



HAL
open science

Nonconvex bundle method with application to a delamination problem

Minh Dao, Joachim Gwinner, Dominikus Noll, Nina Ovcharova

► **To cite this version:**

Minh Dao, Joachim Gwinner, Dominikus Noll, Nina Ovcharova. Nonconvex bundle method with application to a delamination problem. *Computational Optimization and Applications*, 2016, 65 (1), pp.173 - 203. 10.1007/s10589-016-9834-0 . hal-01868450

HAL Id: hal-01868450

<https://hal.science/hal-01868450v1>

Submitted on 22 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NONCONVEX BUNDLE METHOD WITH APPLICATION TO A DELAMINATION PROBLEM

M.N. Dao[†], J. Gwinner^{*}, D. Noll[†], and N. Ovcharova^{*}

ABSTRACT. Delamination is a typical failure mode of composite materials caused by weak bonding. It arises when a crack initiates and propagates under a destructive loading. Given the physical law characterizing the properties of the interlayer adhesive between the bonded bodies, we consider the problem of computing the propagation of the crack front and the stress field along the contact boundary. This leads to a hemivariational inequality, which after discretization by finite elements we solve by a nonconvex bundle method, where upper- C^1 criteria have to be minimized. As this is in contrast with other classes of mechanical problems with non-monotone friction laws and in other applied fields, where criteria are typically lower- C^1 , we propose a bundle method suited for both types of nonsmoothness. We prove its global convergence in the sense of subsequences and test it on a typical delamination problem of material sciences.

Key words. Composite material · delamination · crack front propagation · hemivariational inequality · Clarke directional derivative · nonconvex bundle method · lower- and upper- C^1 function · convergence.

1. INTRODUCTION

We develop a bundle technique to solve nonconvex variational problems arising in contact mechanics and in other applied fields. We are specifically interested in the delamination of composite structures with an adhesive bonding under destructive loading, a failure mode which is studied in the material sciences. When the properties of the interlayer adhesive between the bonded bodies are given in the form of a physical law relating the normal component of the stress vector to the relative displacement between the upper and lower boundaries at the crack tip, the challenge is to compute the displacement and stress fields in order to assess the reactive destructive forces along the contact boundary, as the latter are difficult to measure in situ. This leads to minimization of an energy functional, where a specific form of nonsmoothness arises in the boundary integral at the contact boundary. After discretization via piecewise linear finite elements using the trapezoidal quadrature rule, this leads to a finite-dimensional nonsmooth optimization problem of the form

$$(1) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b \end{array}$$

where f is locally Lipschitz and neither smooth nor convex. Depending on the nature of the frictional forces, the criterion f may be upper- C^1 or lower- C^1 , see e.g. Figure 1. As these two classes of nonsmooth functions behave substantially differently when minimized, we are forced to expand on existing bundle strategies and develop an algorithm general enough to encompass both types of nonsmoothness. We prove its convergence to a critical point in the sense of subsequences, and show that it provides satisfactory numerical results

[†] Institut de Mathématiques, Université de Toulouse, France.

^{*} Institute of Mathematics, Department of Aerospace Engineering, Universität der Bundeswehr München, Germany.

in a simulation of the double cantilever beam test [1], one of the most popular destructive tests to qualify structural adhesive joints.

The difficulty in nonconvex bundling is to provide a suitable cutting plane oracle which replaces the no longer available convex tangent plane. One of the oldest oracles, discussed already in Mifflin [2], and used in the bundle codes of Lemaréchal and Sagastizábal [3, 4], or the BT-codes of Zowe [5, 6], uses the method of *downshifted tangents*. While these authors use linesearch with Armijo and Wolfe type conditions, which allows only weak convergence certificates in the sense that *some* accumulation point of the sequence of serious iterates is critical, we favor proximity control in tandem with a suitable backtracking strategy. This leads to stronger convergence certificates, where *every* accumulation point of the sequence of serious iterates is critical. For instance, in [7, 8, 9] a strong certificate for downshifted tangents with proximity control was proved within the class of lower- C^1 functions, but its validity for upper- C^1 criteria remained open. An oracle for upper- C^1 functions with a rigorous convergence theory can be based on the *model approach* of [7, 8, 10], but the latter is not compatible with the downshift oracle.

To have two strings to one bow is unsatisfactory, as one could hardly expect practitioners to select their strategy according to such a distinction, which might not be easy to make in practice. In this work we will resolve this impasse and present a cutting plane oracle based on downshifted tangents, which leads to a bundle method with strong convergence certificate for both types of nonsmoothness. In its principal components our method agrees with existing strategies for downshifted tangents, like [3, 5, 11, 12], and could therefore be considered as a justification of this technique for a wide class of applications. Differences with existing methods occur in the management of the proximity control parameter, which in our approach has to respect certain rules to assure convergence to a critical point, without impeding good practical performance.

The structure of the paper is as follows. Section 2 gives some preparatory information on lower- and upper- C^1 functions. Section 4 presents the algorithm and comments on its ingredients. Theoretical tools needed to prove convergence are presented and employed in sections 3 and 5. Section 6 gives the main convergence result, while section 7 discusses practical aspects of the algorithm. In section 8, we discuss the delamination problem, which we solve numerically using our bundle algorithm.

Numerical results for contact problem with adhesion based on the bundle-Newton method of L. Lukšan and J. Vlček [13] can be found e.g. in the book of Haslinger et al. [14], in [11, 12], and in the more recent [15, 16]. Mathematical analysis and numerical results for quasistatic delamination problems can be found in [17, 18].

2. LOWER- AND UPPER- C^1 FUNCTIONS

Following Spingarn [19], a locally Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower- C^1 at x_0 , if there exists a compact Hausdorff space K , a neighborhood U of x_0 , and a mapping $F : U \times K \rightarrow \mathbb{R}$ such that both F and $D_x F$ are jointly continuous and

$$(2) \quad f(x) = \max\{F(x, y) : y \in K\}$$

is satisfied for $x \in U$. The function f is upper- C^1 at x_0 if $-f$ is lower- C^1 at x_0 .

In a minimization problem (1), we expect lower- and upper- C^1 functions to behave completely differently. Minimizing a lower- C^1 function ought to lead to real difficulties, as on descending we move *into* the zone of nonsmoothness, which for lower- C^1 goes downward. In contrast, upper- C^1 functions are generally expected to be well-behaved, as intuitively on descending we move *away* from the nonsmoothness, which here goes upward. The present application shows that this argument is too simplistic. Minimization of upper- C^1 functions leads to real difficulties, which we explain subsequently. In delamination for

composite materials we encounter objective functions of the form

$$(3) \quad f(x) = f_s(x) + \int_0^1 \min_{i \in I} f_i(x, t) dt,$$

where f_s gathers the smooth part, while the integral term, due to the minimum, is responsible for the nonsmoothness.

Lemma 1. *Suppose f_s is of class C^1 and the f_i are jointly of class C^1 . Then the function (3) is upper- C^1 and can be represented in the form*

$$(4) \quad f(x) = f_s(x) + \min_{\sigma \in \Sigma} \int_0^1 f_{\sigma(t)}(x, t) dt,$$

where Σ is the set of all measurable mappings $\sigma : [0, 1] \rightarrow I$.

Proof. Let us first prove (4). For $\sigma \in \Sigma$ and fixed $x \in \mathbb{R}^n$ the function $t \mapsto f_{\sigma(t)}(x, t)$ is measurable, and since $\min_{i \in I} f_i(x, t) \leq f_{\sigma(t)}(x, t) \leq \max_{i \in I} f_i(x, t)$, it is also integrable. Hence $F(x, \sigma) = f_s(x) + \int_0^1 f_{\sigma(t)}(x, t) dt$ is well-defined, and clearly $F(x, \sigma) \geq f(x)$, so we have $\inf_{\sigma \in \Sigma} F(x, \sigma) \geq f(x)$.

To prove the reverse estimate, fix $x \in \mathbb{R}^n$ and consider the closed-valued multifunction $\Phi : [0, 1] \rightarrow 2^I$ defined by $\Phi(t) = \{i \in I : f_i(x, t) = \min_{i' \in I} f_{i'}(x, t)\}$. Since the $f_i(x, \cdot)$ are measurable and I is finite, Φ is a measurable multifunction. Choose a measurable selection σ , that is, $\sigma \in \Sigma$ satisfying $\sigma(t) \in \Phi(t)$ for every $t \in [0, 1]$. Then clearly $F(x, \sigma) = f(x)$. This proves (4).

Let us now show that f is upper- C^1 . We consider $\varphi(x, t) = \min_{i \in I} f_i(x, t)$. In view of [19] $\varphi(\cdot, t)$ is upper- C^1 and its Clarke subdifferential $\partial\varphi(\cdot, t)$ is strictly supermonotone uniformly over $t \in [0, 1]$. By Theorem 2 in [20], $\varphi(\cdot, t)$ is approximately concave uniformly over $t \in [0, 1]$. Integration with respect to $t \in [0, 1]$ then yields an approximately concave function with respect to x , which by the equivalences in [20] and [19] is upper- C^1 . \square

Note that the minimum (4) is semi-infinite even though I is finite. Minimization of (3) cannot be converted into a NLP, as would be possible in the min-max case. The representation (4) highlights the difficulty in minimizing (3). Minimizing a minimum has a disjunctive character, and due to the large size of Σ this could lead to a combinatorial situation with intrinsic difficulty.

3. THE MODEL CONCEPT

The model of a nonsmooth function was introduced in [8] and is a key element in understanding the bundle concept.

Definition 1 (Compare [8]). *A function $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is called a model of the locally Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ on the set $\Omega \subset \mathbb{R}^n$ if the following axioms are satisfied:*

- (M₁) *For every $x \in \Omega$ the function $\phi(\cdot, x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, $\phi(x, x) = f(x)$ and $\partial_1\phi(x, x) \subset \partial f(x)$.*
- (M₂) *For every $x \in \Omega$ and every $\epsilon > 0$ there exists $\delta > 0$ such that $f(y) \leq \phi(y, x) + \epsilon\|y - x\|$ for every $y \in B(x, \delta)$.*
- (M₃) *The function ϕ is jointly upper semicontinuous, i.e., $(y_j, x_j) \rightarrow (y, x)$ on $\mathbb{R}^n \times \Omega$ implies $\limsup_{j \rightarrow \infty} \phi(y_j, x_j) \leq \phi(y, x)$. \square*

We recall that every locally Lipschitz function f has the so-called *standard model*

$$\phi^\sharp(y, x) = f(x) + f^0(x, y - x),$$

where $f^0(x, d)$ is the Clarke directional derivative of f at x in direction d . The same function f may in general have several models ϕ , and following [7, 10], the standard ϕ^\sharp is the smallest one. Every model ϕ gives rise to a bundle strategy. The question is then whether this bundle strategy is successful. This depends on the following property of ϕ .

Definition 2. *A model ϕ of f on Ω is said to be strict at $x_0 \in \Omega$ if axiom (M_2) is replaced by the stronger*

$$(\widehat{M}_2) \text{ For every } \epsilon > 0 \text{ there exists } \delta > 0 \text{ such that } f(y) \leq \phi(y, x) + \epsilon\|y - x\| \text{ for all } x, y \in B(x_0, \delta).$$

We say that ϕ is a strict model on Ω , if it is strict at every $x_0 \in \Omega$. □

Remark 1. We may write axiom (M_2) in the form $f(y) \leq \phi(y, x_0) + o(\|y - x_0\|)$ for $y \rightarrow x_0$, and (\widehat{M}_2) as $f(y) \leq \phi(y, x) + o(\|y - x\|)$ for $x, y \rightarrow x_0$. Except for the fact that these concepts are one-sided, this is precisely the difference between differentiability and strict differentiability. Hence the nomenclature.

Lemma 2 (Compare [7, 10]). *Suppose f is upper- C^1 . Then its standard model ϕ^\sharp is strict, and hence every model ϕ of f is strict.* □

Remark 2. For convex f the standard model ϕ^\sharp is in general not strict, but f may be used as its own model $\phi(\cdot, x) = f$. For nonconvex f , a wide range of applications is covered by composite functions $f = g \circ F$ with g convex and F differentiable. Here the so-called natural model $\phi(y, x) = g(F(x) + F'(x)(y - x))$ can be used, because it is strict as soon as F is class C^1 . This includes lower- C^2 functions in the sense of [21], lower- $C^{1,\alpha}$ functions in the sense of [22], or amenable functions in the sense of [23], which allow representations of the form $f = g \circ F$ with F of class $C^{1,1}$.

We conclude with the remark that lower- C^1 functions also admit strict models, even though in that case the construction is more delicate. The strict model in that case cannot be exploited algorithmically, and for lower- C^1 functions we prefer the oracle concept, which will be discussed in section 5.

4. ELEMENTS OF THE ALGORITHM

In this section we briefly explain the main features of the algorithm. This concerns building the working model, computing the solution of the tangent program, checking acceptance, updating the working model after null steps, and the management of the proximity control parameter.

4.1. Working model. At the current serious iterate x the inner loop of the algorithm at counter k computes an approximation $\phi_k(\cdot, x)$ of f in a neighborhood of x , called a first-order working model. The working model is a polyhedral convex function of the form

$$(5) \quad \phi_k(\cdot, x) = \max_{(a,g) \in \mathcal{G}_k} a + g^\top(\cdot - x),$$

where \mathcal{G}_k is a finite set of affine functions $y \mapsto a + g^\top(y - x)$ satisfying $a \leq f(x)$, referred to as *planes*. The set \mathcal{G}_k is updated during the inner loop k . At each step k the following rules have to be respected when updating \mathcal{G}_k into \mathcal{G}_{k+1} :

(R_1) One or several cutting planes at the null step y^k , generated by an abstract cutting plane oracle, are added to \mathcal{G}_{k+1} .

- (R₂) The so-called aggregate plane (a^*, g^*) , which consists of convex combinations of elements of \mathcal{G}_k , is added to \mathcal{G}_{k+1} .
- (R₃) Some older planes in \mathcal{G}_k , which become obsolete through the addition of the aggregate plane, are discarded and not kept in \mathcal{G}_{k+1} .
- (R₄) Every \mathcal{G}_k contains at least one so-called exactness plane (a_0, g_0) , where exactness plane means $a_0 = f(x)$, $g_0 \in \partial f(x)$. This assures $\phi_k(x, x) = f(x)$, hence the name.
- (R₅) We have to make sure that each working model ϕ_k satisfies $\partial_1 \phi_k(x, x) \subset \partial f(x)$.

Once the first-order working model $\phi_k(\cdot, x)$ has been built, the second-order working model $\Phi_k(\cdot, x)$ is of the form

$$(6) \quad \Phi_k(\cdot, x) = \phi_k(\cdot, x) + \frac{1}{2}(\cdot - x)^\top Q(x)(\cdot - x),$$

where $Q(x) = Q(x)^\top$ is a possibly indefinite symmetric matrix, depending only on the current serious iterate x , and fixed during the inner loop k . The second-order term includes curvature information on f , if available.

4.2. Tangent program and acceptance test. Once the second-order working model (6) is formed and the proximity control parameter $\tau_{k-1} \rightarrow \tau_k$ is updated, we solve the tangent program

$$(7) \quad \begin{aligned} & \text{minimize} && \Phi_k(y, x) + \frac{\tau_k}{2} \|y - x\|^2 \\ & \text{subject to} && Ay \leq b \end{aligned}$$

Here the proximity control parameter τ_k satisfies $Q + \tau_k I \succ 0$, which assures that (7) is strictly convex and has a unique solution, y^k , called the *trial step*. The trial step is a candidate to become the new serious iterate x^+ . In order to decide whether y^k is acceptable, we compute the test

$$(8) \quad \rho_k = \frac{f(x) - f(y^k)}{f(x) - \Phi_k(y^k, x)} \stackrel{?}{\geq} \gamma,$$

where $0 < \gamma < 1$ is some fixed parameter. If $\rho_k \geq \gamma$, then $x^+ = y^k$ is accepted and called a *serious step*. In this case the inner loop ends successfully. On the other hand, if $\rho_k < \gamma$, then y^k is rejected and called a *null step*. In this case the inner loop k continues. This means we will update working model $\Phi_k(\cdot, x) \rightarrow \Phi_{k+1}(\cdot, x)$, adjust the proximity control parameter $\tau_k \rightarrow \tau_{k+1}$, and solve (7) again.

Note that the test (8) corresponds to the usual Armijo descent condition used in line-searches, or to the standard acceptance test in trust region methods.

4.3. Updating the working model via aggregation. Suppose the trial step y^k fails the acceptance test (8) and is declared a null step. Then the inner loop has to continue, and we have to improve the working model at the next sweep in order to perform better. Since the second-order part of the working model $\frac{1}{2}(\cdot - x)^\top Q(x)(\cdot - x)$ remains invariant, we will update the first-order part only.

Concerning rule (R₂), by the necessary optimality condition for (7), there exists a multiplier η^* such that

$$0 \in \partial_1 \Phi_k(y^k, x) + \tau_k(y^k - x) + A^\top \eta^*,$$

or what is the same,

$$(Q(x) + \tau_k I)(y^k - x) - A^\top \eta^* \in \partial_1 \phi_k(y^k, x).$$

Since $\phi_k(\cdot, x)$ is by construction a maximum of affine planes, we use the standard description of the convex subdifferential of a max-function. Writing $\mathcal{G}_k = \{(a_0, g_0), \dots, (a_p, g_p)\}$

for $p = \text{card}(\mathcal{G}_k) + 1$, we find non-negative multipliers $\lambda_0, \dots, \lambda_p$ summing up to 1 such that

$$(Q(x) + \tau_k I)(y^k - x) - A^\top \eta^* = \sum_{i=0}^p \lambda_i g_i,$$

and in addition, $a_i + g_i^\top(y^k - x) = \phi_k(y^k, x)$ for all $i \in \{0, \dots, p\}$ with $\lambda_i > 0$. We say that those planes which are active at y^k are *called by the aggregate plane*. In the above rule (R_3) we allow those to be removed from \mathcal{G}_k . We now define the aggregate plane as:

$$a_k^* = \sum_{i=0}^p \lambda_i a_i, \quad g_k^* = \sum_{i=0}^p \lambda_i g_i.$$

Note that by construction the aggregate plane $m_k^*(\cdot, x) = a_k^* + g_k^{*\top}(\cdot - x)$ at null step y^k satisfies $m_k^*(y^k, x) = a_k^* + g_k^{*\top}(y^k - x) = \phi_k(y^k, x)$. This construction is standard and follows the original idea in Kiwiel [24]. It assures in particular that $\Phi_{k+1}(y^k, x) \geq m_k^*(y^k, x) + \frac{1}{2}(y^k - x)^\top Q(x)(y^k - x) = \Phi_k(y^k, x)$.

4.4. Updating the working model by cutting planes and exactness planes. The crucial improvement in the first-order working model is in adding a cutting plane which cuts away the unsuccessful trial step y^k according to rule (R_1). We shall denote the cutting plane as $m_k(\cdot, x) = a_k + g^\top(\cdot - x)$. The only requirement for the time being is that $a_k \leq f(x)$, as this assures $\phi_{k+1}(x, x) \leq f(x)$. Since we also maintain at least one exactness plane of the form $m_0(\cdot, x) = f(x) + g_0^\top(\cdot - x)$ with $g_0 \in \partial f(x)$, we assure $\phi_{k+1}(x, x) = \Phi_{k+1}(x, x) = f(x)$. Later we will also have to check the validity of (R_5).

It is possible to integrate so-called *anticipated cutting planes* in the new working model \mathcal{G}_{k+1} . Here anticipated designates all planes which are not based on the rules exactness, aggregation, cutting planes. Naturally, adding such planes can not be allowed in an arbitrary way, because axioms (R_1) – (R_5) have to be respected.

Remark 3. It may be beneficial to choose a new exactness plane $m_0(\cdot, x) = f(x) + g^\top(\cdot - x)$ after each null step y , namely the one which satisfies $m_0(y, x) = f^0(x, y - x)$. If x is a point of differentiability of f , then all these exactness planes are identical anyway, so no extra work occurs. On the other hand, computing $g \in \partial f(x)$ such that $g^\top(y - x) = f^0(x, y - x)$ is usually cheap. Consider for instance eigenvalue optimization, where $f(x) = \lambda_1(F(x))$, $x \in \mathbb{R}^n$, $F: \mathbb{R}^n \rightarrow \mathbb{S}^m$, and $\lambda_1: \mathbb{S}^m \rightarrow \mathbb{R}$ is the maximum eigenvalue function of \mathbb{S}^m . Then $f^0(x, d) = \lambda_1'(X, D) = \lambda_1(Q^\top D Q)$, where $X = F(x)$, $D = F'(x)d$, and where Q is a $t \times m$ matrix whose columns form an orthogonal basis of the maximum eigenspace of X of dimension t [25]. Then $G = Q Q^\top \in \partial \lambda_1(X)$ attains $\lambda_1'(X, D)$, hence $g = F'(x)^* Q Q^\top$ attains $f'(x, d)$. Since usually $t \ll m$, the computation of g is cheap.

4.5. Management of proximity control. The central novelty of the bundle methods developed in [7, 8, 26] is the discovery that in the absence of convexity the proximity control parameter τ has to follow certain basic rules to assure convergence of the sequence x^j of serious iterates. This is in contrast with convex bundle methods, where τ could in principle be frozen once and for all. More precisely, suppose $\phi_k(\cdot, x)$ has failed and produced only a null step y^k . Having built the new model $\phi_{k+1}(\cdot, x)$, we compute the secondary test

$$(9) \quad \tilde{\rho}_k = \frac{f(x) - \Phi_{k+1}(y^k, x)}{f(x) - \Phi_k(y^k, x)} \stackrel{?}{\geq} \tilde{\gamma},$$

where $0 < \gamma < \tilde{\gamma} < 1$ is fixed. Our decision is

$$(10) \quad \tau_{k+1} = \begin{cases} 2\tau_k & \text{if } \tilde{\rho}_k \geq \tilde{\gamma} \\ \tau_k & \text{if } \tilde{\rho}_k < \tilde{\gamma} \end{cases}$$

The rationale of (9) is to decide whether improving the model by adding planes will suffice, or shorter steps have to be forced by increasing τ .

The denominator in (9) gives the model predicted progress $f(x) - \phi_k(y^k, x) = \phi_k(x, x) - \phi_k(y^k, x) > 0$ at y^k . On the other hand, the numerator $f(x) - \phi_{k+1}(y^k, x)$ gives the progress over x we would achieve at y^k , had we already known the cutting planes drawn at y^k . Due to aggregation we know that $\phi_{k+1}(y^k, x) \geq \phi_k(y^k, x)$, so that $\tilde{\rho}_k \leq 1$, but values $\tilde{\rho}_k \approx 1$ indicate that little to no progress is achieved by adding the cutting plane. In this case we decide that the τ -parameter must be increased to force smaller steps, because that reinforces the agreement between f and $\phi_{k+1}(\cdot, x)$.

In the test (10) we replace $\tilde{\rho}_k \approx 1$ by $\tilde{\rho}_k \geq \tilde{\gamma}$ for some fixed $0 < \gamma < \tilde{\gamma} < 1$. If $\tilde{\rho}_k < \tilde{\gamma}$, then the quotient is far from 1 and we decide that adding planes has still the potential to improve the situation. In that event we do not increase τ .

Let us next consider the management of τ in the outer loop. Since τ can only increase or stay fixed in the inner loop, we allow τ to decrease between serious steps $x \rightarrow x^+$, respectively, $x^j \rightarrow x^{j+1}$. This is achieved by the test

$$(11) \quad \rho_{k_j} = \frac{f(x^j) - f(x^{j+1})}{f(x^j) - \Phi_{k_j}(x^{j+1}, x^j)} \stackrel{?}{\geq} \Gamma,$$

where $0 < \gamma \leq \Gamma < 1$ is fixed. In other words, if at acceptance we have not only $\rho_{k_j} \geq \gamma$, but even $\rho_{k_j} \geq \Gamma$, then we decrease τ at the beginning of the next inner loop $j + 1$, because we may trust the model. On the other hand, if $\gamma \leq \rho_{k_j} < \Gamma$ at acceptance, then we memorize the last τ -parameter used, that is τ_{k_j} at the end of the j th inner loop.

Remark 4. We should compare our management of the proximity control parameter τ with other strategies in the literature. For instance Mäkelä *et al.* [11] consider a very different management of τ , which is motivated by the convex case.

4.6. Statement of the algorithm. We are now ready to give our formal statement of algorithm 1.

5. NONCONVEX CUTTING PLANE ORACLES

In the convex cutting plane method [27, 28] unsuccessful trial steps y^k are cut away by adding a tangent plane to f at y^k into the model. Due to convexity, the cutting plane is below f and can therefore be used to construct an approximation (5) of f . For nonconvex f , cutting planes are more difficult to construct, but several ideas have been discussed. We mention [29, 2]. In [7] we have proposed an axiomatic approach, which has the advantage that it covers the applications we are aware of, and allows a convenient convergence theory. Here we use this axiomatic approach in the convergence proof.

Definition 3 (Compare [7]). *Let f be locally Lipschitz. A cutting plane oracle for f on the set Ω is an operator \mathcal{O} which, with every pair (x, y) , x a serious iterate in Ω , $y \in \mathbb{R}^n$ a null step, associates an affine function $m_y(\cdot, x) = a + g^\top(\cdot - x)$, called the cutting plane at null step y for serious iterate x , so that the following axioms are satisfied:*

- (O₁) For $y = x$ we have $a = f(x)$ and $g \in \partial f(x)$.
- (O₂) Let $y_j \rightarrow x$. Then there exist $\epsilon_j \rightarrow 0^+$ such that $f(y_j) \leq m_{y_j}(y_j, x) + \epsilon_j \|y_j - x\|$.
- (O₃) Let $x_j \rightarrow x$ and $y_j, y_j^+ \rightarrow y$. Then there exists $z \in \mathbb{R}^n$ such that

$$\limsup_{j \rightarrow \infty} m_{y_j^+}(y_j, x_j) \leq m_z(y, x). \quad \square$$

As we shall see, these axioms are aligned with the model axioms $(M_1) - (M_3)$. Not unexpectedly, there is also a strict version of (O_2) .

Algorithm 1. Proximity control algorithm for (1).

Parameters: $0 < \gamma < \Gamma < 1$, $\gamma < \tilde{\gamma} < 1$, $0 < q < \infty$, $q < T < \infty$.

- 1: **Initialize outer loop.** Choose initial guess x^1 with $Ax^1 \leq b$ and an initial matrix $Q_1 = Q_1^\top$ with $-qI \preceq Q_1 \preceq qI$. Fix memory control parameter τ_1^\sharp such that $Q_1 + \tau_1^\sharp I \succ 0$. Put $j = 1$.
- 2: **Stopping test.** At outer loop counter j , stop if $0 \in \partial f(x^j) + A^\top \eta^*$ for some multiplier $\eta^* \geq 0$. Otherwise goto inner loop.
- 3: **Initialize inner loop.** Put inner loop counter $k = 1$ and initialize τ -parameter using the memory element, i.e., $\tau_1 = \tau_j^\sharp$. Choose initial convex working model $\phi_1(\cdot, x^j)$, possibly recycling some planes from previous sweep $j - 1$, and let $\Phi_1(\cdot, x^j) = \phi_1(\cdot, x^j) + \frac{1}{2}(\cdot - x^j)^\top Q_j(\cdot - x^j)$.
- 4: **Trial step generation.** At inner loop counter k solve tangent program

$$\min_{Ay \leq b} \Phi_k(y, x^j) + \frac{\tau_k}{2} \|y - x^j\|^2.$$

The solution is the new trial step y^k .

- 5: **Acceptance test.** Check whether

$$\rho_k = \frac{f(x^j) - f(y^k)}{f(x^j) - \Phi_k(y^k, x^j)} \geq \gamma.$$

If this is the case put $x^{j+1} = y^k$ (serious step), quit inner loop and goto step 8. If this is not the case (null step) continue inner loop with step 6.

- 6: **Update working model.** Build new convex working model $\phi_{k+1}(\cdot, x^j)$ based on null step y^k by adding an exactness plane $m_k^\sharp(\cdot, x^j)$ satisfying $m_k^\sharp(y^k, x^j) = f^0(x^j, y^k - x^j)$, a downshifted tangent $m_k^\downarrow(\cdot, x^j)$, and the aggregate plane $m_k^*(\cdot, x^j)$. Apply rule (R_3) to avoid overflow. Build $\Phi_{k+1}(\cdot, x^j)$, and goto step 7.
- 7: **Update proximity parameter.** Compute

$$\tilde{\rho}_k = \frac{f(x^j) - \Phi_{k+1}(y^k, x^j)}{f(x^j) - \Phi_k(y^k, x^j)}.$$

Put

$$\tau_{k+1} = \begin{cases} \tau_k, & \text{if } \tilde{\rho}_k < \tilde{\gamma} & \text{(bad)} \\ 2\tau_k, & \text{if } \tilde{\rho}_k \geq \tilde{\gamma} & \text{(too bad)} \end{cases}$$

Then increase counter k and continue inner loop with step 4.

- 8: **Update Q_j and memory element.** Update matrix $Q_j \rightarrow Q_{j+1}$, respecting $Q_{j+1} = Q_{j+1}^\top$ and $-qI \preceq Q_{j+1} \preceq qI$. Then store new memory element

$$\tau_{j+1}^\sharp = \begin{cases} \tau_k, & \text{if } \gamma \leq \rho_k < \Gamma & \text{(not bad)} \\ \frac{1}{2}\tau_k, & \text{if } \rho_k \geq \Gamma & \text{(good)} \end{cases}$$

Increase τ_{j+1}^\sharp if necessary to ensure $Q_{j+1} + \tau_{j+1}^\sharp I \succ 0$.

- 9: **Large multiplier safeguard rule.** If $\tau_{j+1}^\sharp > T$ then re-set $\tau_{j+1}^\sharp = T$. Increase outer loop counter j by 1 and loop back to step 2.
-

Definition 4. A cutting plane oracle \mathcal{O} for f is called strict at x_0 if the following strict version of (O_2) is satisfied:

$$(\widehat{O}_2) \text{ Suppose } y_j, x_j \rightarrow x. \text{ Then there exist } \epsilon_j \rightarrow 0^+ \text{ such that } f(y_j) \leq m_{y_j}(y_j, x_j) + \epsilon_j \|y_j - x_j\|. \quad \square$$

We now discuss two versions of the oracle which are of special interest for our applications.

Example 5.1 (Model-based oracle). Suppose ϕ is a model of f . Then we can generate a cutting plane for serious iterate x and trial step y by taking $g \in \partial_1 \phi(y, x)$ and putting

$$m_y(\cdot, x) = \phi(y, x) + g^\top(\cdot - y) = \phi(y, x) + g^\top(x - y) + g^\top(\cdot - x).$$

Oracles generated by a model ϕ in this way will be denoted \mathcal{O}_ϕ . Note that \mathcal{O}_ϕ coincides with the standard oracle if f is convex and $\phi(\cdot, x) = f$, i.e., if the convex f is chosen as its own model. In more general cases, the simple idea of this oracle is that in the absence of convexity, where tangents to f at y are not useful, we simply take tangents of $\phi(\cdot, x)$ at y . Note that the model-based oracle \mathcal{O}_ϕ is strict as soon as the model ϕ is strict. \square

Example 5.2 (Standard oracle). A special case of the model-based oracle is obtained by choosing the standard model ϕ^\sharp . Due to its significance for our present work we call this the standard oracle. The standard cutting plane for serious step x and null step y is $m_y^\sharp(\cdot, x) = f(x) + g^\top(\cdot - x)$, where the Clarke subgradient $g \in \partial f(x)$ is one of those that satisfy $g^\top(y - x) = f^0(x, y - x)$. The standard oracle is strict iff ϕ^\sharp is strict. As was observed before, this is for instance the case when f is upper- C^1 . Note a specificity of the standard oracle: every standard cutting plane $m_y^\sharp(\cdot, x)$ is also an exactness plane at x . \square

Example 5.3 (Downshifted tangents). Probably the oldest oracle used for nonconvex functions are downshifted tangents, which we define as follows. For serious iterate x and null step y let $t(\cdot) = f(y) + g^\top(\cdot - y)$ be a tangent of f at y . That is, $g \in \partial f(y)$. Then we shift $t(\cdot)$ down until it becomes useful for the model (5). Fixing a parameter $c > 0$, this is organized as follows: We define the cutting plane as $m_y^\downarrow(\cdot, x) = t(\cdot) - s$, where the downshift $s \geq 0$ satisfies

$$s = [t(x) - f(x) + c\|y - x\|^2]_+.$$

In other words, $m_y^\downarrow(\cdot, x) = a + g^\top(\cdot - x)$, where $a = \min\{t(x), f(x) - c\|y - x\|^2\}$. Note that this procedure always satisfies axioms (O_1) and (O_3) , whereas axioms (O_2) , respectively, (\widehat{O}_2) , are satisfied if f is lower- C^1 at x_0 . In other words, see [7], for f lower- C^1 this is an oracle, which is automatically strict. \square

Motivated by the previous examples, we now define an oracle which works for both lower- C^1 and upper- C^1 .

Example 5.4 (Modified downshift). Let x be the current serious iterate, y a null step in the inner loop belonging to x . Then we form the downshifted tangent $m_y^\downarrow(\cdot, x) := t(\cdot) - s$, that is, the cutting plane we would get from the downshift oracle, and we form the standard oracle plane $m_y^\sharp(\cdot, x) = f(x) + g^\top(\cdot - x)$, where the Clarke subgradient g satisfies $f^0(x, y - x) = g^\top(y - x)$. Then we define

$$m_y(\cdot, x) = \begin{cases} m_y^\downarrow(\cdot, x) & \text{if } m_y^\downarrow(y, x) \geq m_y^\sharp(y, x) \\ m_y^\sharp(\cdot, x) & \text{else} \end{cases}$$

In other words, among the two candidate cutting planes $m_y^\downarrow(\cdot, x)$ and $m_y^\sharp(\cdot, x)$, we take the one which has the larger value at the null step y .

Note that this is the oracle we use in our algorithm. Theorem 1 clarifies when this oracle is strict. \square

Given an operator \mathcal{O} which with every pair (x, y) of serious step x and null step y associates a cutting plane $m_y(\cdot, x) = a + g^\top(\cdot - x)$, we fix a constant $M > 0$ and define what we call the upper envelope function of the oracle

$$\phi^\uparrow(\cdot, x) = \sup\{m_y(\cdot, x) : \|y - x\| \leq M\}.$$

The crucial property of ϕ^\uparrow is the following

Lemma 3. *Suppose $\mathcal{O} : (x, y) \mapsto m_y(\cdot, x)$ is a cutting plane oracle satisfying axioms $(O_1) - (O_3)$. Then ϕ^\uparrow is a model of f . Moreover, if the oracle satisfies (\widehat{O}_2) , then ϕ^\uparrow is strict. \square*

The proof can be found in [7]. We refer to ϕ^\uparrow as the upper envelope model associated with the oracle \mathcal{O} . Since in turn every model ϕ gives rise to a model-based oracle, \mathcal{O}_ϕ , it follows that having a strict oracle and having a strict model are equivalent properties of f . Note, however, that the model ϕ^\uparrow is in general not practically useful. It is a theoretical tool in the convergence proof.

Remark 5. If we start with a model ϕ , then build \mathcal{O}_ϕ , and go back to ϕ^\uparrow , we get back to ϕ , at least locally.

On the other hand, going from an oracle \mathcal{O} to its envelope model ϕ^\uparrow , and then back to the model based oracle $\mathcal{O}_{\phi^\uparrow}$ does *not* necessarily lead back to the oracle \mathcal{O} .

We are now in the position to check axiom (R_5) .

Corollary 1. *All working models ϕ_k constructed in our algorithm satisfy $\partial_1 \phi_k(x, x) \subset \partial f(x)$. \square*

6. MAIN CONVERGENCE RESULT

In this section we state and prove the main result of this work and give several consequences.

Theorem 1. *Let f be locally Lipschitz and suppose for every $x \in \mathbb{R}^n$, f is either lower- C^1 or upper- C^1 at x . Let x^1 be such that $Ax^1 \leq b$ and $\{x \in \mathbb{R}^n : f(x) \leq f(x^1), Ax \leq b\}$ is bounded. Then every accumulation point x^* of the sequence x^j of serious iterates generated by algorithm 1 is a KKT-point of (1).*

Proof. The result will follow from [7, Theorem 1] as soon as we show that downshifted tangents as modified in Example 5.4 and used in the algorithm is a strict cutting plane oracle in the sense of definition 4. The remainder of the proof is to verify this.

1) Let us denote cutting planes arising from the standard model ϕ^\sharp by $m_y^\sharp(\cdot, x)$, cutting planes obtained by downshift as $m_y^\downarrow(\cdot, x) = t(\cdot) - s$, and the true cutting plane of the oracle as $m_y(\cdot, x)$. Then as we know $m_y(\cdot, x) = m_y^\downarrow(\cdot, x)$ if $m_y^\downarrow(y, x) \geq m_y^\sharp(y, x)$, and otherwise $m_y(\cdot, x) = m_y^\sharp(\cdot, x)$. We have to check (O_1) , (\widehat{O}_2) , (O_3) .

2) The validity of (O_1) is clear, as both oracles provide Clarke tangent planes to f at x for $y = x$.

3) Let us now check (O_3) . Consider $x_j \rightarrow x$, and $y_j, y_j^+ \rightarrow y$. Here y_j^+ is a null step at serious step x_j . Passing to a subsequence, we may distinguish case I, where $m_{y_j^+}(\cdot, x_j) = m_{y_j^+}^\sharp(\cdot, x_j)$ for every j , and case II, where $m_{y_j^+}(\cdot, x_j) = m_{y_j^+}^\downarrow(\cdot, x_j)$ for every j .

Consider case I first. Let $m_{y_j^+}^\sharp(y_j, x_j) = f(x_j) + g_j^\top(y_j - x_j)$, where $g_j \in \partial f(x_j)$ satisfies $f^0(x_j, y_j^+ - x_j) = g_j^\top(y_j^+ - x_j)$. Passing to yet another subsequence, we may assume $g_j \rightarrow g$, and upper semi-continuity of the Clarke subdifferential gives $g \in \partial f(x)$. Therefore $m_{y_j^+}^\sharp(y_j, x_j) = f(x_j) + g_j^\top(y_j - x_j) \rightarrow f(x) + g^\top(y - x) \leq m_y^\sharp(y, x) \leq m_y(y, x)$. So here (O_3) is satisfied with $z = y$.

Now consider case II. Here we have $m_{y_j^+}(\cdot, x_j) = t_{g_j}(y_j) - s_j$, where $t_{g_j}(\cdot)$ is a tangent to f at y_j^+ with subgradient $g_j \in \partial f(y_j^+)$, and s_j is the corresponding downshift

$$s_j = [t_{g_j}(x_j) - f(x_j) + c\|y_j^+ - x_j\|^2]_+.$$

Passing to a subsequence, we may assume $g_j \rightarrow g$, and by upper semi-continuity of ∂f we have $g \in \partial f(y)$. Therefore $s_j \rightarrow [t_g(x) - f(x) + c\|y - x\|^2]_+ =: s$, where uniform convergence $t_{g_j}(y_j) \rightarrow t_g(y)$ occurs due to the boundedness of ∂f . But now we see that s is the downshift for the pair (x, y) when $g \in \partial f(y)$ is used. Hence $m_{y_j^+}(y_j, x_j) \rightarrow m_y^\downarrow(y, x)$, and since $m_y^\downarrow(y, x) \leq m_y(y, x)$, we are done. So again the z in (O_3) equals y here.

4) Let us finally check axiom (\widehat{O}_2) . Let $x_j, y_j \rightarrow x$ be given. We first consider the case when f is upper- C^1 at x . We have to find $\epsilon_j \rightarrow 0^+$ such that $f(y_j) \leq m_{y_j}(y_j, x_j) + \epsilon_j\|y_j - x_j\|$ as $j \rightarrow \infty$, and by the definition of the oracle, it clearly suffices to show $f(y_j) \leq m_{y_j}^\sharp(y_j, x_j) + \epsilon_j\|y_j - x_j\|$. By Spingarn [19], or Daniilidis and Georgiev [20], $-f$, which is lower- C^1 at x , has the following property: For every $\epsilon > 0$ there exists $\delta > 0$ such that for all $0 < t < 1$ and $y, z \in B(x, \delta)$,

$$f(y) \leq f(z) + t^{-1}(f(z + t(y - z)) - f(z)) + \epsilon(1 - t)\|z - y\|.$$

Taking the limit superior $t \rightarrow 0^+$ implies

$$f(y) \leq f(z) + f'(z, y - z) + \epsilon\|y - z\| \leq f(z) + f^0(z, y - z) + \epsilon\|y - z\|.$$

Choosing $z = x_j$, $y = y_j$, $\delta_j = \|y_j - z_j\| \rightarrow 0$, we can find $\epsilon_j \rightarrow 0^+$ such that $f(y_j) \leq f(x_j) + f^0(x_j, y_j - x_j) + \epsilon_j\|y_j - x_j\|$, hence $f(y_j) \leq m_{y_j}^\sharp(y_j, x_j) + \epsilon_j\|y_j - x_j\|$ by the definition of $m_{y_j}^\sharp(\cdot, x_j)$. That settles the upper- C^1 case.

Now consider the case where f is lower- C^1 at x . We have to find $\epsilon_j \rightarrow 0^+$ such that $f(y_j) \leq m_{y_j}(y_j, x_j) + \epsilon_j\|y_j - x_j\|$ as $j \rightarrow \infty$, and it suffices to show $f(y_j) \leq m_{y_j}^\downarrow(y_j, x_j) + \epsilon_j\|y_j - x_j\|$. Since $m_{y_j}^\downarrow(y_j, x_j) \geq f(y_j) - s_j$, where s_j is the downshift $s_j = [t(x_j) - f(x_j) + c\|y_j - x_j\|^2]_+$, and $t(\cdot) = f(y_j) + g_j^\top(\cdot - y_j)$ for some $g_j \in \partial f(y_j)$, it suffices to exhibit $\epsilon_j \rightarrow 0^+$ such that $f(y_j) \leq f(y_j) - s_j + \epsilon_j\|y_j - x_j\|$, or what is the same, $s_j \leq \epsilon_j\|y_j - x_j\|$. For that it suffices to arrange $[t(x_j) - f(x_j)]_+ \leq \epsilon_j\|y_j - x_j\|$, because once this is verified, we get $s_j \leq [t(x_j) - f(x_j)]_+ + c\|y_j - x_j\|^2 \leq (\epsilon_j + c\|y_j - x_j\|)\|y_j - x_j\| =: \tilde{\epsilon}_j\|y_j - x_j\|$. Note again that by [19, 20] f has the following property at x : For every $\epsilon > 0$ there exists $\delta > 0$ such that $f(tz + (1 - t)y) \leq tf(z) + (1 - t)f(y) + \epsilon t(1 - t)\|z - y\|$ for all $y, z \in B(x, \delta)$. Dividing by $t > 0$ and passing to the limit $t \rightarrow 0^+$ gives $f^0(y, z - y) \leq f(z) - f(y) + \epsilon\|y - z\|$, using the fact that f is locally Lipschitz. But for every $g \in \partial f(y)$, $g^\top(z - y) \leq f^0(y, z - y)$. Using $\|y_j - x_j\| =: \delta_j \rightarrow 0$ and taking $y = y_j$, $z = x_j$, this allows us to find $\epsilon_j \rightarrow 0^+$ such that $g_j^\top(x_j - y_j) \leq f(x_j) - f(y_j) + \epsilon_j\|y_j - x_j\|$. Substituting this above gives $t(x_j) - f(x_j) = f(y_j) - f(x_j) + g_j^\top(x_j - y_j) \leq \epsilon_j\|y_j - x_j\|$ as desired. That settles the lower- C^1 case. \square

7. PRACTICAL ASPECTS OF THE ALGORITHM

In this section we discuss several technical aspects of the algorithm, which are important for its performance.

7.1. Stopping. The stopping test in step 2 of the algorithm is stated in this form for the sake of the convergence proof. In practice we delegate stopping to the inner loop using the following two-stage procedure.

If the inner loop at serious iterate x^j finds the new serious step x^{j+1} such that

$$\frac{\|x^{j+1} - x^j\|}{1 + \|x^j\|} < \text{tol}_1, \quad \frac{|f(x^{j+1}) - f(x^j)|}{1 + |f(x^j)|} < \text{tol}_2,$$

then we decide that x^{j+1} is optimal. In consequence, the $(j + 1)$ st inner loop will not be executed. On the other hand, if the inner loop has difficulties terminating and produces

five consecutive null steps y^k where

$$\frac{\|y^k - x^j\|}{1 + \|x^j\|} < \text{tol}_1, \quad \frac{|f(y^k) - f(x^j)|}{1 + |f(x^j)|} < \text{tol}_2,$$

or if a maximum number k_{\max} of allowed steps in the inner loop is reached, then we decide that x^j is optimal. In our experiments we use $\text{tol}_1 = 10^{-5}$, $\text{tol}_2 = 10^{-5}$, and $k_{\max} = 50$.

7.2. Recycling of planes. At the beginning of a new inner loop at serious step x^{j+1} , we do not want to start building the working model $\phi_1(\cdot, x^{j+1})$ from scratch. It is more efficient to recycle some of the planes $(a, g) \in \mathcal{G}_{k_j}$ in the latest working model $\phi_{k_j}(\cdot, x^j)$. In the convex cutting plane method, this is self-understood, as cutting planes are affine minorants of f , and can at leisure stay on in the sets \mathcal{G} at all times j, k . Without convexity, we need the following recycling procedure:

Given a plane $m(\cdot, x^j) = a + g^\top(\cdot - x^j)$ in the latest set \mathcal{G}_{k_j} , we form the new downshifted plane

$$m(\cdot, x^{j+1}) = m(\cdot, x^j) - s,$$

where the downshift is organized as

$$s = [m(x^{j+1}, x^j) - f(x^{j+1}) + c\|x^{j+1} - x^j\|^2]_+.$$

In other words, we treat $m(\cdot, x^j)$ like a tangent to f at null step x^j with respect to the serious step x^{j+1} in the downshift oracle. We put

$$m(\cdot, x^{j+1}) = a + g^\top(\cdot - x^j) - s = a - s + g^\top(x^{j+1} - x^j) + g^\top(\cdot - x^{j+1}),$$

and we accommodate $(a - s + g^\top(x^{j+1} - x^j), g) \in \mathcal{G}_1$ at the beginning of the $(j+1)$ st inner loop. In the modified version we only keep a plane of this type in \mathcal{G}_1 after comparing it to the exactness plane $m_0(\cdot, x^{j+1}) = f(x^{j+1}) + g^\top(\cdot - x^{j+1})$, $g \in \partial f(x^{j+1})$, which satisfies $g^\top(x^j - x^{j+1}) = f^0(x^{j+1}, x^j - x^{j+1})$. Indeed, when $m(x^j, x^{j+1}) \geq m_0(x^j, x^{j+1})$, then we keep the downshifted plane, otherwise we add $m_0(\cdot, x^{j+1})$ as additional exactness plane.

8. THE DELAMINATION BENCHMARK PROBLEM

The interface behavior of laminated composite materials is modeled by a non-monotone multi-valued function ∂j , characteristic of the interlayer adhesive placed at the contact boundary Γ_c . In more precise terms, ∂j is the physical law which holds between the normal component $-S_n(s)|_{\Gamma_c}$ of the stress vector and the relative displacement $u_2(s)|_{\Gamma_c}$, or *jump*, between the upper and lower boundaries. A typical law ∂j for an interlayer adhesive is shown in Figure 1 (left). In the material sciences, the knowledge of ∂j is crucial for the understanding of the basic failure modes of the composite material.

The adhesive law ∂j is usually determined experimentally using the double cantilever beam test [1] or other destructive testing methods. The result of a typical experiment is shown schematically in Figure 3 from [1], where three probes with different levels of contamination have been exposed. While the intact material shows stable propagation of the crack front (dashed curve), the 10% contaminated specimen shows a typical zig-zag profile (bold solid curve), indicating unstable crack front propagation. Indeed, when reaching the critical load $P = 140\text{N}$, the crack starts to propagate. Since by the growth of the crack-elongation, the compliance of the structure increases, the crack propagation slows down and the crack is "caught", i.e., stops at $u_2 = 0.25\text{mm}$ and the load P in the structure drops from $P = 140\text{N}$ to $P = 40\text{N}$. Thereafter, due to the continuously increased load, the crack starts again to propagate until reaching another critical load level at $P = 90\text{N}$ and $u_2 = 5\text{mm}$. This phenomenon occurs five to six times, as seen in Figure 3.

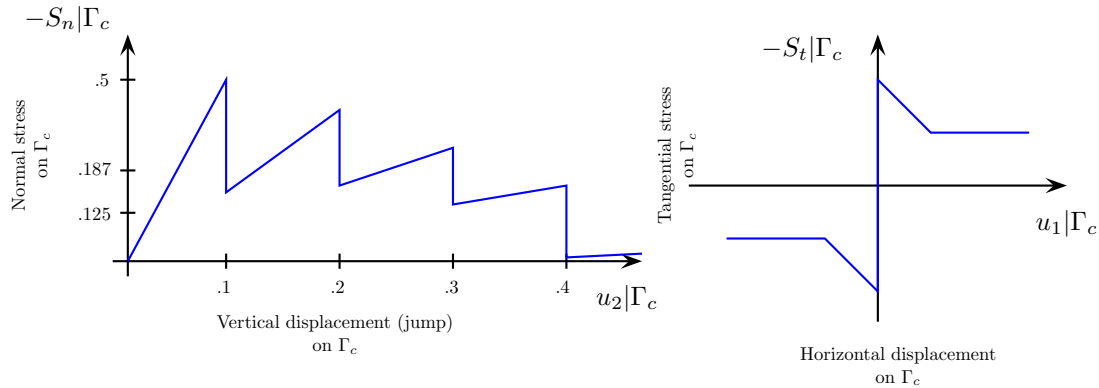


FIGURE 1. Left image shows non-monotone delamination law ∂j , leading to an upper- C^1 objective. Right image shows non-monotone friction law, leading to a lower- C^1 objective.

The 50% contaminated specimens (dotted curve) shows micro-cracks that appear at a finer level and are not visible in the Figure 3. The lower level of the adhesive energy, which is represented by the area below the load-displacement curve, indicates now that this specimen is of minor resistance.

Even though the displacement u_2 in Figure 3 can only be measured at the crack tip, in order to proceed one now *stipulates* the law ∂j all along $s \in \Gamma_c$ by assuming that the normal stresses $S_n(s)|\Gamma_c$ follow the measured behavior

$$(12) \quad -S_n(s) \in \partial j(s, u_2(s)), \quad s \in \Gamma_c.$$

Under this hypothesis one now solves the variational inequality for the unknown displacement field $\mathbf{u} = (u_1, u_2)$, and then validates (12). Note that $S_n(s)|\Gamma_c$ is the *truly* relevant information, as it indicates the action of the destructive forces along Γ_c , explaining eventual failure of the composite. In current practice in the material sciences, this information cannot be assessed by direct measurement, and is therefore estimated by heuristic formulae [1]. Our approach could be interpreted as one such estimation technique based on mathematical modeling.

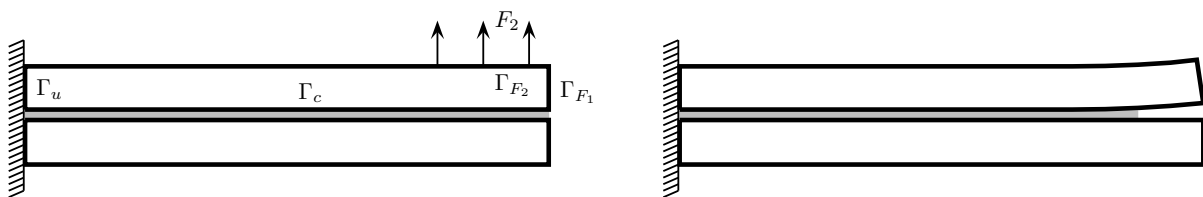


FIGURE 2. Schematic view of cantilever beam testing. Under applied traction force F_2 the crack front propagates to the left. In program (16) traction force F_2 and crack front length are given, while the corresponding displacement u and reactive forces $-S_n|_{\Gamma_c}$ along the contact boundary Γ_c have to be computed.

8.1. Delamination study. Within the framework of plane linear elasticity we consider a symmetric laminated structure with an interlayer adhesive under loading (see Fig. 2). Because of the symmetry of the structure, it suffices to consider only the upper half of the specimen, represented by $\Omega \subset \mathbb{R}^2$. The Lipschitz boundary Γ of Ω consists of four

disjoint parts Γ_u , Γ_c , Γ_{F_1} and Γ_{F_2} . The body is fixed on Γ_u , i.e.,

$$u_i = 0 \text{ on } \Gamma_u, \quad i = 1, 2.$$

On Γ_{F_1} the traction forces \mathbf{F} are constant and given as

$$\mathbf{F} = (0, F_2) \quad \text{on } \Gamma_{F_1}.$$

The part Γ_{F_2} is load-free. We adopt standard notation from linear elasticity and introduce the bilinear form of linear elasticity

$$(13) \quad a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \varepsilon(\mathbf{u}) : \sigma(\mathbf{v}) \, dx,$$

where $\mathbf{u} = (u_1, u_2)$ is the displacement vector, $\varepsilon(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \mathbf{u})^T)$ the linearized strain tensor, and $\sigma(\mathbf{v}) = \mathbf{C} : \varepsilon(\mathbf{v})$ the stress tensor. Here, \mathbf{C} is the elasticity tensor with symmetric positive L^∞ coefficients. The bilinear form is symmetric and due to the first Korn inequality, coercive. The linear form $\langle \mathbf{g}, \cdot \rangle$ is defined by

$$\langle \mathbf{g}, \mathbf{v} \rangle = F_2 \int_{\Gamma_{F_1}} v_2 \, ds.$$

On the contact boundary Γ_c we have the unilateral constraint

$$u_2 \geq 0 \quad \text{a.e. on } \Gamma_c$$

and we apply the non-monotone multi-valued adhesive law

$$(14) \quad -S_n(s) \in \partial j(s, u_2(s)) \quad \text{for a.a. } s \in \Gamma_c.$$

Here $S_n = \sigma_{ij} n_j n_i$, where $\mathbf{n} = (n_1, n_2)$ is the outward unit normal vector to Γ_c .

A typical non-monotone law $\partial j(s, \cdot)$ for delamination, describing the behavior of the adhesive, is shown in Fig. 1. This law is derived from a nonconvex and a nonsmooth locally Lipschitz super-potential j expressed in terms of a minimum function. In particular, $j(s, \cdot)$ is a minimum of four convex quadratic and one linear function.

We also assume that tangential traction can be neglected on Γ_c , i.e., $S_t(s) = 0$. The weak formulation of the delamination problem is then given by the following hemivariational inequality: Find $\mathbf{u} \in K$ such that

$$(15) \quad a(\mathbf{u}, \mathbf{v} - \mathbf{u}) + \int_{\Gamma_c} j^0(s, u_2(s); v_2(s) - u_2(s)) \, ds \geq \langle \mathbf{g}, \mathbf{v} - \mathbf{u} \rangle \quad \forall \mathbf{v} \in K,$$

where $j^0(s, u; d)$ is the Clarke directional derivative of $j(s, \cdot)$ at u in direction d , K is the nonempty, closed convex set of all admissible displacements defined by

$$K = \{\mathbf{v} \in V : v_2 \geq 0 \text{ on } \Gamma_c\},$$

contained in the function space

$$V = \{\mathbf{v} \in H^1(\Omega; \mathbb{R}^2) : \mathbf{v} = 0 \text{ on } \Gamma_u\}.$$

The potential energy of the problem is

$$\Pi(\mathbf{v}) = \frac{1}{2}a(\mathbf{v}, \mathbf{v}) + J(\mathbf{v}) - \langle \mathbf{g}, \mathbf{v} \rangle,$$

where $J : V \rightarrow \mathbb{R}$ defined by

$$J(\mathbf{v}) = \int_{\Gamma_c} j(s, v_2(s)) \, ds$$

is the term responsible for the nonsmoothness. Using the potential energy, the hemivariational inequality (15) can be transformed to the following nonsmooth, nonconvex constrained optimization problem of the form (1)

$$(16) \quad \begin{aligned} & \text{minimize} && \Pi(\mathbf{u}) \\ & \text{subject to} && \mathbf{u} \in K \end{aligned}$$

where the objective is upper- C^1 , because the super-potential $j(s, \cdot)$ is a minimum. In particular, we have an objective of the form (3), where the smooth part f_s comprises $\frac{1}{2}a(\mathbf{v}, \mathbf{v}) - \langle \mathbf{g}, \mathbf{v} \rangle$, while the nonsmooth part $J(\mathbf{v}) = \int_{\Gamma_c} j(s, v_2(s)) ds$ has the form (3) with a finite index set I once the boundary integral is suitably parametrized.

According to the existence theory in [30], problem (16) has at least one Clarke critical point \mathbf{u}^* satisfying the necessary optimality condition

$$0 \in \partial \Pi(\mathbf{u}^*) + N_K(\mathbf{u}^*),$$

where $N_K(\mathbf{u})$ is the normal cone to K at \mathbf{u} , and vice versa, by a result in [12] every critical point of Π on K is a solution of (15) (see also [11]).

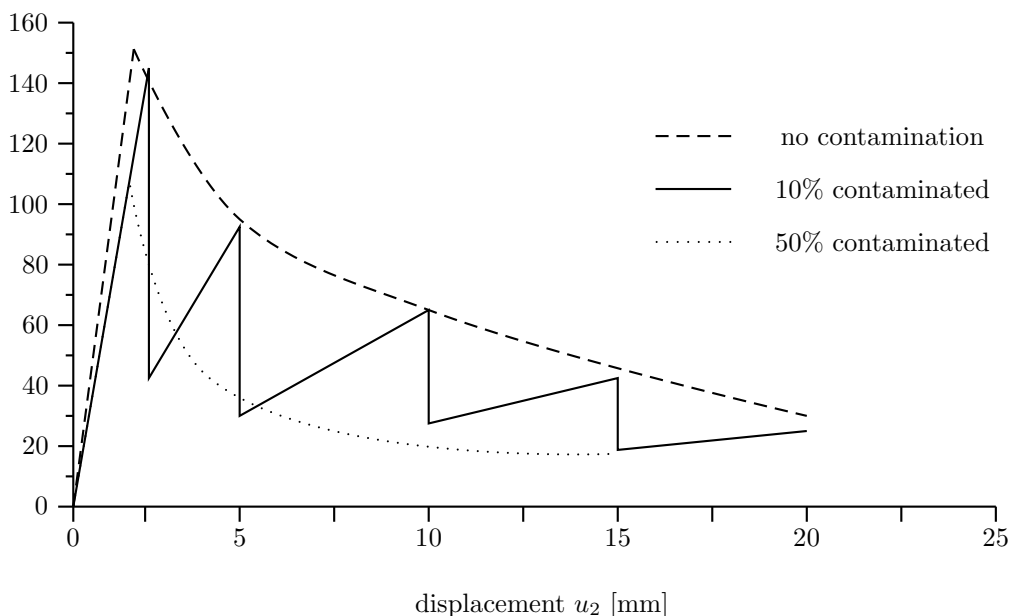


FIGURE 3. Load-displacement curve determined by double cantilever beam test. Dashed curve shows stable behavior for material without contamination. The 10% contaminated specimen (bold solid curve) shows unstable crack growth. After initial linear growth, when the critical load $P = 140\text{N}$ is reached, the crack starts to propagate. But then the propagation speed slows down, since by the crack the compliance of the specimen increases, and the crack is "caught" at $u_2 = 0.25\text{mm}$. The load P drops from $P = 140\text{N}$ to $P = 40\text{N}$. Then, by the constantly applied traction force, there is a linear growth of the load P from $P = 40\text{N}$ to the critical load $P = 90\text{N}$, where the crack propagates again and stops at $u_2 = 5\text{mm}$, with the load now reduced to $P = 30\text{N}$. The 50% contaminated specimen exhibits micro-cracks not visible at the chosen scale.

8.2. Discrete problem. We consider a regular triangulation $\{\mathcal{T}_h\}$ of Ω , where we first divide Ω into small squares of size h and then each square by its diagonal into two triangles.

To approximate V and K we use a piecewise linear finite element approximation and set

$$V_h = \{v_h \in C(\bar{\Omega}; \mathbb{R}^2) : v_h|_T \in (\mathbf{P}_1)^2, \forall T \in \mathcal{T}_h, v_h|_{\Gamma_u} = 0\},$$

$$K_h = \{v_h \in V_h : v_{h2}(s_\nu) \geq 0 \quad \forall s_\nu \in \bar{\Gamma}_c \setminus \bar{\Gamma}_u\}.$$

Similar to low order finite element approximations of nonsmooth convex contact problems [37, 39], we use the trapezoidal quadrature rule to approximate the functional J by

$$(17) \quad J_h(v_h) = \frac{1}{2} \sum_{s_\nu \in \bar{\Gamma}_c \setminus \bar{\Gamma}_u} |s_\nu s_{\nu+1}| [j(s_\nu, v_{h2}(s_\nu)) + j(s_{\nu+1}, v_{h2}(s_{\nu+1}))],$$

where we are summing over the nodes s_ν on the contact boundary $\bar{\Gamma}_c \setminus \bar{\Gamma}_u$, with $s_{\nu+1}$ being the neighbor of node s_ν on Γ_c in the sense of integration. This can be regrouped as

$$J_h(v_h) = \sum_{s_\nu \in \bar{\Gamma}_c \setminus \bar{\Gamma}_u} c_\nu j(s_\nu, v_{h2}(s_\nu)) = \sum_{s_\nu \in \bar{\Gamma}_c \setminus \bar{\Gamma}_u} c_\nu \min_{i \in I} j_i(s_\nu, v_{h2}(s_\nu))$$

with appropriate weights $c_\nu > 0$. Here, I is the set of zig-zags in the graph of ∂j .

The bundle algorithm is applied to minimize the discrete functional

$$(18) \quad \Pi_h(v_h) = \frac{1}{2} a(v_h, v_h) + J_h(v_h) - \langle g, v_h \rangle \quad \text{on } K_h.$$

Introducing an index set N for the nodes s_ν on the contact boundary $\bar{\Gamma}_c$, we may pull out the minimum from under the sum, which leads to the expression

$$\Pi_h(v_h) = \frac{1}{2} a(v_h, v_h) + \min_{i(\cdot) \in I^N} \sum_{\nu \in N} c_\nu j_{i(\nu)}(s_\nu, v_{h2}(s_\nu)) - \langle g, v_h \rangle.$$

This is the discrete version of (4), where $\frac{1}{2} a(v_h, v_h) - \langle g, v_h \rangle$ is the smooth term f_s , and J_h the nonsmooth part.

While computation of Clarke subgradients is straightforward here, we still have to explain how the matrix $Q = Q(v)$ in the second-order working model (6) is chosen. Discretizing the quadratic form of linear elasticity as $a(v_h, v_h) = v_h^\top \mathbf{A} v_h$ with the symmetric stiffness matrix \mathbf{A} , and observing that $\langle g, v_h \rangle = \mathbf{g}^\top v_h$ is linear, we choose $Q(v) = \mathbf{A} + \sum_{\nu \in N} \nabla^2 j_{i(\nu)}(s_\nu, v_h(s_\nu))$, where $i(\nu) \in I$ is one of those indices, where the minimum $\min_{i \in I} j_i(s_\nu, v_{h2}(s_\nu))$ is attained.

For convergence of the lowest-order finite element approximation used here we refer to the results in [31]. Higher-order approximations with no limitation in the polynomial degree, which lead to nonconforming approximation of unilateral constraints, have only recently been analyzed for monotone contact problems, see [40].

8.3. Numerical results. We present numerical results obtained in a delamination simulation with modulus of elasticity $E = 210$ GPa and Poisson ratio $\nu = 0.3$ corresponding to a steel specimen. In all examples we use the benchmark model of [35] with geometrical characteristics $(0, 100) \times (0, 10)$ in [mm] and thickness 5mm. We apply our bundle method to (16) and compare the results to those obtained by the regularization technique in [31, 33]. All computations use piecewise linear functions and the discretization 40×4 corresponding to $h = 0.25$ cm. In this case, the number of the unknowns in the discrete problem (18) is 80.

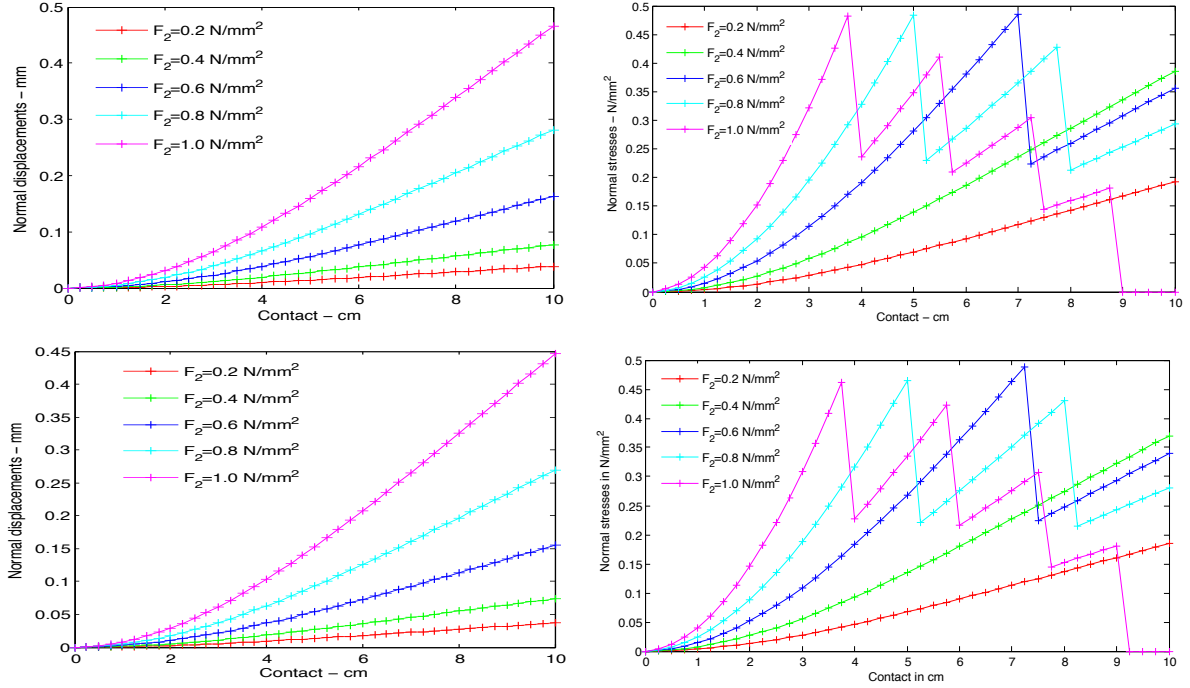


FIGURE 4. Upper: regularization method of [31, 33]. Lower: optimization method. Left image shows vertical displacement u_2 for 5 different values of F_2 . Right image shows vertical component of reactive force along contact boundary for same 5 scenarios.

TABLE 1. Regularization. Vertical displacement [mm] at 4 intermediate points for same 5 scenarios.

F_2 [N/mm ²]	$u_2(x_1)$	$u_2(x_2)$	$u_2(x_3)$	$u_2(x_4)$
0.2	4.154500e-06	1.394500e-05	2.601700e-05	3.858700e-05
0.4	8.308100e-06	2.788800e-05	5.202800e-05	7.716600e-05
0.6	1.633200e-05	5.622700e-05	1.080000e-04	1.640000e-04
0.8	2.792500e-05	9.663100e-05	1.860000e-04	2.810000e-04
1.0	4.600600e-05	1.590000e-04	3.080000e-04	4.660000e-04

TABLE 2. Optimization. Vertical displacement [mm] at four intermediate points for same 5 scenarios.

F_2 [N/mm ²]	$u_2(x_1)$	$u_2(x_2)$	$u_2(x_3)$	$u_2(x_4)$
0.2	4.022500e-06	1.345400e-05	2.499300e-05	3.691900e-05
0.4	8.069300e-06	2.698800e-05	5.013300e-05	7.404900e-05
0.6	1.564800e-05	5.373900e-05	1.030000e-04	1.550000e-04
0.8	2.691300e-05	9.297200e-05	1.790000e-04	2.700000e-04
1.0	4.414000e-05	1.530000e-04	2.940000e-04	4.470000e-04

CONCLUSION

We have presented a bundle method based on the mechanism of downshifted tangents which is suited to optimize upper- and lower- C^1 functions. Our method allows to integrate second-order information, if available, and gives a convergence certificate in the

TABLE 3. Regularization. Horizontal displacement [mm] at four intermediate points for same 5 scenarios.

$F_2 [N/mm^2]$	$u_2(x_1)$	$u_2(x_2)$	$u_2(x_3)$	$u_2(x_4)$
0.2	1.481900e-06	2.251300e-06	2.474400e-06	2.499500e-06
0.4	2.963600e-06	4.502200e-06	4.948300e-06	4.998500e-06
0.6	5.918500e-06	9.400600e-06	1.077100e-05	1.097500e-05
0.8	1.015200e-05	1.625600e-05	1.866400e-05	1.904000e-05
1.0	1.674400e-05	2.690100e-05	3.100500e-05	3.167000e-05

TABLE 4. Optimization. Horizontal displacement [mm] at four intermediate points for same 5 scenarios.

$F_2 [N/mm^2]$	$u_2(x_1)$	$u_2(x_2)$	$u_2(x_3)$	$u_2(x_4)$
0.2	1.432200e-06	2.161500e-06	2.356100e-06	2.368400e-06
0.4	2.872700e-06	4.335000e-06	4.724700e-06	4.748800e-06
0.6	5.663400e-06	8.957000e-06	1.023200e-05	1.041100e-05
0.8	9.777300e-06	1.561000e-05	1.787700e-05	1.822600e-05
1.0	1.606400e-05	2.578000e-05	2.970700e-05	3.034700e-05

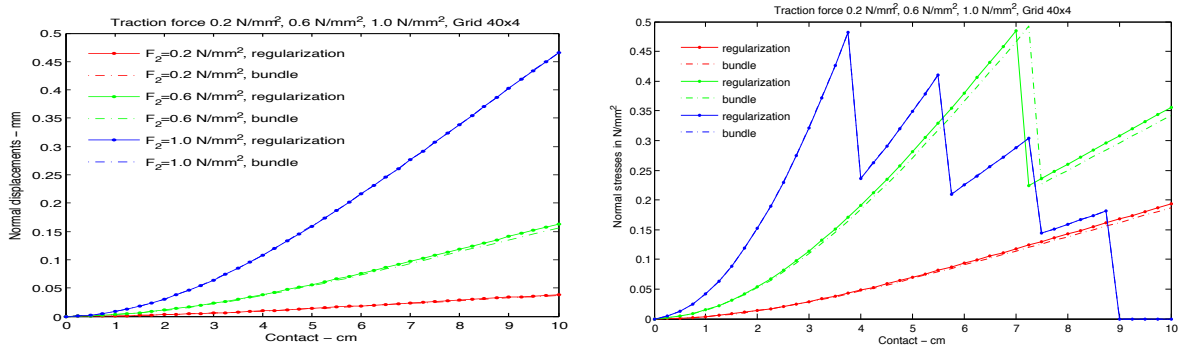
FIGURE 5. Comparison of regularization (bold solid curves) and optimization (dashed) for 3 different values of F_2 . Left vertical displacement, right reactive force.

TABLE 5. Comparison of optimal value obtained by regularization and optimization

$F_2 [N/m^2]$	$\Pi_{h_{reg}} [Nm]$	$\Pi_{h_{opt}} [Nm]$
200000	-1.32894	-1.29271
400000	-2.35224	-2.30025
600000	-3.83972	-3.74609
800000	-5.08164	-5.05389
1000000	-5.66771	-5.66770

sense of subsequences. Every accumulation point of the sequence of serious iterates with an arbitrary starting point is critical. We have successfully applied our method to a delamination problem arising in the material sciences, where upper- C^1 functions have to be minimized. Results obtained by optimization were compared to results obtained by the regularization technique of [31, 33], and both methods are in good agreement.

ACKNOWLEDGMENT

The authors thank H.-J. Gudladt for many useful discussions. The authors were partially supported by Bayerisch-Französisches Hochschulzentrum (BFHZ).

REFERENCES

- [1] M. Wetzel, J. Holtmannspötter, H.-J. Gudladt, J. v. Czarnecki: Sensitivity of double cantilever beam test to surface contamination and surface pretreatment. *International Journal of Adhesion & Adhesives*, Vol. 46, 114-121 (2013)
- [2] R. Mifflin: A modification and an extension of Lemaréchal's algorithm for nonsmooth minimization. *Math. Progr. Study* 17, 77-90 (1982)
- [3] C. Lemaréchal: Bundle methods in nonsmooth optimization. In *Nonsmooth optimization (Proc. IIASA Workshop, Laxenburg, 1977)*, pp. 79-102, IIASA Proc. Ser., 3, Pergamon, Oxford-Elmsford, N.Y., 1978.
- [4] C. Lemaréchal, C. Sagastizábal: Variable metric bundle methods: from conceptual to implementable forms. *Math. Programming* 76 (1997), no. 3, Ser. B, 393-410.
- [5] J. Zowe: The BT-Algorithm for minimizing a nonsmooth functional subject to linear constraints, in *Nonsmooth Optimization and Related Topics*, F. H. Clarke, V. F. Demyanov, F. Gianessi (eds.), Plenum Press (1989)
- [6] H. Schramm, J. Zowe: A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results, *SIAM J. Optim.* 2, 121 - 152 (1992)
- [7] D. Noll: Cutting plane oracles to minimize nonsmooth nonconvex functions. *Set-Valued Var. Anal.* 18 (3-4), 531-568 (2010)
- [8] D. Noll, O. Prot, A. Rondepierre: A proximity control algorithm to minimize nonsmooth nonconvex functions. *Pacific J. Optim.* 4 (3), 569-602 (2008)
- [9] D. Alazard, M. Gabarrou, D. Noll: Design of a flight control architecture using a nonconvex bundle method. *Math. Control Sign. Syst.* 25 (2), 257-290 (2013)
- [10] D. Noll: Convergence of nonsmooth descent methods using the Kurdyka-Łojasiewicz inequality. *J. Optim. Theory Appl.* (DOI) 10.1007/s10957-013-0391-8.
- [11] M.M. Mäkelä, M. Miettinen, L. Lukšan, J. Vlček: Comparing nonsmooth nonconvex bundle methods in solving hemivariational inequalities, *Journal of Global Optimization* 14 (2), 117-135 (1999).
- [12] M. Miettinen, M.M. Mäkelä, J. Haslinger: On numerical solution of hemivariational inequalities by nonsmooth optimization methods, *Journal of Global Optimization* 6 (4), 401-425 (1995).
- [13] L. Lukšan, J. Vlček: A Bundle-Newton method for nonsmooth unconstrained minimization, *Math. Progr.* 83, 373 - 391 (1998)
- [14] J. Haslinger, M. Miettinen, P.D. Panagiotopoulos: *Finite Element Methods for Hemivariational Inequalities*, Kluwer Academic Publishers (1999)
- [15] J. Czepiel: Proximal Bundle Method for a Simplified Unilateral Adhesion Contact Problem of Elasticity, *Schedae Informaticae* 20, 115-136 (2011)
- [16] L. Neseemann, E.P. Stephan: Numerical solution of an adhesion problem with FEM and BEM, *Appl. Numer. Math.* 62 (5), 606-619 (2012)
- [17] M. Kočvara, A. Mielke, T. Roubíček: A rate-independent approach to the delamination problem, *Math. Mech. Solids* 11, No. 4, 423-447 (2006)
- [18] T. Roubíček, V. Mantic, Panagiotopoulos, C.G.: A quasistatic mixed-mode delamination model, *Discrete Contin. Dyn. Syst., Ser. S* 6, No. 2, 591-610 (2013)
- [19] J. E. Spingarn: Submonotone subdifferentials of Lipschitz functions. *Trans. Amer. Math. Soc.* 264, 77-89 (1981)
- [20] A. Daniilidis, P. Georgiev: Approximate convexity and submonotonicity. *J. Math. Anal. Appl.* 291, 117-144 (2004)
- [21] R. T. Rockafellar, R. J-B. Wets: *Variational Analysis*. Springer Verlag (2004)
- [22] A. Daniilidis, J. Malick: Filling the gap between lower- C^1 and lower- C^2 functions. *Journal of Convex Analysis* 12(2), 2005, pp. 315 - 329.
- [23] R. A. Poliquin, R. T. Rockafellar: Prox-regular functions in variational analysis, *Trans. Amer. Math. Soc.* 348 (5), 1805 - 1838 (1996)
- [24] K.C. Kiwiel: An aggregate subgradient method for nonsmooth convex minimization, *Math. Programming* 27, 320 - 341 (1983)
- [25] J. Cullum, W.E. Donath, P. Wolfe: The minimization of certain nondifferential sums of eigenvalues of symmetric matrices. *Math. Progr. Stud.* 3, 35-55 (1975)

- [26] P. Apkarian, D. Noll, O. Prot: A trust region spectral bundle method for nonconvex eigenvalue optimization, *SIAM J. Optim.* 10 (1), 281-306 (2008)
- [27] A. Ruszczyński: *Nonlinear optimization*, Princeton University Press (2006)
- [28] J.-B. Hiriart-Urruty, C. Lemaréchal: *Convex Analysis and Minimization Algorithms*, vol. I and II: *Advanced Theory and Bundle Methods*, vol. 306 of *Grundlehren der mathematischen Wissenschaften*, Springer Verlag, New York, Heidelberg, Berlin (1993)
- [29] W. L. Hare, C. Sagastizabal: Computing proximal points of nonconvex functions, *Math. Programming series B* 116, 221-258 (2009)
- [30] Z. Naniewicz, P.D. Panagiotopoulos: *Mathematical Theory of Hemivariational Inequalities and Applications*, New York (1995).
- [31] N. Ovcharova: *Regularization Methods and Finite Element Approximation of Hemivariational Inequalities with Applications to Nonmonotone Contact Problems*, PhD Thesis, Universität der Bundeswehr München, Cuvillier Verlag, Göttingen (2012).
- [32] N. Ovcharova, J. Gwinner: A study of regularization techniques of nondifferentiable optimization in view of application to hemivariational inequalities, accepted for publication in *JOTA*, JOTA-D-13-00163
- [33] Ovcharova, N., Gwinner. J: On the regularization method in nondifferentiable optimization applied to hemivariational inequalities, *Constructive Nonsmooth Analysis and Related Topics*, Springer, 59-70 (2013).
- [34] C.C. Baniotopoulos, J. Haslinger, Z. Morávková: Contact problems with nonmonotone friction: discretization and numerical realization, *Comput. Mech.* 40, 157-165 (2007)
- [35] C.C. Baniotopoulos, J. Haslinger, Z. Morávková: Mathematical modeling of delamination and nonmonotone friction problems by hemivariational inequalities, *Applications of Mathematics* 50 (1), 1-25 (2005)
- [36] S. Carl, V.K. Le, D. Motreanu: *Nonsmooth Variational Problems and Their Inequalities*, Springer (2007)
- [37] R. Glowinski: *Numerical Methods for Nonlinear Variational Problems*, Springer, New York (1984)
- [38] D. Goeleven, D. Motreanu, Y. Dumont, M. Rochdi, M.: *Variational and Hemivariational Inequalities: Theory, Methods and Applications*, Vol. I: *Unilateral Analysis and Unilateral Mechanics*, Vol. II: *Unilateral problems*, Kluwer (2003)
- [39] J. Gwinner: Finite-element convergence for contact problems in plane linear elastostatics, *Quarterly of Applied Mathematics*, Vol. 50, 11-25 (1992)
- [40] J. Gwinner: hp-FEM convergence for unilateral contact problems with Tresca friction in plane linear elastostatics, *J. Comput. Appl. Math.*, Vol. 254, 175-184 (2013)
- [41] P.D. Panagiotopoulos: *Hemivariational inequalities. Applications in mechanics and engineering*, Berlin, Springer (1993)
- [42] P.D. Panagiotopoulos: *Inequality problems in mechanics and application. Convex and nonconvex energy functions*, Basel, Birkhäuser (1998)
- [43] M. Sofonea, A. Matei: *Variational Inequalities with Applications*, Springer (2009)