On solution of contact shape optimization problem by proximal bundle method

BEREMLIJSKI Petr, PLÍVOVÁ Nikola

VŠB-TU Ostrava, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, CZ
E-mail: petr.beremlijski@vsb.cz

Abstract: From the mathematical point of view, the contact shape optimization is a problem of nonlinear (usually nonsmooth) optimization with a specific structure which can be exploited in its solution. In this paper, we show how to overcome the difficulties related to the nonsmooth cost function by using the proximal bundle method. To illustrate the performance of the presented algorithm, we solve a shape optimization problem associated with the discretized two-dimensional contact problem with Coulomb's friction.

Keywords: nonsmooth optimization, Clarke calculus, proximal bundle method, shape optimization.

1 Introduction

Shape optimization problems arise naturally in mechanical engineering whenever the design requirements include an optimal performance of a machine comprising several bodies in mutual contact. From the mathematical point of view, these problems can be characterized by a locally Lipschitz continuous cost function which is differentiable in most but not all points. Shape optimization problems have the following form:

$$\left.\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \subset \mathbb{R}^n. \end{array}\right\} \tag{1}$$

The solution of such problems can be obtained by a suitable iterative algorithm – its typical structure reads as in Tab 1. The hardest difficulty is the direction searching in Step 2 since the cost function f is not differentiable but only locally Lipschitz continuous. This implies that to minimize the function f, we can choose an algorithm from the following two classes: derivative-free methods (like genetic algorithms) and methods that use the subgradient information (like subgradient or bundle methods). Since the subgradient information is available for our problem, we have chosen the latter class of algorithms. In this paper, the proximal bundle method (see [4] or [6]) is presented. This method needs the function value $f(x)$ and one (arbitrary) Clarke subgradient of f at x in every step of the iteration process.

Table 1: Basic iterative algorithm.

Step 0: (Initialization)
Find a feasible starting point $x_1 \in \Omega$ and set $k = 1$.

Step 1: (Stopping criterion)
If x_k is "close enough" to the required solution then STOP.

Step 2: (Direction finding)
Find a feasible descent direction $d_k \in \mathbb{R}^n$:
$f(x_k + td_k) < f(x_k)$ and $x_k + td_k \in \Omega$ for some $t > 0$.

Step 3: (Line search)
Find a step size $t_k > 0$ such that
$t_k \approx \arg\min_{t>0}\{f(x_k + td_k)\}$ and $x_k + td_k \in \Omega$.

Step 4: (Updating)
Set $x_{k+1} = x_k + t_k d_k, k = k + 1$ and go on to Step 1.

Main definitions are introduced in the beginning of this article and then the following section presents description of the proximal bundle method. To show the functionality of the presented algorithm, we solve a shape optimization problem with the discretized two-dimensional contact problem with Coulomb's friction in the last part.

2 Nonsmooth analysis - calculus of Clarke

We start this section with definition of Lipschitz continuity and generalized gradient.

Definition 1 *A function $f\colon \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is said to be Lipschitz continuous on Ω if there exists some constant $L = L(\Omega) > 0$ such that*

$$|f(x) - f(y)| \leq L\|x - y\|, \quad \forall x, y \in \Omega. \tag{2}$$

Definition 2 *A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be Lipschitz continuous at $x \in \mathbb{R}^n$ if there exists a neighbourhood U of x and a constant $L = L(U) > 0$ such that*

$$|f(x) - f(y)| \leq L\|x - y\|, \quad \forall y \in U. \tag{3}$$

Definition 3 *A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be locally Lipschitz continuous in $\mathbb{R}^n$ if this function f is Lipschitz continuous at $x \in \mathbb{R}^n$ for every $x \in \mathbb{R}^n$.*

Definition 4 *Let $\Omega \subset \mathbb{R}^n$. Then* conv (Ω) *denotes the convex hull of the set Ω, which is defined by*

$$\text{conv}\,(\Omega) = \left\{ \sum_{i=1}^{n} \lambda_i x_i \;\middle|\; n \in \mathbb{N}, \lambda \in \mathbb{R}^n, x_1, \cdots, x_n \in \Omega, \lambda_i \geq 0, \forall i, \sum_{i=1}^{n} \lambda_i = 1 \right\}. \tag{4}$$

Definition 5 *Let the objective function $f\colon \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous (in $\mathbb{R}^n$). The generalized gradient of the objective function f at $x \in \mathbb{R}^n$ is the set*

$$\partial f(x) = \text{conv}\left\{ g \in \mathbb{R}^n \;\middle|\; g = \lim_{i\to\infty} \nabla f(x_i), x_i \to x, x_i \notin \Omega_f \right\}, \tag{5}$$

where $\Omega_f = \{x \in \mathbb{R}^n, f$ is not differentiable in $x\}$. Each element $g \in \partial f(x)$ is called a subgradient of the objective function f at x.

We now illustrate the previous definition Def. 5. Let us consider the function

$$f(x) = |x-1| \ + \ |x| \ + \ |x+1|.$$

Figure 1 shows the graph of the function f (left) and graph of its general gradient.

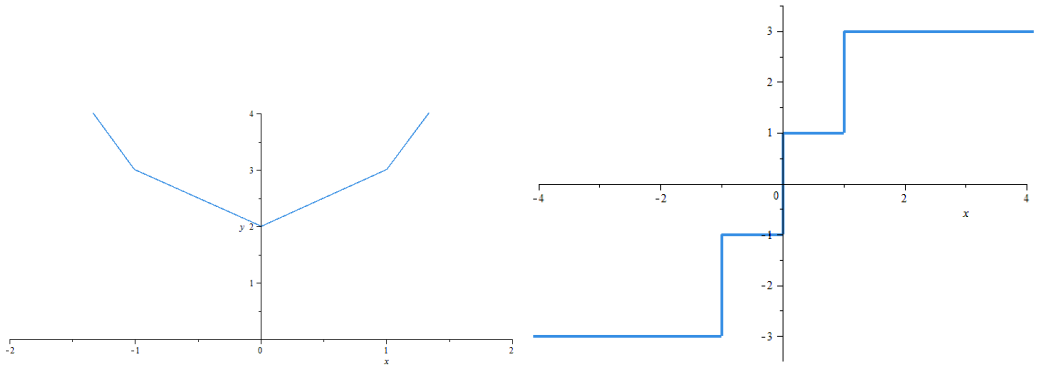

Figure 1: Graph of function f (left) and graph of its general gradient (right).

3 Description of the proximal bundle method

Consider the following nonlinear constrained optimization problem

$$\left.\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Cx \leq b, \\ & x_{\min} \leq x \leq x_{\max}, \end{array}\right\} \tag{6}$$

where the objective function $f\colon \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous in $\mathbb{R}^n$, $C \in \mathbb{R}^{m\times n}$ is an constraint matrix, $b \in \mathbb{R}^m$ is a right-hand side vector and $x_{\max} \in \mathbb{R}^n$, $x_{\min} \in \mathbb{R}^n$ are bound vectors. To make these notations simple we suppose that the simple bounds $x_{\min}, x_{\max}$ are included in the linear system $Cx \leq b$. For further details on the proximal bundle method the interested reader is reffered to [4].

3.1 Direction finding

Our aim is to solve the problem with respect to $d \in \mathbb{R}^n$

$$\left.\begin{array}{ll} \text{minimize} & f(x_k + d) - f(x_k), \\ \text{subject to} & x_k + d \in \Omega, \end{array}\right\} \tag{7}$$

where $\Omega = \{x \in \mathbb{R}^n \mid Cx \leq b\}$ and d is the descent direction .

Suppose that we have some starting point $x_1 \in \Omega$, the current iteration point $x_k \in \Omega$ and that we have subgradients $g_j^f \in \partial f(y_j)$ for all $j \in J_f^k$, where $J_f^k \subset \{1, \ldots, k\}$ is a nonempty index set and where $y_j \in \Omega$ is an auxiliary point. Denoting

$$f_j^k = f\left(y_j\right) + \left(g_j^f\right)^T \left(x_k - y_j\right), \tag{8}$$

the linearization of our cost function is

$$\overline{f}_j(x) = f_j^k + \left(g_j^f\right)^T (x - x_k) \text{ for all } j \in J_f^k, \tag{9}$$

but we can rewrite the formulation (8) into its recursive form

$$f_j^{k+1} = f_j^k + \left(g_j^f\right)^T (x_{k+1} - x_k) \text{ for all } j \in J_f^k. \tag{10}$$

Moreover, we can employ this linearization for polyhedral approximation of the objective function (e.g. in Fig. 2)

$$\hat{f}^k(x) = \max\{\, \overline{f}_j(x) \mid j \in J_f^k \,\} \tag{11}$$

and then we can define the improved polyhedral function $\hat{H}^k$

$$\hat{H}^k(x) = \hat{f}^k(x) - f(x_k) \text{ for all } x \in \mathbb{R}^n. \tag{12}$$

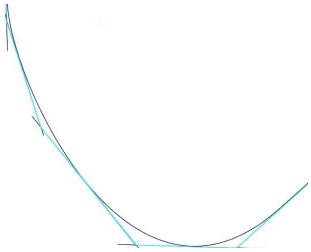

Figure 2: Illustration of the linearization.

By employing the proximal bundle idea [1] and after a series of adjustments, we can rewrite the whole problem (7) into its dual form

$$\left.\begin{array}{ll} \min\limits_{\lambda,\nu} & \frac{1}{2u_k}\| \sum\limits_{j\in J_f^k} \lambda_j g_j^k + \sum\limits_{i\in I} \nu_i C_i\|^2 + \sum\limits_{j\in J_f^k} \lambda_j \alpha_{f,j}^k + \sum\limits_{i\in I} \nu_i \alpha_{C,i}^k \\ \text{subject to} & \sum\limits_{j\in J_f^k} \lambda_j = 1 \text{ and } \lambda, \nu \geq 0, \end{array}\right\} \tag{13}$$

where u_k is the weight, $\alpha_{f,j}^k$ are subgradient errors $\left(\alpha_{f,j}^k = f(x_k) - f_j^k, \text{ for } j \in J_f^k\right)$ and $\alpha_{C,i}^k$ are errors of the constraints subgradients $\left(\alpha_{C,i}^k = -C_i x_k + b_i, \text{ for } i \in I = \{1, \ldots, m\}\right)$. We denote the solution of the problem (13) as vector (λ^k, ν^k). The descent direction d_k is given as

$$d_k = -\frac{1}{u_k}\left(\sum_{j\in J_f^k} \lambda_j^k g_j^k + \sum_{i\in I} \nu_i^k C_i\right) \tag{14}$$

and the awaited decrease v_k can be computed as

$$v_k = -\frac{1}{u_k}\left\|\sum_{j\in J_f^k} \lambda_j^k g_j^k\right\|^2 - \sum_{j\in J_f^k} \lambda_j^k \alpha_{f,j}^k - \sum_{i\in I} \nu_i^k C_i < 0. \tag{15}$$

[1] The idea of adding a penalty to be able to limit the step length.

3.2 Subgradient aggregation

There is still one hidden but equally important difficulty in the problem (13). Let us consider the index set J_f^k. The simplest way to choose this set seems to let

$$J_f^k = \{1, \ldots, k\}\,. \tag{16}$$

However, this is not the right idea. Because, in every iteration step, the index set will enlarge which causes larger and larger memory requirements. In 1985, Kiwiel presented the subgradient aggregation strategy. The idea is to aggregate the constraints made up by the previous subgradients. This strategy allows us to keep the quantity of constraints bounded. We denote the aggregate subgradient by p_f^k. For more details see [4].

3.3 Nonconvexity

Let us recall that $\alpha_{f,j}^k = f(x_k) - f_j^k$ is the linearization error. If f is convex, then $\alpha_{f,j}^k \geq 0$ for all $j \in J_f^k$ and $\overline{f}_j(x) \leq f(x)$ for all $x \in \Omega$. It means that our linearization approximates the cost function f from bellow and $\alpha_{f,j}^k$ indicates how good our linearization is. But this is true only if the cost function f is convex. Unfortunately, in the nonconvex case, the inequality $\overline{f}_j(x) \leq f(x)$ is not valid at every $x \in \Omega$. The linear approximation can be above the cost function f and the linearization error may takes values less then zero (see Fig. 3).

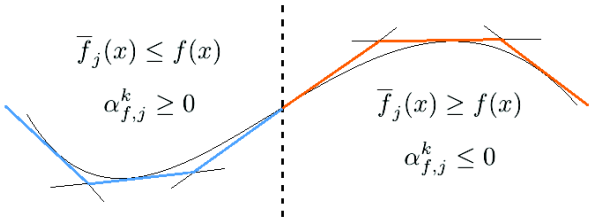

Figure 3: Linear approximation of a nonconvex function.

We have to generalize the subgradient error $\alpha_{f,j}^k$. To achieve this, we will need some information about the distance between the trial point y_j and the actual iteration point x_k.

Definition 6 *Let us define the distance measure at every iteration k by*

$$s_j^k = \begin{cases} \|x_j - y_j\| + \sum\limits_{i=j}^{k-1} \|x_{i+1} - x_i\| & \text{for } j = 1, \ldots, k-1 \\ \|x_k - y_k\| & \text{for } j = k \end{cases} \tag{17}$$

And now we are able to define the subgradient locality measure.

Definition 7 *At every iteration step k, the subgradient locality measure is defined by*

$$\beta_j^k = \max\left\{|\alpha_{f,j}^k|, \gamma\left(s_j^k\right)^2\right\} \text{ for all } j \in J_f^k, \tag{18}$$

where $\gamma \geq 0$ is the distance measure parameter which is equal to zero, when the cost function is convex.

We choose the parameter γ heuristically. We denote the aggregate subgradient locality measure by $\tilde{\beta}_{f,p}^k$.

3.4 Line search

The descent direction d_k is known. But we do not know yet how far we can go in the direction d_k to evaluate the next value x_{k+1}. A solution to this problem was presented by Kiwiel in 1990 in his contribution [3].

3.5 Weight update

One of the last but still very important question is the choice of weight update u_k. We cannot keep u_k constant, because it could make some difficulties (e.g. if the parameter u_k is large, values $|v_k|$ and $\|d_k\|$ will be very small and therefore the decrease will be small). This difficulty was also solved by Kiwiel in 1990. The whole weight update strategy can be found in the book [4] and in the article [3].

3.6 Several conclusion notes about proximal bundle method algorithm

At the beginning of our algorithm, we need to set several parameters such as stopping tolerance $\varepsilon_S > 0$, which is used in the stopping criterion, the maximum number of stored subgradients $M_g \geq 2$ and distance parameter $\gamma > 0$.

In the next step of the algorithm, we should find multipliers λ_j^k by solving the dual problem (13). In the algorithm, there is also implemented the stopping criterion. We need to evaluate whether $w_k \leq \varepsilon_S$, where $w_k = \frac{1}{2}\|p_f^k\|^2 + \tilde{\beta}_{f,p}^k$, holds or not. If so, the algorithm stops and we obtain the desired result. Otherwise the algorithm continues by line search and after finding the step size, we make the linearization update.

The final part of the algorithm consists of the weight update and the index set updating $J_f^{k+1} = J_f^k \cup \{k+1\}$, but if the size of $J_f^{k+1} > M_g$, we set $J_f^{k+1} = J_f^{k+1} \setminus \left\{\min\ j \ \middle|\ j \in J_f^{k+1}\right\}$. Now it remains to increase the iteration counter k by 1 and to repeat the whole algorithm from the part with the dual problem.

4 Numerical experiment

The proximal bundle method described in the previous section will now be used to solve a model example. We chose the shape optimization of a discretized two-dimensional contact problem with Coulomb friction as the model example. Shape optimization is a part of the optimal control in which the control variables are linked to the geometry of elastic bodies that are in contact. The aim of the problem on the lower level which is contact problem with friction is to find the set of the state variables for the fixed vector of control variables. The state vector contains variables which describe the displacements and the normal stress on the contact boundary. Hereafter the contact problem with Coulomb friction will be considered as the state problem. The mapping describing the solution of the state problem for the prescribed control variable is named as the control–state mapping. A typical feature of the contact shape optimization with Coulomb friction is its nonsmooth character due to the fact that the respective control–state mapping is typically nondifferentiable. Shape optimization of a discretized 2D contact problem with Coulomb friction was considered in [1]. Shape optimization of a discretized 3D contact problem with Coulomb friction was considered in [2]. Sensitivity analysis (computation of the subgradients of the minimized function) with help of calculus of Clarke (for 2D case) and calculus of Mordukhovich (for 3D case) was proposed in [1], [2]. In this contribution, we approximate subgradients only numerically by the forward finite difference approximation.

Example 1 *Now let us deal with the shape optimization of a discretized two-dimensional contact problem with Coulomb friction only briefly. Let $\mathcal{J}$ be a cost function. The shape optimization problem is defined generally as follows*

$$\left.\begin{array}{l} \textit{minimize } \mathcal{J}(\alpha, \mathcal{S}(\alpha)) \\ \textit{subject to } \alpha \in U_{ad}, \end{array}\right\} \tag{19}$$

where the admissible set U_{ad} is given by

$$\begin{aligned} U_{ad} := \; & \big\{\alpha \in \mathbb{R}^d :\; 0 \le \alpha^i \le C_0,\; i = 0, 1, \ldots, d-1;\; |\alpha^{i+1} - \alpha^i| \le C_1 h,\; i = 0, 1, \ldots, d-2; \\ & C_{21} \le \text{meas } \Omega(\alpha) \le C_{22}\big\}. \end{aligned}$$

We will try to smooth down the peaks of the normal contact stress distribution. To this aim, we should minimize the max-norm of the discrete normal contact stress λ. The objective function $\mathcal{J}$, however, must be continuously differentiable to ensure that the composite function $\mathcal{J}(\alpha, \mathcal{S}(\alpha))$ is locally Lipschitz, so we will use the p power of the p norm of the vector λ with $p = 4$ as the objective function $\mathcal{J}$. The shape optimization problem then reads as follows:

$$\left.\begin{array}{l} \textit{minimize } \|\lambda\|_4^4 \\ \textit{subject to } \alpha \in U_{ad}. \end{array}\right\} \tag{20}$$

The vector α denotes the control vector, u denotes the displacement and λ denotes the normal stress and mapping $\mathcal{S}\colon \alpha \in U_{ad} \subset \mathbb{R}^d \to (u, \lambda) \in \mathbb{R}^{3p}$ denotes the control–state mapping. Number d is the dimension of the control vector α, p is the number of the nodes of the discretized elastic body $\Omega(\alpha)$ and U_{ad} is the set of the admissible control variables. For more detailed description, see [1].

The shape of the elastic body $\Omega(\alpha)$, $\alpha \in U_{ad}$, is defined through a Bézier function F_α as follows (cf. Fig. 4):

$$\Omega(\alpha) = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \in (0, a),\; F_\alpha(x_1) < x_2 < b\},$$

where the vector α contains the control points of the Bézier function F_α.

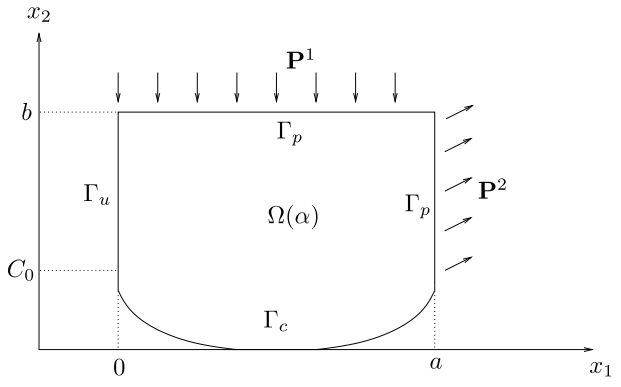

Figure 4: The elastic body and applied loads.

From Fig. 4 we can also see the distribution of external pressures on the boundary Γ_P, given as $P^1 = (0; -200\ MPa)$ on $(0, a) \times \{b\}$, while $P^2 = (100\ MPa; 40\ MPa)$ on $\{a\} \times (0, b)$ Further, Γ_u is the part of the boundary where the zero displacements are prescribed.

The set of the admissible designs U_{ad} and the elastic body $\Omega(\alpha)$ is specified as follows: $a = 2$, $b = 1$ and $C_0 = 0.75$, $C_1 = 1$, $C_{21} = 1.8$, $C_{22} = 2$. The Young modulus $E = 1$ GPa and the Poisson constant $\sigma = 0.3$ are used for the definition of the mapping $\mathcal{S}$. The value of the coefficient of the Coulomb friction is 0.25. The state problem on $\Omega(\alpha)$ is discretized by isoparametric quadrilateral elements of Lagrange type. The total number of nodes (vertices of quadrilaterals) is 3976 for any $\alpha \in U_{ad}$. The dimension of the control vector α, generating the Bézier function and defining $\Omega(\alpha)$, is $d = 8$.

The stopping tolerance was set to $\varepsilon_S = 1 \cdot 10^{-6}$. This required precision was reached after 11 iterations. We depict the initial shape and the distribution of the von Mises stress in the loaded initial body in Fig. 5 . Figure 6 shows the optimal shape and the von Mises stress in the deformed optimal body. Finally, figure Fig. 7 compare the contact normal stresses for the initial and optimal shape, respectively. Note that during the optimization process the initial value $\mathcal{J}(\alpha_0) = 2.8612 \cdot 10^{11}$ of the cost functional dropped to $\mathcal{J}(\alpha_{opt}) = 1.0695 \cdot 10^{11}$. The decrease of the peak stress is also quite significant. The experiment was carried out in Mathworks Matlab.

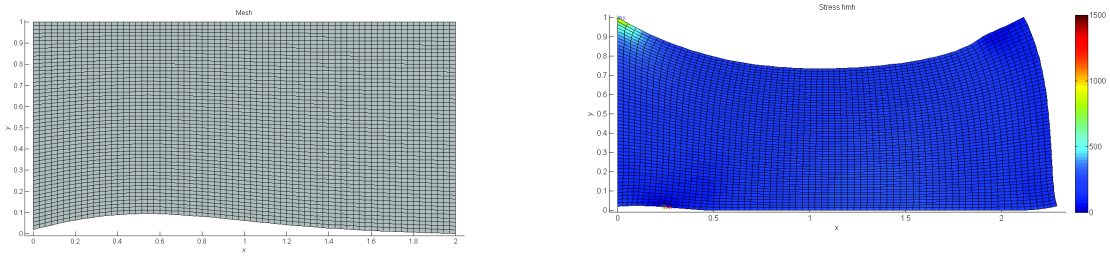

Figure 5: Example, initial design – the initial shape of the body (left) and the distribution of the von Mises stress in the deformed initial body (right).

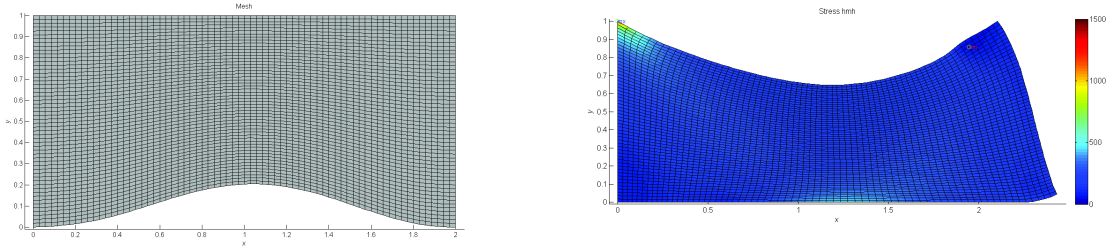

Figure 6: Example, optimal design – the optimal shape of the body (left) and the distribution of the von Mises stress in the deformed optimal body (right).

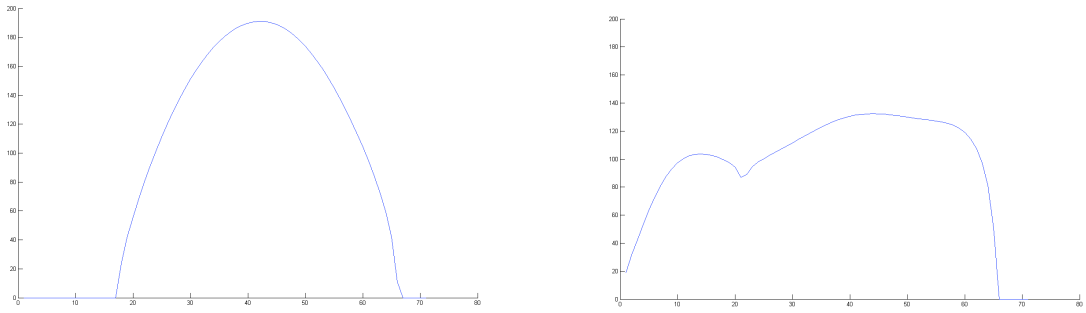

Figure 7: Example, normal stress for initial (left) and optimal (right) design.

Conclusion

In this contribution we have briefly introduced the proximal bundle method for nonsmooth optimization problems with linear constraints and with simple bounds. We outlined the implemented algorithm, which was employed to solve our model example. Then we tried to deal with the shape optimization of a discretized 2D contact problem with Coulomb friction.
The work is partially supported by Grant of SGS No. SP2017/122, VŠB–Technical University of Ostrava, Czech Republic.

References

[1] BEREMLIJSKI, P., J. HASLINGER, M. KOČVARA and J. OUTRATA. Shape optimization in contact problems with Coulomb friction. *SIAM Journal on Optimization*. 2002, vol. 13, iss. 2, pp. 561–587. ISSN 0233-1934.

[2] BEREMLIJSKI, P., J. HASLINGER, M. KOČVARA, R. KUČERA and J. OUTRATA. Shape optimization in three-dimensional contact problems with Coulomb friction. *SIAM Journal on Optimization*. 2009, vol. 20, iss. 1, pp. 416–444. ISSN 1052-6234.

[3] KIWIEL, K.C. Proximity Control in Bundle Methods for Convex Nondifferentiable Minimization. *Mathematical Programming*. 1990, vol. 46, iss. 1, pp. 105–122. ISSN 0025-5610.

[4] MÄKELÄ, M. M. and P. NEITTAANMÄKI. *Nonsmooth Optimization, Analysis and Algorithms with Application to Optimal Control*. Singapore: World Scientific, 1992. ISBN 9810207735.

[5] OUTRATA, J., M. KOČVARA and J. ZOWE. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints, Theory, Applications and Numerical Results*. Berlin: Springer Science+Business Media, B.V., 1998. ISBN 978-1-4419-4804-5.

[6] PLÍVOVÁ, N. and P. BEREMLIJSKI. Proximal Bundle Method for Contact Shape Optimization Problem. *Advances in Electrical and Electronic Engineering*. 2017, vol. 15, iss. 2, pp. 242–250.

[7] ZOWE, J. *Nondifferentiable Optimization*. University of Bayreuth, 1985.

Řešení úlohy tvarové optimalizace pro kontaktní úlohu pomocí proximal bundle metody

Abstrakt: Úlohu tvarové optimalizace pro kontaktní úlohu můžeme popsat jako úlohu nelineární optimalizace. Velmi často jde o úlohu nehladké optimalizace. V tomto příspěvku si ukážeme, jak minimalizovat cenovou funkci, která je nediferencovatelná. K tomu použijeme proximal bundle metodu. V příspěvku popíšeme postup minimalizace nediferencovatelné funkce, včetně linearizace, hledání směru poklesu, výpočtu délky kroku a návrhu ukončující podmínky. Abychom ukázali efektivitu této metody, použijeme ji pro řešení úlohy tvarové optimalizace pro 2D kontaktní úlohu s Coulombovým třením.

Klíčová slova: nehladká optimalizace, Clarkeův kalkul, proximal bundle metoda, tvarová optimalizace.