must be solved. The differentiable ($\mu > 0$) nonlinear system (7)–(8) can be solved using any Newton algorithm, in the numerical examples the algorithm of Pang (1990) was used. It was found that this algorithm worked excellently also when $\mu = 0$.

In Newton's method a system of linear equations must be solved to find the next iterate. Assuming that the state problem has a unique solution, smoothing has the benefit that the linear system cannot be singular, which may happen if smoothing is not used. There is also evidence that combining smoothing and Netwon's method may be a very effective method for solving contact problems, see Leung *et al.* (1998). The numerical experiments reported in Sect. 6 do not contradict this; a very accurate solution was on the average achieved within less than 6 iterations. The fast *average* convergence is partly due to that when solving the structural optimization problem a large number of contact problems with quite small differences in design must be solved. Hence it is possible to get a good starting iterate for the Newton solver, which implies fast convergence.

5 Sensitivity analysis

In the main step in the HSP, (P_μ) is to be solved by a selected gradient-based optimization algorithm. To do this, besides $\mathbf{u}_\mu(\mathbf{s})$ and $\mathbf{p}_\mu(\mathbf{s})$, also their derivatives with respect to $\mathbf{s}$ must be obtained. The latter is the objective of the sensitity analysis.

Assuming that (i)–(v) of Theorem 1 holds and that $\mu > 0$, the derivatives of $\mathbf{u}_\mu$ and $\mathbf{p}_\mu$ can be obtained by implicitly differentiating (7)–(8) with respect to the design variable $\mathbf{s}$. This yields the following system of linear equations for the derivatives $\partial \mathbf{u}_\mu / \partial s_k$ and $\partial \mathbf{p}_\mu / \partial s_k$, $k = 1, \ldots, n_s$ [for convenience the dependence of $\mathbf{K}$, $\mathbf{C}$, $\mathbf{f}$, $\mathbf{g}$, $\mathbf{u}_\mu$, and $\mathbf{p}_\mu$ on $\mathbf{s}$ is not written out]:

$$\mathbf{K}\frac{\partial \mathbf{u}_\mu}{\partial s_k} + \mathbf{C}^T \frac{\partial \mathbf{p}_\mu}{\partial s_k} = \frac{\partial \mathbf{f}}{\partial s_k} - \frac{\partial \mathbf{K}}{\partial s_k}\mathbf{u}_\mu - \left(\frac{\partial \mathbf{C}}{\partial s_k}\right)^T \mathbf{p}_\mu\,,$$

$$D_1\{\phi_\mu\}((\mathbf{p}_\mu)_i, (\mathbf{g} - \mathbf{C}\mathbf{u}_\mu)_i)\left(\frac{\partial \mathbf{p}_\mu}{\partial s_k}\right)_i +$$

$$D_2\{\phi_\mu\}((\mathbf{p}_\mu)_i, (\mathbf{g}-\mathbf{C}\mathbf{u}_\mu)_i)\left[\frac{\mathrm{d}\mathbf{g}}{\mathrm{d}s_k} - \right.$$

$$\left.\mathrm{row}_i\left(\frac{\mathrm{d}\mathbf{C}}{\mathrm{d}s_k}\right)\mathbf{u}_\mu - \mathrm{row}_i(\mathbf{C})\frac{\partial \mathbf{u}_\mu}{\partial s_k}\right] = 0\,,$$

$$i = 1, \dots, n_c\,, \tag{12}$$

where $D_j\{\phi_\mu\}$ is the derivative of ϕ_μ with respect to its j-th argument,

$$D_1\{\phi_\mu\}(a,b) = 1 - (a-b)/\sqrt{(a-b)^2 + 4\mu^2}\,,$$

$$D_2\{\phi_\mu\}(a,b) = 1 + (a-b)/\sqrt{(a-b)^2 + 4\mu^2}\,.$$

In the case that $\mu = 0$, the above equations for the derivative are valid as long as not both $p_i = 0$ and $(\mathbf{g} - \mathbf{C}\mathbf{u})_i = 0$ for any $i = 1, \dots, n_c$.

6
Numerical evaluation

To evaluate the efficiency with which the HSP helps a gradient-based optimization algorithm to avoid non-global local optima, the following test was carried out. Starting with an initial design $\mathbf{s}^0$ the gradient optimization algorithm was applied both *with* and *without* the HSP to a structural optimization problem.

Two different gradient algorithms were used in the tests of the HSP: (i) the implicit programming algorithm (alternative variant) (IMPAA), see Hilding, Klarbring and Pang (1999), and (ii) the method of moving asymptotes (MMA), see Svanberg (1987). The IMPAA is essentially an SQP algorithm with line-search. Its main feature is that it is designed to solve also nondifferentiable problems such as (P). The MMA is a sequential programming algorithm for differentiable nonlinear programs. The latter algorithm is known to work well for a large variety of structural optimization problems.

The structural optimization problem in the test is to minimize the maximum force from the unilateral constraints. The structure is a truss and the cross-section areas of the bars are taken as design variables. Design constraints in the form of a maximum volume constraint and upper and lower limits on the cross-section areas are used. The problem can be written as follows:

$$\min_{\mathbf{s}} \quad \max_{i\in[1,\dots,sn_c]} p_i(\mathbf{s})\,,$$

$$\text{subject to} \quad s_i^{\mathrm{low}} \le s_i \le s_i^{\mathrm{upp}}\,, \quad i = 1, \dots, n_s\,,$$

$$\sum_{i=1}^{n_s} s_i \ell_i \le V_{\max}\,, \tag{13}$$

where ℓ_i is the length of member i, $\mathbf{s}^{\mathrm{low}}$ and $\mathbf{s}^{\mathrm{upp}}$ are lower, respectively upper bounds on the cross-sectional areas, and $V_{\max}$ is the maximum allowed total volume of the bars .

The cost function in (13) is continuous, but not differentiable. This poses no problem for IMPAA, but MMA may have difficulties with nondifferentiable cost functions. Therefore, MMA is instead used together with the following alternate form of (13), which has a differentiable cost function:

$$\min_{\mathbf{s}} \; r\,,$$

$$\text{subject to} \quad s_i^{\mathrm{low}} \le s_i \le s_i^{\mathrm{upp}}\,, \quad i = 1, \dots, sn_s\,,$$

$$\sum_{i=1}^{n_s} s_i \ell_i \le V_{\max}\,, \quad p_i(\mathbf{s}) \le r\,, \quad i = 1, \dots, sn_c\,. \tag{14}$$

Test results are reported for two different trusses: (i) a truss resting on a frictionless support; $\mathbf{p}$ is in this case the contact force; and (ii) a truss supported by nonextensible ropes; $\mathbf{p}$ is in this case the force in the ropes. For truss 1, the test was performed for several different sizes of the truss.

Illustrations of the trusses can be found in Figs. 4 and 5. In the figures solid lines represent bars, dotted lines ropes, and the letter F a force with magnitude 1 MN. All bars have a Young's modulus of 200 GPa. For both trusses, the following design limits where used: $s_i^{\mathrm{low}} = 0.001\,\mathrm{m}^2$ and $s_i^{\mathrm{upp}} = 0.1\,\mathrm{m}^2$, $i = 0, \dots, n_s$. For both trusses the initial design was set to $s_i^0 = 0.02\,\mathrm{m}^2$, $i = 0, \dots, n_s$, and $V_{\max}$ was set to 1.5 times the volume of

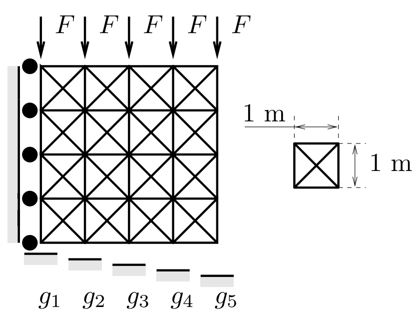

Fig. 4 Truss 1, a truss in contact with a number of rigid obstacles (5×5 nodes case shown)

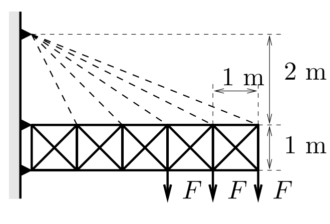

Fig. 5 Truss 2, a truss supported by a number of nonelastic ropes

the initial design ($\sum_{i=1}^{n_s} s_i^0 \ell_i$). The gap between the stair shaped obstacle and structure in truss 1 is given by $g_i = d(i-1)^2$, where $d = 0.001$ m. The initial slack of the ropes in truss 2 is given by $g_i = 0.005$ m.

A stopping criteria is needed for MMA and IMPAA in Step 1 of the HSP. The iterations in MMA/IMPAA in Step 1 of the HSP where stopped when the decrease in the cost function was less than 1×10^{-6} MN between iterations or when the number of iterations exceeded 30.

Truss 2 fulfills all the prerequisites of Theorem 1. Truss 1 can make rigid body motions and therefore $\mathbf{K}(\mathbf{s})$ is not positive definite, thus the proof of Theorem 1 does not hold for truss 1. Though unproven the approach and implementation seem to be valid anyway. In fact, it is likely that the conclusions of Theorem 1 hold as long as the equilibrium problem has a unique solution (which it has for truss 1).

6.1 Numerical results

The results of the test can be found in Tables 1 and 2. The tables show that the HSP had a good effect on both MMA and IMPAA for this structural optimization problem; the HSP lead to a better optima for all tested trusses and gradient algorithms (IMPAA/MMA).

Due to the mechanical situation, for truss 1 the cost function value at a global optimum is at least 1 MN. When HSP was used, both MMA and IMPAA succeeded in finding a global optimum, except in one case (5×5) where MMA got stuck in a local optima. For truss 2, when the HSP was used, IMPAA gave a slightly better result than MMA. This was due to slow convergence of MMA. If more iterations had been allowed then MMA would have given the same result as IMPAA for truss 2.

When the HSP was not used, both MMA and IMPAA got stuck in (nonglobal) local minima well within the 30 iterations for all tested problems.

Table 1 Cost function after optimization of truss 1

Problem size	$\max_{i\in[1,\dots,sn_c]} p_i$ (MN) with HSP IMPAA	with HSP MMA	without HSP IMPAA	without HSP MMA
3×3	1.000	1.001	3.000	3.000
4×4	1.000	1.000	2.000	2.000
5×5	1.000	1.250	2.500	2.500
6×6	1.213	1.500	3.000	3.000

Table 2 Cost function after optimization of truss 2

$\max_{i\in[1,\dots,sn_c]} p_i$ (MN) with HSP IMPAA	with HSP MMA	without HSP IMPAA	without HSP MMA
0.882	0.902	1.130	1.130

7 Conclusions

For the problems tested the HSP performed well, resulting in better optima, for both the IMPAA and MMA algorithms. This indicates that the choice of optimization algorithm used together with the HSP is not very important. The HSP is also easy to implement, as it does not require any modification to the chosen optimization algorithm. Furthermore, if the cost function is differentiable, then the HSP requires the solution of a constrained differentiable optimization problem, for which a lot of standard software exists. A better scheme for updating the smoothing parameter μ_k might be beneficial, even though the HSP performed well with the present very simple scheme.

Finally, neither problem (P) nor the test problems include behavioural constraints, i.e. constraints involving the state variables, such as a von Mises stress limit. Including such constraints poses no formal difficulties, it is only necessary to add the behavioural constraints to (P) and (P_μ). It is also possible to extend Theorem 1 to cover behavioural constraints, even though (P_μ) must be slightly modified to accomplish this [it is necessary to add the behavioural constraints to (P_μ) in a "relaxed" way, analogously to Chang (1992), Rozvany and Sobieszczanski-Sobieski (1992) or Cheng and Guo (1997)].

Acknowledgements This work was based on research supported by the Swedish Research Council for Engineering Sciences (TFR).

References

Bazaraa, M.S.; Sherali, H.D.; Shetty, C.M. 1993: *Nonlinear programming: theory and algorithms.* New York: John Wiley & Sons

Chang, K.J. 1992: Optimality criteria methods using K-S functions. *Struct. Optim.* **4**, 213–217

Chen, C.; Mangasarian, O.L. 1995: Smoothing methods for convex inequalities and linear complementarity problems. *Math. Prog.* **71**, 51–69

Cheng, G.D.; Guo, X.M. 1997: ε-relaxed approach in structural topology optimization. *Struct. Optim.* **13**, 258–266

Cottle, R.W.; Pang, J.-S.; Stone, R.E. 1992: *The linear complementarity problem.* San Diego: Academic Press

Facchinei, F.; Jiang, H.; Qi, L. 1999: A smoothing method for mathematical programs with equilibrium constraints. *Math. Prog.* **85**, 107–134

Fiacco, A.V. 1983: *Introduction to sensitivity and stability analysis in nonlinear programming*. Orlando, FL: Academic Press

Harker, P.T.; Pang, J.-S. 1990: Finite dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Math. Prog.* **48**, 161–220

Haslinger, J.; Neittaanmäki, P. 1988: *Finite element approximation for optimal shape design: theory and applications*. Chichester: John Wiley & Sons

Hilding, D.; Klarbring, A.; Pang, J.-S. 1999: Minimization of maximum unilateral force. *Comp. Meth. Appl. Mech. Eng.* **177**, 215–234

Hilding, D.; Klarbring, A.; Petersson, J. 1999: Optimization of structures in unilateral contact. *Appl. Mech. Rev.* **52**, 139–160

Klarbring, A.; Rönnqvist, M. 1995: Nested approach to structural optimization in nonsmooth mechanics. *Struct. Optim.* **10**, 79–86

Leung, A.Y.T.; Guoquing, C.; Wanji, C. 1998: Smoothing Newton method for solving two- and three-dimensional frictional contact problems. *Int. J. Num. Meth. Eng.* **41**, 1001–1027

Pang, J.-S. 1990: Newton's method for B-differentiable equations. *Math. Oper. Res.* **15**, 311–341

Rozvany, G.I.N.; Sobieszczanski-Sobieski, J. 1992: New optimality criteria methods: forcing uniqueness of the adjoint strains by corner-rounding at constraint intersections. *Struct. Optim.* **4**, 244–246

Rudin, W. 1976: *Principles of mathematical analysis*, 3-rd edition. New York: McGraw-Hill

Svanberg, K. 1987: The method of moving asymptotes – a new method for structural optimization. *Int. J. Num. Meth. Eng.* **24**, 359–373

Appendix

The proof of Theorem 1.

Proof:

For convenience the dependence of $\mathbf{K}$, $\mathbf{C}$, $\mathbf{f}$, $\mathbf{g}$, $\mathbf{u}$, $\mathbf{p}$, $\mathbf{u}_\mu$, and $\mathbf{p}_\mu$ on $\mathbf{s}$ is not written out in (a)–(d).

(a) Basic matrix operations yield that (2)–(5) has the same solution set as

$$(\mathbf{Mp}+\mathbf{q})_i \geq 0\,, \tag{15}$$

$$p_i \geq 0\,,\ (\mathbf{Mp}+\mathbf{q})_i p_i = 0\,, \tag{16}$$

$$\mathbf{u} = \mathbf{K}^{-1}(\mathbf{f}-\mathbf{Cp})\,, \quad i=1,\dots,n_c\,, \tag{17}$$

where $\mathbf{M} = \mathbf{CK}^{-1}\mathbf{C}^T$ is positive definite due to (ii)–(iii) and $\mathbf{q} = \mathbf{g}-\mathbf{CK}^{-1}\mathbf{f}$. Because $\mathbf{M}$ is positive definite, it follows from Theorem 3.1.6 by Cottle *et al.* (1992) that the linear complementarity problem (15)–(16) has an unique solution $\mathbf{p}$. The existence and uniqueness of $\mathbf{u}$ follows from that of $\mathbf{p}$ together with (17). q.e.d.

(b) Basic matrix operations together with (10) yield that (7)–(8) has the same solution set as

$$(\mathbf{Mp}_\mu+\mathbf{q})_i \geq 0\,, \tag{18}$$

$$(\mathbf{p}_\mu)_i \geq 0\,, \quad (\mathbf{Mp}_\mu+\mathbf{q})_i(\mathbf{p}_\mu)_i = \mu^2\,, \tag{19}$$

$$\mathbf{u}_\mu = \mathbf{K}^{-1}(\mathbf{f}-\mathbf{Cp}_\mu)\,, \quad i=1,\dots,n_c\,, \tag{20}$$

where $\mathbf{M}$ and $\mathbf{q}$ are the same as in (a). Because $\mathbf{M}$ is positive definite, the prerequisites of Theorem 5.9.13 of Cottle *et al.* (1992) are fulfilled, which yields that the problem (18)–(19) has an unique solution $\mathbf{p}_\mu$ for all μ. The existence and uniqueness of $\mathbf{u}_\mu$ follows from that of $\mathbf{p}_\mu$ together with (20). q.e.d.

(c) Matrix $\mathbf{M}$ in (15)–(16) is positive definite, hence it is a $\mathbf{P}$-matrix. From Proposition 5.10.5 in Cottle *et al.* (1992) applied to (15)–(16) it then follows that

$$\|\mathbf{p}-\mathbf{p}_\mu\|_\infty \leq \frac{1+\|\mathbf{M}\|_\infty}{c(\mathbf{M})}\|\min(\mathbf{p}_\mu, \mathbf{Mp}_\mu+\mathbf{q})\|_\infty, \tag{21}$$

where [Proposition 5.10.10 by Cottle *et al.* (1992)]

$$c(\mathbf{M}) \geq \lambda(\mathbf{M})/n_c. \tag{22}$$

Here $\lambda(\mathbf{M})$ is the smallest eigenvalue of $\mathbf{M}$. As $\mathbf{p}_\mu$ fulfills (18)–(19) it holds that

$$\|\min(\mathbf{p}_\mu, \mathbf{Mp}_\mu+\mathbf{q})\|_\infty \leq \mu. \tag{23}$$

Combining (21)–(23) results in:

$$\|\mathbf{p}-\mathbf{p}_\mu\|_\infty \leq \frac{(1+\|\mathbf{M}\|_\infty)n_c}{\lambda(\mathbf{M})}\mu. \tag{24}$$

Since S is compact and $\|\mathbf{M}\|_\infty$ and $\lambda(\mathbf{M})$ are continuous on S, it holds that there exists a constant D, which is independent of μ and $\mathbf{s}\in S$, such that

$$\|\mathbf{p}-\mathbf{p}_\mu\|_\infty \leq D\mu. \tag{25}$$

Relation (25) implies that $\mathbf{p}_\mu$ converges uniformly to $\mathbf{p}$ on S as $\mu \to 0$. That $\mathbf{u}_\mu$ converges uniformly to $\mathbf{u}$ follows from (17) and (20) together with the uniform convergence of $\mathbf{p}_\mu$ to $\mathbf{p}$. q.e.d.

(d) Due to (ii)–(iii) and (10), (7)–(8) are the necessary and sufficient Karush-Kuhn-Tucker conditions for the solution of the following optimization problem, cf. Bazaraa *et al.* (1993),

$$\min_{\mathbf{u}_\mu} \frac{1}{2}\mathbf{u}_\mu^T\mathbf{K}\mathbf{u}_\mu - \mathbf{f}^T\mathbf{u}_\mu - \mu^2\sum_{i=1}^{n_c}\log((\mathbf{g}-\mathbf{Cu}_\mu)_i)\,. \tag{26}$$

Further, due to (iv), assumptions (ii) and (iii) must hold also in the closure of some open set S' containing S. Thus, $\mathbf{u}_\mu$ is well-defined on the closure of S'.

Corollary 3.2.5 of Fiacco (1983) (specialized to the case of an optimization problem without constraints) applied to (26) yields that $\mathbf{u}_\mu$ is i times continuously differentiable on S'.

Now define $\mathbf{p}_\mu$ on S' from $\mathbf{u}_\mu$ through the following equation:

$$\mathbf{C}\mathbf{C}^T\mathbf{p}_\mu = \mathbf{C}(\mathbf{f} - \mathbf{K}\mathbf{u}_\mu). \tag{27}$$

Here $\mathbf{C}\mathbf{C}^T$ is invertible on the closure of S', because $\mathbf{C}$ has full row rank on the closure of S'. By inserting (27) into (7)–(8) it can be checked that the $\mathbf{p}_\mu$ defined by (27) indeed is the unique solution of (7)–(8) on (the closure of) S'. By applying standard implicit function theorems onto (27), the differentiability of $\mathbf{p}_\mu$ on S' follows from (iv) and the differentiability of $\mathbf{u}_\mu$. q.e.d.

(e) From (c) and (d), there are series of continuous functions $\mathbf{u}_\mu$ and $\mathbf{p}_\mu$ that converge uniformly to $\mathbf{u}$ and $\mathbf{p}$ respectively. By Theorem 7.12 of (Rudin 1976) it follows that $\mathbf{u}$ and $\mathbf{p}$ are continuous on S. q.e.d.

(f) From (d) and (e): $\mathbf{u}, \mathbf{p}, \mathbf{u}_\mu$, and $\mathbf{p}_\mu$ (for all μ) are continuous on S. Thus both (P) and (P_μ) are the problem of finding a point for which a continuous function, c, attains its smallest value on a nonempty compact (metric) set, S. It then follows from Theorem 4.16 (Rudin 1976), that they have a solution. q.e.d.

(g) For the sake of contradiction; assume that (g) is not true, then there is an infinite series $\{\mathbf{s}^*_{\mu_k}\}$ of solutions to (P_{μ_k}), where $\mu_k \neq 0$ and $\mu_k \to 0$, such that the smallest distance (in the ∞-norm) to the solution set of (P) is greater than some $\epsilon > 0$ for all k.

As $\{\mathbf{s}^*_{\mu_k}\}$ is contained in the compact, (i), set S there is a convergent sub-sequence of $\{\mathbf{s}^*_{\mu_k}\}$. This convergent sub-sequence is for simplicity denoted $\{\mathbf{s}^*_{\mu_k}\}$ and its limiting point $\mathbf{s}^*_{\mu_\infty}$. (Let $\bar{c}(\mathbf{s}) = c[\mathbf{u}(\mathbf{s}), \mathbf{p}(\mathbf{s}), \mathbf{s}]$, $\bar{c}_\mu(\mathbf{s}) = c(\mathbf{u}_\mu(\mathbf{s}), \mathbf{p}_\mu(\mathbf{s}), \mathbf{s})$, and $\mathbf{s}^*$ be a solution of (P).) As $\mathbf{s}^*_{\mu_\infty}$ is the limit point of $\{\mathbf{s}^*_{\mu_k}\}$, the smallest distance between $\mathbf{s}^*_{\mu_\infty}$ and the solution set of (P) is greater than ϵ, i.e. $\mathbf{s}^*_{\mu_\infty}$ is not a solution of (P). Hence there exist an $\eta > 0$ such that $\bar{c}(\mathbf{s}^*_{\mu_\infty}) \geq \bar{c}(\mathbf{s}^*) + \eta$.

Due to the uniform convergence, (c), of $\mathbf{u}_\mu$ and $\mathbf{p}_\mu$ and the continuouty of c, (v), it holds that

$$\lim_{k\to+\infty} \bar{c}_{\mu_k}(\mathbf{s}^*_{\mu_k}) = \bar{c}(\mathbf{s}^*_{\mu_\infty}) \geq \bar{c}(\mathbf{s}^*) + \eta\,.$$

It further holds that $\mathbf{s}^*$ is feasible to (P_μ) for all μ. Consider the series $\{\bar{c}_{\mu_k}(\mathbf{s}^*)\}$. Due to (c),

$$\lim_{k\to+\infty} \bar{c}_{\mu_k}(\mathbf{s}^*) = \bar{c}(\mathbf{s}^*)\,.$$

This implies that for large enough k: $\bar{c}_{\mu_k}(\mathbf{s}^*_{\mu_k}) > \bar{c}_{\mu_k}(\mathbf{s}^*)$, i.e. $\mathbf{s}^*_{\mu_k}$ is not a solution of (P_{μ_k}), but this is a contradiction because it was assumed that $\mathbf{s}^*_{\mu_k}$ is a solution of (P_{μ_k}). Thus (g) holds. q.e.d. $\square$