Comput Optim Appl manuscript No.
(will be inserted by the editor)

Nonconvex bundle method with application to a delamination problem

Minh N. Dao · Joachim Gwinner · Dominikus Noll · Nina Ovcharova

Received: date / Accepted: date

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.

M.N. Dao
Department of Mathematics and Informatics,
Hanoi National University of Education, Vietnam
and Institut de Mathématiques, Université de Toulouse, France
E-mail: minhdn@hnue.edu.vn

J. Gwinner
Institute of Mathematics, Department of Aerospace Engineering,
Universität der Bundeswehr München, Germany
E-mail: joachim.gwinner@unibw.de

D. Noll
Institut de Mathématiques, Université de Toulouse, France
E-mail: dominikus.noll@math.univ-toulouse.fr

N. Ovcharova
Institute of Mathematics, Department of Aerospace Engineering,
Universität der Bundeswehr München, Germany
E-mail: nina.ovcharova@unibw.de

Keywords Composite material · delamination · hemivariational inequality · regularization · Clarke directional derivative · nonconvex bundle method · lower- and upper-C^1 functions · 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 boundary 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

$$\begin{array}{l} \text{minimize} \quad f(x) \\ \text{subject to } Ax \leqslant b \end{array} \tag{1}$$

where f is locally Lipschitz, but neither smooth nor convex. Depending on the nature of the non-monotone contact problem, the criterion f may be upper-C^1 or lower-C^1, see e.g. Fig. 2. 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 in a simulation of the double cantilever beam test [43], one of the most popular destructive tests used in the material sciences 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 [29], and used in the bundle codes of Lemaréchal and Sagastizábal [22,23], or the BT-codes of Zowe [44,41], uses the method of *downshifted tangents*. While these authors use linesearch with Armijo and Wolfe type conditions, which in the nonconvex case 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 [31,8,9] 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 [34,31,33], but the latter is not compatible with the downshift oracle. While these considerations concern the nonconvex case, we recall that according to [16, Theorem XV.3.2.2(ii)] in the convex bundle method the entire sequence of serious iterates converges to a single minimum if the proximity control parameter stays bounded away from zero. For bundle methods which can deal with inexact function values or subgradients, we refer the reader to [32,14].

It is unsatisfactory to have two strings to one bow, 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 [22,44,27,25], 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. Then a representation formula for a typical nonsmooth upper-C^1 objective as encountered in delamination problems is given, highlighting the difficulty. Section 4 presents the algorithm and comments on its ingredients. Theoretical tools for 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.

Related contact problems with adhesion modeled by upper-C^1 objective function have been considered in [27,25] and in the book of Haslinger et al. [15], where numerical results based on proximal bundle method by Mäkelä [26] and bundle-Newton method by Lukšan and Vlček [24] have been presented.

For related, but different interface crack problems, their study and numerical treatment by primal-dual active-set methods, we refer to [21] and the references therein.

2 Lower- and upper-C^1 functions

Following Spingarn [42], a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ is lower-C^1 at x_0, if there exist a compact Hausdorff space K, a neighborhood U of x_0, and a mapping $F : U \times K \to \mathbb{R}$ such that both F and its partial derivative $D_x F$ are jointly continuous and

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

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.

Note that if f is lower-C^1, then it is semismooth [28, Definition 1] and Clarke regular. This follows from [42, Prop. 2.4 and Thm. 3.9]. A counterexample in [42, Section II] shows that the converse is not true; see also [14, Section 1.1 (3)]. Now observe that if f is upper-C^1, then by the above $-f$ is semismooth, and then so is f, as follows directly from the definition of semismoothness.

In the 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

$$f(x) = f_s(x) + \int_0^1 \min_{i \in I} f_i(x,t)\, dt, \tag{2}$$

where f_s gathers the smooth part, while the integral term, due to the minimum, is responsible for the nonsmoothness. For the following, let I be a finite set.

Lemma 1 *Suppose f_s is upper-C^1, the f_i are jointly continuous, and each $f_i(\cdot,t)$ is upper-C^1. Then the function (2) is upper-C^1 and can be represented in the form*

$$f(x) = f_s(x) + \min_{\sigma \in \Sigma} \int_0^1 f_{\sigma(t)}(x,t)\, dt, \tag{3}$$

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

Proof Let us first prove (3). For $\sigma \in \Sigma$ and fixed $x \in \mathbb{U}$ the function $t \mapsto f_{\sigma(t)}(x,t)$ is measurable, and since $\min_{i\in I} f_i(x,t) \leqslant f_{\sigma(t)}(x,t) \leqslant \max_{i\in I} f_i(x,t)$, 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 $F(x,\sigma) \geqslant f(x)$, so we have $\inf_{\sigma\in\Sigma} F(x,\sigma) \geqslant f(x)$.

To prove the reverse estimate, fix $x \in \mathbb{R}^n$ and consider the closed-valued multifunction $\Phi : [0,1] \to 2^I$ 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 measurable, cf. [3, Theorem 8.2.1]. 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 gives (3).

We show that f is upper-C^1. Let $\varphi(x,t) = \min_{i\in I} f_i(x,t)$ and $\Phi(x) = \int_0^1 \varphi(x,t)dt$. Every $\varphi(\cdot,t)$ is upper-C^1 as a finite minimum of upper-C^1 functions. By [6, Corollary 3] the $-\varphi(\cdot,t)$ are approximately convex in the sense of [30], and by a standard compactness argument using joint continuity of φ,

they are uniformly approximately convex over $t \in [0,1]$ in the following sense: For every $x \in \mathbb{R}^n$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all $y, z \in B(x,\delta)$ and every $t \in [0,1]$, the inequality

$$\varphi(\lambda y + (1-\lambda)z, t) \geqslant \lambda\varphi(y,t) + (1-\lambda)\varphi(z,t) - \varepsilon\lambda(1-\lambda)\|y-z\|$$

is satisfied for $0 < \lambda < 1$. Integration readily gives $\Phi(\lambda y + (1-\lambda)z) \geqslant \lambda\Phi(y) + (1-\lambda)\Phi(z) - \varepsilon\lambda(1-\lambda)\|y-z\|$, and as x and $\varepsilon > 0$ were arbitrary, this proves that $-\Phi$ is approximately convex. Again by [6, Corollary 3], $-\Phi$ is lower-C^1, and so $f = f_s + \Phi$ is upper-C^1. □

Note that the minimum (3) is semi-infinite even though I is finite. Minimization of (2) cannot be converted into an NLP, as would be possible in the min-max case. The representation (3) highlights the difficulty in minimizing (2). 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.

Remark 1 We mention that semismoothness of integral functionals of the form (2) is also discussed in [38, Section 3].

3 The model concept

The following extension of the first-order Taylor expansion to nonsmooth functions was proposed in [34], and is used to analyze nonconvex bundle methods.

Definition 1 (Compare [34]**)** A function $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a first-order model of the locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ on the set $\Omega \subset \mathbb{R}^n$ if the following axioms are satisfied:

(M_1) For every $x \in \Omega$ the function $\phi(\cdot, x) : \mathbb{R}^n \to \mathbb{R}$ is convex, $\phi(x,x) = f(x)$ and $\partial_1\phi(x,x) \subset \partial f(x)$.

(M_2) For every $x \in \Omega$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that $f(y) \leqslant \phi(y,x) + \varepsilon\|y-x\|$ for every $y \in B(x,\delta)$.

(M_3) The function ϕ is jointly upper semicontinuous, i.e., $(y_j, x_j) \to (y,x)$ on $\mathbb{R}^n \times \Omega$ implies $\limsup_{j\to\infty} \phi(y_j, x_j) \leqslant \phi(y,x)$. □

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

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

where $f^\circ(x,d)$ 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 [31,33], the standard model $\phi^\sharp$ is the smallest one. As we shall see, 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 first-order 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

($\widehat{M}_2$) For every $\varepsilon > 0$ there exists $\delta > 0$ such that $f(y) \leqslant \phi(y, x) + \varepsilon\|y - x\|$ 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 2 We may write axiom (M_2) in the form $f(y) \leqslant \phi(y, x_0) + \mathrm{o}(\|y - x_0\|)$ for $y \to x_0$, and ($\widehat{M}_2$) as $f(y) \leqslant \phi(y, x) + \mathrm{o}(\|y - x\|)$ for $x, y \to x_0$. 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 [31,33]**)** *Suppose f is upper-C^1. Then its standard model $\phi^\sharp$ is strict, and hence every model ϕ of f is strict.* □

Remark 3 For convex f the standard model $\phi^\sharp$ 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 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, and we have termed this the *natural model*. This model can be used for lower-C^2 functions in the sense of [39], lower-$C^{1,\alpha}$ functions in the sense of [7], or amenable functions in the sense of [37], 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. Nonetheless, the link between more principled cutting plane oracles and the model concept corroborates the importance of the latter.

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 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, which shall be called a first-order *working model*. Working models have of course been at the core of any bundle method and are being used since [19,16]. Here we distinguish between the ideal model ϕ, and its approximation ϕ_k, hence the terminology.

Definition 3 A first-order working model of f at the serious iterate x is a polyhedral convex function of the form

$$\phi_k(\cdot, x) = \max_{(a,g)\in\mathcal{G}_k} a + g^\top(\cdot - x), \tag{4}$$

where $\mathcal{G}_k$ is a finite set of pairs (a, g) representing affine functions $y \mapsto a + g^\top(y - x)$ satisfying $a \leqslant f(x)$, referred to as *cutting planes*. The set $\mathcal{G}_k$ is updated during the inner loop k.

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

$$\Phi_k(\cdot, x) = \phi_k(\cdot, x) + \tfrac{1}{2}(\cdot - x)^\top Q(x)(\cdot - x), \tag{5}$$

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 at x, if available.

4.2 Tangent program and acceptance test

Once the second-order working model (5) is formed, we solve the tangent program with proximity control

$$\begin{array}{ll} \text{minimize} & \Phi_k(y, x) + \frac{\tau_k}{2}\|y - x\|^2 \\ \text{subject to} & Ay \leqslant b. \end{array} \tag{6}$$

Here the proximity control parameter τ_k satisfies $Q(x)+\tau_k I \succ 0$, which assures that (6) 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

$$\rho_k = \frac{f(x) - f(y^k)}{f(x) - \Phi_k(y^k, x)} \overset{?}{\geqslant} \gamma, \tag{7}$$

where $0 < \gamma < 1$ is a fixed parameter. If $\rho_k \geqslant \gamma$, 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 the working model $\Phi_k(\cdot, x) \to \Phi_{k+1}(\cdot, x)$, adjust the proximity control parameter $\tau_k \to \tau_{k+1}$, and solve (6) again.

Note that the test (7) corresponds to the usual Armijo descent condition used in a linesearch, or to the standard acceptance test in trust region methods, see e.g. [19,16,40].

4.3 Updating the working model via aggregation

Suppose the trial step y^k fails the acceptance test (7) and is declared a null step. Then the inner loop k 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. At each inner loop step k, the following rules have to be respected when updating ϕ_k to ϕ_{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 $\mathcal{G}_{k+1}$.
- (R_2) The so-called *aggregate plane* (a^*, g^*) according to [18], which consists of convex combinations of elements of $\mathcal{G}_k$, is added to $\mathcal{G}_{k+1}$.
- (R_3) Some older planes in $\mathcal{G}_k$, which become obsolete through the addition of the aggregate plane, are discarded and not kept in $\mathcal{G}_{k+1}$.
- (R_4) 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.
- (R_5) Every working model ϕ_k satisfies $\partial_1 \phi_k(x, x) \subset \partial f(x)$.

Concerning rule (R_2), by the necessary optimality condition for (6), there exists a multiplier $\eta^* \geqslant 0$ such that

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

or equivalently,

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

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 $\mathcal{G}_k = \{(a_0, g_0), \ldots, (a_p, g_p)\}$ for $p = \operatorname{card}(\mathcal{G}_k) + 1$, we find non-negative multipliers $\lambda_0, \ldots, \lambda_p$ summing up to 1 such that

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

and in addition, $a_i + g_i^\top (y^k - x) = \phi_k(y^k, x)$ for all $i \in \{0, \ldots, p\}$ with $\lambda_i > 0$.

We now define the aggregate plane through the pair (a^*, g^*) with

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

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 [18]. It assures in particular that $\Phi_{k+1}(y^k, x) \geqslant m_k^*(y^k, x) + \frac{1}{2}(y^k - x)^\top Q(x)(y^k - x) = \Phi_k(y^k, x)$.

We say that those planes which are active at y^k in (6) are *called* by the aggregate plane. With this notion rule (R_3) can now be specified as follows.

Planes called by the aggregate plane can be removed from $\mathcal{G}_k$ and be represented by the aggregate plane, which has to be included in $\mathcal{G}_{k+1}$. Inactive planes can altogether be removed, but at least one exactness plane and the latest cutting plane at y^k have to stay in $\mathcal{G}_{k+1}$.

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 $m_k(\cdot, x) = a_k + g_k^\top(\cdot - x)$. The only requirement for the time being is that $a_k \leqslant f(x)$, as this assures $\phi_{k+1}(x, x) \leqslant f(x)$. 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 $\mathcal{G}_{k+1}$. This term designates all planes which are not based on the rules exactness, aggregation, cutting planes. Naturally, adding such planes cannot be allowed in an arbitrary way, because axioms $(R_1) - (R_5)$ have to be respected. For further details, see also [31, Example 5].

Remark 4 A typical example of anticipated cutting planes arises, e.g., when a function $f = \max\{f_i : i \in I\}$ is minimized. Here a cutting plane like $m^\downarrow(\cdot, x)$ at some unsuccessful trial step y would require a downshifted tangent of one of those branches f_i which are active at y. But it may be beneficial to also include downshifted tangents of some inactive (but nearly active) branches f_j. The importance of adding anticipated cuts was as already observed in [1, Section VI. G] for a different type of oracle.

Remark 5 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 $m_0(y, x) = f^\circ(x, y - x)$. If x is a point of strict differentiability of f, then $\partial f(x) = \{\nabla f(x)\}$ [39, Theorem 9.18], and all these exactness planes are then identical, so no extra work occurs.

In many cases computation of a subgradient $g \in \partial f(x)$ satisfying $g^\top(y - x) = f^\circ(x, y - x)$ is cheap. Consider for instance eigenvalue optimization, where $f(x) = \lambda_1(F(x))$, $x \in \mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{S}^m$, and $\lambda_1 : \mathbb{S}^m \to \mathbb{R}$ is the maximum eigenvalue function of $\mathbb{S}^m$. Then $f^\circ(x, d) = \lambda_1'(X, D) = \lambda_1(Q^\top D Q)$, 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 [5]. Then $G = QQ^\top \in \partial\lambda_1(X)$ attains $\lambda_1'(X, D)$, hence $g = F'(x)^* QQ^\top$ attains $f'(x, d)$. Since usually $t \ll m$, the computation of g is cheap.

On the other hand, situations where even computation of a single $g \in \partial f(x)$ is expensive are not unusual. This includes for instance classical applications in Lagrangian relaxation or stochastic programming. Here the use of cutting plane oracles which require only computation of a single subgradient $g \in \partial_1\phi(y, x)$, respectively, $g \in \partial f(y)$, is mandatory.

4.5 Management of proximity control

The central novelty of the bundle methods developed in [2,34,31] 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

$$\widetilde{\rho}_k = \frac{f(x) - \Phi_{k+1}(y^k, x)}{f(x) - \Phi_k(y^k, x)} \overset{?}{\geqslant} \widetilde{\gamma}, \tag{8}$$

where $0 < \gamma < \widetilde{\gamma} < 1$ are fixed. Our decision whether τ_k should be increased or not is

$$\tau_{k+1} = \begin{cases} 2\tau_k & \text{if } \widetilde{\rho}_k \geqslant \widetilde{\gamma}, \\ \tau_k & \text{if } \widetilde{\rho}_k < \widetilde{\gamma}. \end{cases} \tag{9}$$

The rationale of (9) is as follows. In case $\widetilde{\rho}_k < \widetilde{\gamma}$, improving the model by adding cutting planes may still suffice. On the other hand, if $\widetilde{\rho}_k \geqslant \widetilde{\gamma}$, then we will have to force shorter steps. Namely, the denominator in (8) 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. 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 $\phi_{k+1}(y^k, x) \geqslant \phi_k(y^k, x)$, so that $\widetilde{\rho}_k \leqslant 1$. Values $\widetilde{\rho}_k \approx 1$ indicate that little to no progress is achieved by adding the cutting plane. In this case the τ-parameter must be increased to force smaller steps, because that reinforces the agreement between f and $\phi_{k+1}(\cdot, x)$, hence increases the chances to find a serious step.

Remark 6 In the test (8) we replace $\widetilde{\rho}_k \approx 1$ by $\widetilde{\rho}_k \geqslant \widetilde{\gamma}$ for some fixed $0 < \gamma < \widetilde{\gamma} < 1$. If $\widetilde{\rho}_k < \widetilde{\gamma}$, 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 $x \to x^+$, respectively, $x^j \to x^{j+1}$. This is achieved by the test

$$\rho_{k_j} = \frac{f(x^j) - f(x^{j+1})}{f(x^j) - \Phi_{k_j}(x^{j+1}, x^j)} \overset{?}{\geqslant} \Gamma,$$

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

Remark 7 We should compare our management of the proximity control parameter τ to other strategies in the literature. For instance [29,18,24] use linesearch strategies, and [25] consider a different management of τ, which is motivated by the convex case. Our strategy assures a global convergence in the sense that every accumulation point of the sequence of serious iterates is critical.

4.6 Statement of the algorithm

We are now ready to give our formal statement of Algorithm 1 (see next page).

5 Nonconvex cutting plane oracles

In the convex cutting plane method [40,16], 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 (4) of f. For nonconvex f, cutting planes are more difficult to construct, but several ideas have been discussed, see e.g. [29,13]. In [31] we have proposed an axiomatic approach, which has the advantage that it covers most applications we are aware of, and allows a convenient convergence theory. Here we use this axiomatic approach in the convergence proof.

Definition 4 (Compare [31]**)** Let f be locally Lipschitz. A cutting plane oracle for f on the set Ω is an operator $\mathscr{O}$ 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:

(O_1) For $y = x$ we have $a = f(x)$ and $g \in \partial f(x)$.
(O_2) Suppose $y_j \to x$. Then there exist $\varepsilon_j \to 0^+$ such that $f(y_j) \leqslant m_{y_j}(y_j,x) + \varepsilon_j\|y_j - x\|$.
(O_3) Let $x_j \to x$ and $y_j, y_j^+ \to y$. Then there exists $z \in \mathbb{R}^n$ such that $\limsup_{j\to\infty} m_{y_j^+}(y_j,x_j) \leqslant m_z(y,x)$. □

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).

Definition 5 A cutting plane oracle $\mathscr{O}$ for f is called strict at x_0 if the following strict version of (O_2) is satisfied:

($\widehat{O}_2$) Suppose $y_j, x_j \to x$. Then there exist $\varepsilon_j \to 0^+$ such that $f(y_j) \leqslant m_{y_j}(y_j,x_j) + \varepsilon_j\|y_j - x_j\|$. □

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

Algorithm 1 Proximity control algorithm for (1)

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

▷ **Step 1** (`Initialize outer loop`)**.** Choose initial guess x^1 with $Ax^1 \leqslant b$ and an initial matrix $Q_1 = Q_1^\top$ with $-qI \preceq Q_1 \preceq qI$. Fix memory control parameter $\tau_1^\sharp$ such that $Q_1 + \tau_1^\sharp I \succ 0$. Put $j = 1$.

◇ **Step 2** (`Stopping test`)**.** At outer loop counter j, stop if $0 \in \partial f(x^j) + A^\top \eta^*$ for some Lagrange multiplier $\eta^* \geqslant 0$. Otherwise goto inner loop.

▷ **Step 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 $\Phi_1(\cdot, x^j) = \phi_1(\cdot, x^j) + \frac{1}{2}(\cdot - x^j)^\top Q_j(\cdot - x^j)$.

▷ **Step 4** (`Trial step generation`)**.** At inner loop counter k solve tangent program

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

The solution is the new `trial step` y^k.

◇ **Step 5** (`Acceptance test`)**.** Check whether

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

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.

▷ **Step 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 $m_k^\sharp(y^k, x^j) = f^\circ(x^j, y^k - x^j)$, 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.

◇ **Step 7** (`Update proximity parameter`)**.** Compute

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

Put

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

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

◇ **Step 8** (`Update` Q_j `and memory element`)**.** Update matrix $Q_j \to Q_{j+1}$, respecting $Q_{j+1} = Q_{j+1}^\top$ and $-qI \preceq Q_{j+1} \preceq qI$. Then store new memory element

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

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

◇ **Step 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.

Example 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 $\mathscr{O}_\phi$. Note that $\mathscr{O}_\phi$ coincides with the standard oracle [22,18] if f is convex and $\phi(\cdot,x) = f$, i.e., if the convex f is chosen as its own strict model. This means that our axiomatic approach encompasses the classical convex cutting plane method.

In more general cases, the simple idea of the model-based 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 $\mathscr{O}_\phi$ is strict as soon as the model ϕ is strict. □

Example 2 (Standard oracle) A special case of the model-based oracle is obtained by choosing the standard model $\phi^\sharp$. 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 $g^\top(y-x) = f^\circ(x, y-x)$. The standard oracle is strict iff $\phi^\sharp$ 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. This is no longer true for other models of f. □

Example 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 (4). Fixing a parameter $c > 0$, with $c = 1$ a typical value, this is organized as follows: We define the cutting plane as $m_y^\downarrow(\cdot,x) = t(\cdot) - s$, where the downshift $s \geqslant 0$ satisfies

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

Put differently, $m_y^\downarrow(\cdot,x) = a + g^\top(\cdot - x)$, where $a = \min\{t(x), f(x) - c\|y-x\|^2\}$. Note that this procedure aways 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 [31], for f lower-C^1 this is an oracle, which is automatically strict. □

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

Example 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 $f^\circ(x, y-x) = g^\top(y-x)$. Then we define

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

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. □

Suppose we are given an operator $\mathscr{O}$ which with every pair (x, y) of a serious step x and a null step y associates an abstract cutting plane $m_y(\cdot, x) = a + g^\top(\cdot - x)$. Fixing a constant $M > 0$, we define the following function

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

The crucial property of $\phi^\uparrow$ is the following

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

The proof can be found in [31]. We refer to $\phi^\uparrow$ as the *upper envelope model* associated with the oracle $\mathscr{O}$. Since in turn every model ϕ gives rise to a model-based oracle, $\mathscr{O}_\phi$, it follows that having a strict oracle and having a strict model are equivalent properties of f. Note, however, that the model $\phi^\uparrow$ is in general not practically useful. It is a theoretical tool in the convergence proof.

Remark 8 If we start with a model ϕ, then build $\mathscr{O}_\phi$, and go back to $\phi^\uparrow$, we get back to ϕ.

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

Remark 9 The definition of $\phi^\uparrow$ leaves the choice of the parameter $M > 0$ free. As $\phi^\uparrow$ is only used as a theoretical instrument in the convergence proof, we will adjust $M > 0$ such that $B(x, M)$ contains all trial steps y^k computed by the algorithm in the inner loop at serious iterate x, and this is possible under a mild coercivity assumption on f in tandem with the following Remark 10.

Remark 10 In the inner loop at serious step x boundedness of the set $\{y^k : y^k \text{ trial step at } x\}$ can be seen as follows. The necessary optimality conditions for (6) are

$$(Q + \tau_k I)(x - y^k) - v_k \in \partial_1 \phi_k(y^k, x), \quad v_k = A^\top \eta_k^*, \quad \eta_k^* \geqslant 0.$$

From the subgradient inequality we obtain

$$(x - y^k)^\top (Q + \tau_k I)(x - y^k) - v_k^\top (x - y^k) \leqslant \phi_k(x, x) - \phi_k(y^k, x).$$

By (R_4) we have $\phi_k(x,x) = f(x)$ and $\phi_k(y^k,x) \geqslant m_0(y^k,x)$, where $m_0(\cdot,x) = f(x) + g_0^\top(\cdot - x)$ with $g_0 \in \partial f(x)$ is an exactness plane. Combining with $v_k^\top(x - y^k) \leqslant 0$, we have

$$(x-y^k)^\top(Q+\tau_k I)(x-y^k) \leqslant g_0^\top(x-y^k) \leqslant \|g_0\|\|x-y^k\|.$$

As τ_k is never decreased in the inner loop, we have $Q + \tau_k I \succeq Q + \tau_1 I \succeq \kappa I$ for some $\kappa > 0$. That implies $\|x - y^k\| \leqslant \kappa^{-1}\|g_0\|$, hence boundedness of the sequence y^k.

In the outer loop a similar argument in tandem with a mild coercivity hypothesis on f implies boundedness of the set $\{y^k : y^k \text{ trial step at } x^j, j = 1,2,\ldots,\tau_k \geqslant 2q\}$. Boundedness of both sets is required to define the upper model envelope $\phi^\uparrow$ correctly.

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

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

6 Main convergence result

In this section we state and prove convergence of our algorithm when it is operated with the modified downshift oracle of Example 4.

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 $Ax^1 \leqslant b$ and $\{x \in \mathbb{R}^n : f(x) \leqslant f(x^1), Ax \leqslant b\}$ is bounded. Then every accumulation point x^* of the sequence x^j of serious iterates generated by Algorithm 1 is a Karush–Kuhn–Tucker point of (1).*

Proof The result will follow from [31, Theorem 1] as soon as we show that downshifted tangents as modified in Example 4 and used in the algorithm is a strict cutting plane oracle in the sense of Definition 5. The remainder of the proof is to verify this.

1) Let us denote cutting planes arising from the standard model $\phi^\sharp$ 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 $m_y^\downarrow(y,x) \geqslant m_y^\sharp(y,x)$, 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 $x_j \to x$, and $y_j, y_j^+ \to y$. Here y_j^+ is a null step at serious step x_j. Passing to subsequences, we may distinguish between 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 $m^\sharp_{y_j^+}(y_j, x_j) = f(x_j) + g_j^\top(y_j - x_j)$, where $g_j \in \partial f(x_j)$ satisfies $f^\circ(x_j, y_j^+ - x_j) = g_j^\top(y_j^+ - x_j)$. Passing to yet another subsequence, we may assume $g_j \to g$, and upper semi-continuity of the Clarke subdifferential gives $g \in \partial f(x)$. Therefore $m_{y_j^+}(y_j, x_j) = f(x_j) + g_j^\top(y_j - x_j) \to f(x) + g^\top(y - x) \leqslant m_y^\sharp(y, x) \leqslant m_y(y, x)$. So here (O_3) is satisfied with $z = y$.

Next consider case II. Here we have $m_{y_j^+}(y_j, x_j) = t_{g_j}(y_j) - s_j$, 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

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

Passing to a subsequence, we may assume $g_j \to g$, and using $y_j^+ \to y$ in tandem with upper semi-continuity of ∂f, we deduce $g \in \partial f(y)$. Therefore, $s_j \to \left[t_g(x) - f(x) + c\|y - x\|^2\right]_+ =: s$, where uniform convergence $t_{g_j}(y_j) \to t_g(y)$ 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 $m_{y_j^+}(y_j, x_j) \to m_y^\downarrow(y, x)$, and since $m_y^\downarrow(y, x) \leqslant m_y(y, x)$, we are done. So again the z in (O_3) equals y here.

4) Let us finally check axiom $(\widehat{O}_2)$. Let $x_j, y_j \to x$ be given. We first consider the case when f is upper-C^1 at x. We have to find $\varepsilon_j \to 0^+$ such that $f(y_j) \leqslant m_{y_j}(y_j, x_j) + \varepsilon_j\|y_j - x_j\|$ as $j \to \infty$, and by the definition of the oracle, it clearly suffices to show $f(y_j) \leqslant m^\sharp_{y_j}(y_j, x_j) + \varepsilon_j\|y_j - x_j\|$. By Spingarn [42], or Daniilidis and Georgiev [6], $-f$, which is lower-C^1 at x, has the following property: For every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 < t < 1$ and $y, z \in B(x, \delta)$,

$$f(y) \leqslant f(z) + t^{-1}\left(f(z + t(y - z)) - f(z)\right) + \varepsilon(1 - t)\|z - y\|.$$

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

$$f(y) \leqslant f(z) + f'(z, y - z) + \varepsilon\|y - z\| \leqslant f(z) + f^\circ(z, y - z) + \varepsilon\|y - z\|.$$

Choosing $z = x_j$, $y = y_j$, $\delta_j = \|y_j - z_j\| \to 0$, we can find $\varepsilon_j \to 0^+$ such that $f(y_j) \leqslant f(x_j) + f^\circ(x_j, y_j - x_j) + \varepsilon_j\|y_j - x_j\|$, hence $f(y_j) \leqslant m^\sharp_{y_j}(y_j, x_j) + \varepsilon_j\|y_j - x_j\|$ by the definition of $m^\sharp_{y_j}(\cdot, x_j)$. That settles the upper-C^1 case.

Now consider the case where f is lower-C^1 at x. We have to find $\varepsilon_j \to 0^+$ such that $f(y_j) \leqslant m_{y_j}(y_j, x_j) + \varepsilon_j\|y_j - x_j\|$ as $j \to \infty$, and it suffices to show $f(y_j) \leqslant m^\downarrow_{y_j}(y_j, x_j) + \varepsilon_j\|y_j - x_j\|$. Since $m^\downarrow_{y_j}(y_j, x_j) \geqslant f(y_j) - s_j$, where s_j is the downshift $s_j = \left[t(x_j) - f(x_j) + c\|y_j - x_j\|^2\right]_+$, 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 $\varepsilon_j \to 0^+$ such that $f(y_j) \leqslant f(y_j) - s_j + \varepsilon_j\|y_j - x_j\|$, or equivalently, $s_j \leqslant \varepsilon_j\|y_j - x_j\|$. For that it suffices to arrange $[t(x_j) - f(x_j)]_+ \leqslant \varepsilon_j\|y_j - x_j\|$, because once this is verified, we get

$$s_j \leqslant [t(x_j) - f(x_j)]_+ + c\|y_j - x_j\|^2 \leqslant \widetilde{\varepsilon}_j\|y_j - x_j\|$$

with $\widetilde{\varepsilon}_j := \varepsilon_j + c\|y_j - x_j\|$. Note again that by [42,6], f has the following property at x: For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $y, z \in B(x, \delta)$

$$f(tz + (1-t)y) \leqslant tf(z) + (1-t)f(y) + \varepsilon t(1-t)\|z - y\|.$$

Dividing by $t > 0$ and passing to the limit $t \to 0^+$ gives $f^\circ(y, z-y) \leqslant f(z) - f(y) + \varepsilon\|y - z\|$, using the fact that f is locally Lipschitz. But for every $g \in \partial f(y)$, $g^\top(z-y) \leqslant f^\circ(y, z-y)$. Using $\|y_j - x_j\| =: \delta_j \to 0$ and taking $y = y_j$, $z = x_j$, this allows us to find $\varepsilon_j \to 0^+$ such that $g_j^\top(x_j - y_j) \leqslant f(x_j) - f(y_j) + \varepsilon_j\|y_j - x_j\|$. Substituting this above gives $t(x_j) - f(x_j) = f(y_j) - f(x_j) + g_j^\top(x_j - y_j) \leqslant \varepsilon_j\|y_j - x_j\|$ as desired. That settles the lower-C^1 case. □

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

$$\frac{|f(x^{j+1}) - f(x^j)|}{1 + |f(x^j)|} < \mathrm{tol}_1, \quad \frac{\|P_C(-g_{j+1}^*)\|}{1 + |f(x^{j+1})|} < \mathrm{tol}_2, \quad \frac{\|x^{j+1} - x^j\|}{1 + \|x^j\|} < \mathrm{tol}_3, \quad (10)$$

then we decide that x^{j+1} is optimal. Here g_{j+1}^* is the aggregate subgradient at acceptance of x^{j+1}, and P_C is the projection operator onto the constraint set $C = \{x : Ax \leqslant b\}$. 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 $k_{\min}$ consecutive null steps y^k, where

$$\frac{|f(y^k) - f(x^j)|}{1 + |f(x^j)|} < \mathrm{tol}_1, \quad \frac{\|P_C(-g_k^*)\|}{1 + |f(x^j)|} < \mathrm{tol}_2, \quad \frac{\|y^k - x^j\|}{1 + \|x^j\|} < \mathrm{tol}_3, \quad (11)$$

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 $\mathrm{tol}_1 = 10^{-3}$, $\mathrm{tol}_2 = 10^{-2}$, $\mathrm{tol}_3 = 10^{-3}$, $k_{\min} = 3$, and $k_{\max} = 50$. Typical values in Algorithm 1 are $\gamma = 0.01$, $\widetilde{\gamma} = 0.4$ and $\Gamma = 0.6$.

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

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

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 accomodate $(a - s + g^\top(x^{j+1} - x^j), g) \in \mathcal{G}_1$ 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 $g^\top(x^j - x^{j+1}) = f^\circ(x^{j+1}, x^j - x^{j+1})$. Indeed, when $m(x^j, x^{j+1}) \geqslant m_0(x^j, x^{j+1})$, then we keep the downshifted plane, otherwise we add $m_0(\cdot, x^{j+1})$ as additional exactness plane.

8 Application to delamination benchmark problem

8.1 Modeling of delamination

The interface behavior of bonded laminated composite benchmark structures is usually determined experimentally using the double cantilever beam test [43] or other destructive testing methods. The result of a typical experiment is shown schematically in Fig. 1 from [43], where three probes with different levels of contamination of the adhesive are displayed, starting with the critical load $P = 150\,\mathrm{N}$. While the intact material shows stable propagation of the crack front (dashed curve), the specimen with 10% contamination on average shows a typical zig-zag profile (bold solid curve), indicating unstable crack front propagation. Indeed, when reaching the critical load, here $P = 140\,\mathrm{N}$, a crack in the adhesive layer opens at the loaded tip of the double beam and propagates starting with initial crack length 0 mm. By the growth of the crack-elongation, the compliance of the structure increases. When the crack meets a zone of contamination, the load P in the structure drops immediately, here from $P = 140\,\mathrm{N}$ to $P = 40\,\mathrm{N}$. At the latest when leaving the zone of

contamination, the crack propagation slows down and the crack is *caught*, i.e., stops at $u_2 = 0.25\,\mathrm{mm}$. Thereafter, by the constantly applied traction force, there is a linear growth of the load from $P = 40\,\mathrm{N}$ to the critical load $P = 90\,\mathrm{N}$, where the crack starts again to propagate and stops at $u_2 = 5\,\mathrm{mm}$ with the load now reduced to $P = 30\,\mathrm{N}$. This phenomenon occurs five to six times, as seen in Fig. 1. Note that in the material sciences, knowledge of the interface behavior is crucial for the understanding of the basic failure modes of real-world bonded composite structures in aerospace and automotive industry.

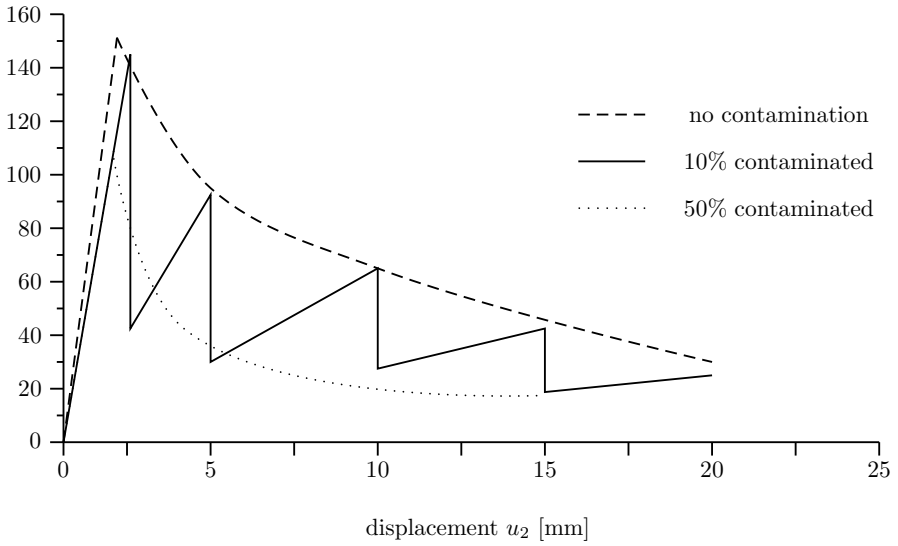

Fig. 1 Load-displacement curve determined by double cantilever beam test. Dashed curve shows stable behavior for material without contamination. The 10% contaminated specimen (bold solid curve) shows unstable crack growth. The 50% contaminated specimen exhibits micro-cracks not visible at the chosen scale. The adhesive energy, which is represented by the area below the load-displacement curve, indicates that a contaminated specimen is of minor resistance.

To describe the process of structural adhesive bonding in mathematical terms we use the tools of nonsmooth analysis. We model the interface behavior of a bonded composite structure by a non-monotone multivalued law, which can be expressed by means of the Clarke subdifferential of a nonconvex nonsmooth, but locally Lipschitz function. More precisely, the physical law governing the relation between the normal component $S_n(s)|\Gamma_c$ of the boundary stress vector on the one hand, and the relative displacement $u_2(s)|\Gamma_c$, or *jump*, between the upper and lower boundaries on the other hand, is described by

$$-S_n(s) \in \partial j(s, u_2(s)),\ s \in \Gamma_c, \tag{12}$$

where j is a locally Lipschitz function. A typical law ∂j for an interlayer adhesive is shown in Fig. 2 (left). The superpotential j of ∂j, which is a

locally Lipschitz minimum-type function, is the nonsmooth part of the total potential energy functional of the composite structure.

Even though the displacement u_2 in Fig. 1 can only be measured at the crack tip, in order to proceed we *stipulate* the law ∂j along the whole boundary Γ_c. Under this hypothesis we minimize the discretized total potential energy and thus compute the unknown displacement and stress fields of the entire mechanical structure. For validation of the adhesive law the numerical results at the crack tip can be compared with experimental data using the initial crack length as variable input parameter of the delamination process.

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 [43]. In contrast, our approach should be understood as an *estimation technique* of the process of delamination including contamination effects, that is based on a rigorous mathematical contact model from continuum mechanics.

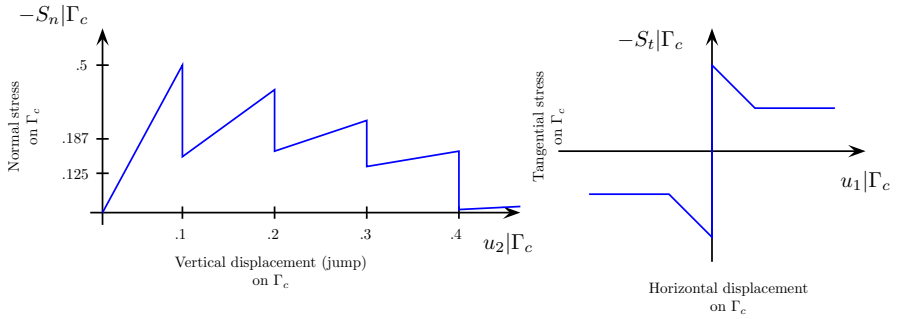

Fig. 2 Left image shows non-monotone delamination law ∂j, leading to an upper-C^1 objective. Right image shows non-monotone friction law, leading to a lower-C^1 objective

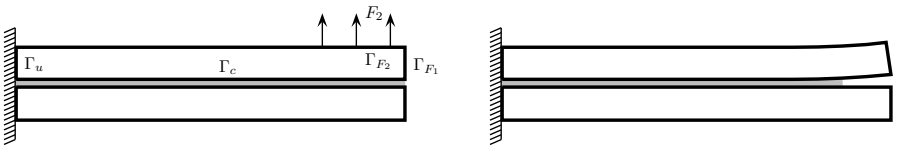

Fig. 3 Schematic view of cantilever beam testing. Under applied traction force F_2 the crack front propagates to the left. In program (17) traction force F_2 and crack front length are given, while the corresponding displacement u and reactive forces $-S_n|\Gamma_c$ along the contact boundary Γ_c have to be computed

8.2 Study of a delamination problem - continuous formulation

Within the framework of plane linear elasticity we consider a two-dimensional symmetric laminated structure with an interlayer adhesive under loading (see Fig. 3). Because of the symmetry of the structure and by assuming that the forces applied to the upper and lower part are the same, it suffices to consider only the upper half of the specimen, represented by $\overline{\Omega} \subset \mathbb{R}^2$. The Lipschitz boundary Γ of the bounded domain Ω consists of four disjoint parts Γ_u, Γ_c, Γ_{F_1} and Γ_{F_2}. We adopt the standard notations from linear elasticity (see e.g. [17]) and introduce the bilinear form

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

where $\mathbf{u} = (u_1, u_2)$ is the displacement vector decomposed in horizontal and vertical displacements, and $\varepsilon(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)$, $\sigma(\mathbf{v}) = \mathbf{C} : \varepsilon(\mathbf{v})$ stand for the linearized strain tensor and the stress tensor, respectively. Here, $\mathbf{C}$ is the Hooke tensor, assumed to be uniformly positive definite with $L^\infty(\Omega)$-coefficients. The bilinear form of linear elasticity (13) is symmetric and due to the first Korn inequality, coercive. The equation of the equilibrium state of Ω is defined as follows

$$\sigma_{ij,j} = 0 \text{ in } \Omega,\ i = 1, 2, \tag{14}$$

since no volume forces are assumed. In our computations we use

$$\sigma_{ij}(\mathbf{u}) = \frac{E\nu}{1-\nu^2}\left(\varepsilon_{11}(\mathbf{u}) + \varepsilon_{22}(\mathbf{u})\right)\delta_{ij} + \frac{E}{1+\nu}\varepsilon_{ij}(\mathbf{u}),$$

where E and ν are the Young modulus and the Poisson's ratio, respectively.

Next, we define the boundary conditions. On $\Gamma_{F_1} \cup \Gamma_{F_2}$ the traction forces are prescribed

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

where F_2 is a constant force. The linear form $\langle \mathbf{g}, \cdot \rangle$ is now defined as

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

Next, the body is fixed along the boundary Γ_u:

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

and because of the non-penetration of the laminae we have also the unilateral constraint

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

in the vertical direction along the contact boundary.

Furthermore, we assume that the non-monotone multivalued law ∂j, displayed schematically in Fig. 2 (left), describes the behavior of the binding

interlayer material along Γ_c. We also assume that the tangential component of the boundary stress vector is negligible, i.e., $\mathbf{S}_t(s) = 0$. In our computations, the superpotential j of ∂j is a minimum of four convex quadratic and one linear function.

The normal component S_n and the tangential component S_t of the stress vector $\mathbf{S}$ are defined, respectively, by $S_n = \sigma_{ij} n_j n_i$ and $S_{t_i} = \sigma_{ij} n_j - S_n\, n_i$, where $\mathbf{n} = (n_1, n_2)$ is the outward unit normal vector to Γ_c.

We define the space

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

Let $K \subset V$ be the nonempty, closed and convex set of all admissible displacements defined by

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

Due to the symmetry of the bilinear form, the potential energy of the considered mechanical problem has the form

$$\Pi(\mathbf{v}) = \tfrac{1}{2} a(\mathbf{v}, \mathbf{v}) + J(\mathbf{v}) - \langle \mathbf{g}, \mathbf{v} \rangle, \tag{15}$$

where $J : V \to \mathbb{R}$ is the nonsmooth functional defined by

$$J(\mathbf{v}) = \int_{\Gamma_c} j(s, v_2(s))\, ds. \tag{16}$$

Using the potential energy functional, we can formulate the following nonsmooth, nonconvex constrained optimization problem:

$$\begin{aligned} &\text{minimize} \ \ \Pi(\mathbf{v}) \\ &\text{subject to } \mathbf{v} \in K, \end{aligned} \tag{17}$$

where the objective is upper-C^1 locally Lipschitz function, because the superpotential j is a minimum-type function. In particular, we have an objective of the form (2), where the smooth part f_s comprises $\frac{1}{2} a(\mathbf{v}, \mathbf{v}) - \langle \mathbf{g}, \mathbf{v} \rangle$, while $J(\mathbf{v}) = \int_{\Gamma_c} j(s, v_2(s))\, ds$ is the nonsmooth part in (2).

Taking into account the nonconvexity and the nonsmoothness of the potential energy functional, we can only pose critical point problems, i.e., find $\mathbf{u}^* \in K$ such that

$$0 \in \partial\, \Pi(\mathbf{u}^*) + N_K(\mathbf{u}^*), \tag{18}$$

where $N_K(\mathbf{u}^*)$ is the normal cone to K at $\mathbf{u}^*$. Moreover, according to [27] every local minimizer of the problem (17) is a critical point of Π on K in the sense of (18).

8.3 Discrete problem

We consider a regular triangulation $\{\mathcal{T}_h\}$ of $\overline{\Omega}$, where we first divide $\overline{\Omega}$ into small rectangles and then each rectangle by its diagonal into two triangles. To approximate V and K we use a piecewise linear finite element approximation and set V_h and K_h, respectively, by

$$V_h = \{v_h \in C(\overline{\Omega};\mathbb{R}^2) : v_{h|_T} \in (\mathbf{P}_1)^2 \quad \forall\, T \in \mathcal{T}_h,\, v_{h|_{\Gamma_u}} = 0\},$$
$$K_h = \{v_h \in V_h : v_{h2}(s_\nu) \geqslant 0 \quad \forall\, s_\nu \in \Gamma_c\backslash\Gamma_u\}.$$

Here, $\mathbf{P}_1$ is the set of all polynomials of degree one and s_ν are the nodes of $\mathcal{T}_h$ lying on $\Gamma_c\backslash\Gamma_u$. Similar to low order finite element approximations of nonsmooth convex contact problems [10,11], we use the trapezoidal quadrature rule to approximate the functional J in (16) by

$$J_h(v_h) = \frac{1}{2} \sum_{s_\nu \in \Gamma_c\backslash\Gamma_u} |s_\nu s_{\nu+1}| \left[j(s_\nu, v_{h2}(s_\nu)) + j(s_{\nu+1}, v_{h2}(s_{\nu+1})) \right],$$

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

$$J_h(v_h) = \sum_{s_\nu \in \Gamma_c\backslash\Gamma_u} c_\nu j(s_\nu, v_{h2}(s_\nu)) = \sum_{s_\nu \in \Gamma_c\backslash\Gamma_u} c_\nu \min_{i\in I} j_i(s_\nu, v_{h2}(s_\nu))$$

with appropriate weights $c_\nu > 0$. Here, I indexes the smooth branches contributing to the nonsmooth function $j(\cdot)$, so for instance in the case of Fig. 2 (left) with $|I| = 5$, this corresponds to four zig-zags in the graph of ∂j.

The bundle algorithm is applied to minimize the discrete functional

$$\Pi_h(v_h) = \tfrac{1}{2} a(v_h, v_h) + J_h(v_h) - \langle g, v_h\rangle \quad \text{on} \quad K_h, \tag{19}$$

where we pull out the minimum from under the sum over the N quadrature nodes s_ν, which leads to the expression

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

which is the discretized version of (3). In (19), $\frac{1}{2} a(v_h, v_h) - \langle g, v_h\rangle$ is the smooth term and J_h is the nonsmooth part.

The discrete critical point problem on K_h reads as follows: Find $u_h \in K_h$ such that

$$0 \in \partial\Pi_h(u_h) + N_{K_h}(u_h). \tag{20}$$

Due to the results in [27], the discrete problem (20) is solvable and its solutions converge in subsequences to the solutions of the continuous critical point problem (18).

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 $\langle g, v_h\rangle =$

$\mathbf{g}^\top v_h$, a Clarke subgradient of (19) at v_h is readily computed as $\mathbf{A}v_h + \sum_{\nu\in N} \nabla j_{i(\nu)}(s_\nu, v_{h2}(s_\nu)) - g$, 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. Similarly, the matrix $Q = Q(v)$ in the second-order working model (5) is chosen as $Q(v) = \mathbf{A} + \sum_{\nu\in N} \nabla^2 j_{i(\nu)}(s_\nu, v_h(s_\nu))$, where $i(\nu) \in I$ is the same active index.

Note that here we use the lowest-order finite element approximation which satisfies $K_h \subset K$ and is thus conforming. Higher-order approximations with no limitation in the polynomial degree, leading to nonconforming approximation of unilateral constraints, have only recently been analyzed for monotone contact problems, see [12].

8.4 Numerical results

We present numerical results obtained in a delamination simulation of a symmetric system of two bodies in an adhesive contact. The used material is characterized by the modulus of elasticity $E = 210\,\mathrm{GPa}$ and Poisson's ratio $\nu = 0.3$ corresponding to a steel specimen. In all examples we use the benchmark model of [4] (see Fig. 3 (left)) with geometrical characteristics $\Omega = (0, 100)\times(0, 10)$ in [mm]. The applied loads F_2 are $0.2, 0.4, 0.6, 0.8, 1.0\,\mathrm{N/mm^2}$, respectively. The volume forces are neglected. The non-monotone adhesive law is defined by the multivalued function shown in Fig. 2 (left). We apply our bundle method to (19) and compare the results to those obtained by the regularization method in [35,36]. All computations use piecewise linear functions on a 40×4 triangulation $\mathcal{T}_h$ of $\overline{\Omega}$, corresponding to the step length $h = 2.5$ in [mm] along the boundary Γ_c. Thus, the number of the unknowns on the boundary in the discrete problem (19) is 80. All tests have been performed with MATLAB R2013b on OS Windows 7 Home Premium with CPU Intel Core i5-2410M (2.30 GHz) and 4GB of RAM.

We note that in our experiments, the bundle algorithm usually terminates due to the stopping test (10). For example, in the case $F_2 = 1\,\mathrm{N/mm^2}$ with the starting point 0.0 in all the components, Algorithm 1 terminated based on (10), where the quantities in (10) are respectively 2.2656×10^{-9}, 3.2573×10^{-3}, 5.9716×10^{-8}. Fig. 4 displays a typical performance profile of algorithm 1.

For comparison we use the regularization method from [35,36]. The regularization parameter ε is set to $\varepsilon = 0.1$, and the discrete regularized problem is solved using the following steps. Firstly, (a) we use a condensation technique based on a Schur complement to reduce the total number of unknowns to obtain a problem formulated only in terms of the displacements at the free nodes on the contact boundary Γ_c. Secondly, (b) the reduced discrete regularized problem is re-written as a mixed complementarity problem, which by means of the Fischer–Burmeister function $f(a, b) = \sqrt{a^2 + b^2} - (a + b)$ is reformulated as a system of nonlinear equations. Finally, (c) by applying an appropriate merit function we obtain an equivalent smooth, unconstrained minimization problem, which is numerically solved by using the *lsqnonlin* -

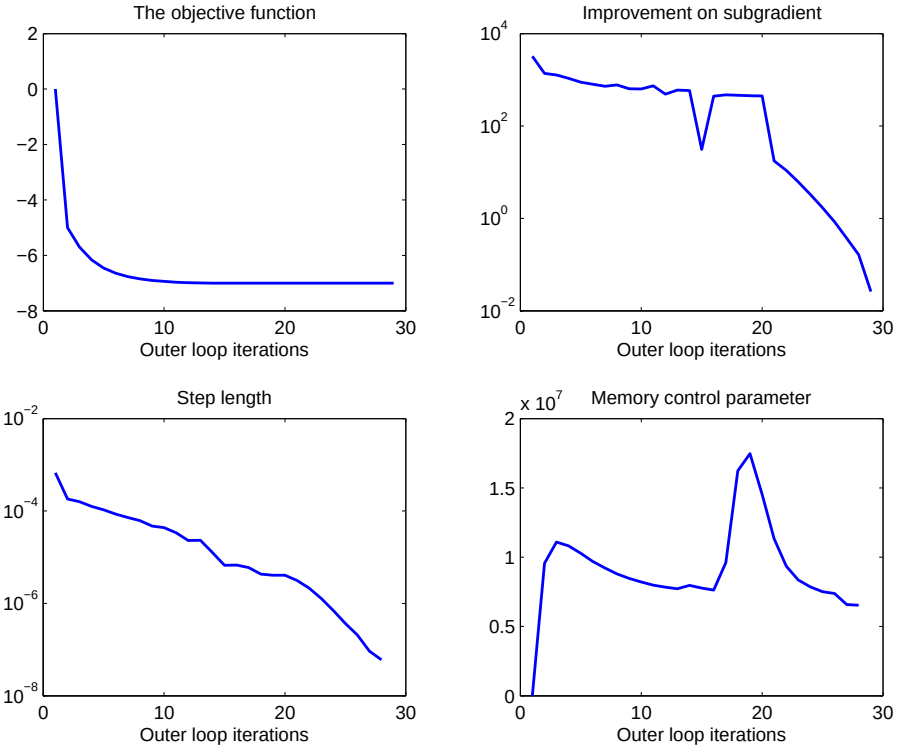

Fig. 4 Performance of Algorithm 1 for a typical run of (19) with 160 unknowns in the case $F_2 = 1\,\mathrm{N/mm^2}$. *Top left* $j \mapsto f(x^j)$. *Top right* $j \mapsto \|P_C(-g_j^*)\|$. *Lower left* $j \mapsto \|x^j - x^{j+1}\|$. *Lower right* $j \mapsto \tau_j^\sharp$, the evolution of the memory control parameter at serious steps

MATLAB function based on a trust region method. The maximal number of iterations in *lsqnonlin* has been fixed to 100. While theoretical convergence of the regularization method occurs with $\varepsilon \to 0$, for details see [35], the choice $\varepsilon = 0.1$ is based on the observation that smaller values, while increasing the CPU times, do not further improve the solution from a mechanics point of view.

The computed values of the vertical and horizontal displacements of the upper lamina on Γ_c obtained by our bundle method and by the regularization method at four intermediate points on Γ_c, $(25, 0), (50, 0), (75, 0), (100, 0)$ in [mm], are presented in Tables 1 and 2. As default, the starting point is 0.1 in all components. Vertical displacements and normal stresses along Γ_c obtained by both methods are compared in Fig. 5 for different loads $F_2 = 0.2, 0.6, 1.0\,\mathrm{N/mm^2}$. Fig. 5 shows qualitatively similar results.

Tables 3–5 give the final results obtained by our bundle method and the regularization technique for 3 different starting points, where horizontal and vertical displacements at the grid points on the boundary are chosen as $0.0, 0.1, 1.0$'s, respectively. We compare optimal objective function values

Table 1 Vertical displacement u_2 in [mm] at four intermediate points for 5 different scenarios obtained by bundle and regularization method

$u_2(x_1)$		$u_2(x_2)$		$u_2(x_3)$		$u_2(x_4)$	
Bundle	Reg.	Bundle	Reg.	Bundle	Reg.	Bundle	Reg.
4.9482e-06	3.3519e-06	1.5175e-05	9.8040e-06	2.7094e-05	1.6938e-05	3.9281e-05	2.4120e-05
9.8965e-06	8.0897e-06	3.0350e-05	2.5206e-05	5.4189e-05	4.6307e-05	7.8562e-05	6.8888e-05
1.9562e-05	3.0080e-05	6.2890e-05	1.0121e-04	1.1680e-04	1.9327e-04	1.7354e-04	2.9167e-04
3.4125e-05	4.6784e-05	1.1216e-04	1.5649e-04	2.1098e-04	2.9665e-04	3.1583e-04	4.4538e-04
5.3916e-05	6.1456e-05	1.7939e-04	2.0452e-04	3.3982e-04	3.8604e-04	5.1063e-04	5.7827e-04

From top to bottom, $F_2 = 0.2, 0.4, 0.6, 0.8, 1.0\,\mathrm{N/mm^2}$, respectively

Table 2 Horizontal displacement u_1 in [mm] at four intermediate points for 5 different scenarios obtained by bundle and regularization method

$u_1(x_1)$		$u_1(x_2)$		$u_1(x_3)$		$u_1(x_4)$	
Bundle	Reg.	Bundle	Reg.	Bundle	Reg.	Bundle	Reg.
1.6362e-06	1.0674e-06	2.2616e-06	1.3801e-06	2.4081e-06	1.4164e-06	2.4124e-06	1.4128e-06
3.2723e-06	2.6925e-06	4.5232e-06	3.8784e-06	4.8162e-06	4.3860e-06	4.8248e-06	4.4818e-06
6.6925e-06	1.0658e-05	9.9396e-06	1.6727e-05	1.1139e-05	1.9253e-05	1.1291e-05	1.9654e-05
1.1879e-05	1.6532e-05	1.8113e-05	2.5639e-05	2.0543e-05	2.9175e-05	2.0922e-05	2.9711e-05
1.8958e-05	2.1644e-05	2.9305e-05	3.3306e-05	3.3458e-05	3.7733e-05	3.4122e-05	3.8403e-05

From top to bottom, $F_2 = 0.2, 0.4, 0.6, 0.8, 1.0\,\mathrm{N/mm^2}$, respectively

and running times in seconds. The computations show that both methods are numerically stable and reliable. Even for very different starting points we obtain almost identical final results.

Table 3 Comparison of bundle and regularization method. Starting point all 0.0's

$F_2[\mathrm{N/mm^2}]$	Initial Π_h	Optimal Π_h		Running time	
		Bundle	Reg.	Bundle	Reg.
0.2	0.000346	-0.169337	-0.145855	6.264	11.6895
0.4	0.000346	-0.678386	-0.584459	0.148	12.1158
0.6	0.000346	-1.684544	-1.605770	5.656	12.2706
0.8	0.000346	-3.633644	-3.199650	0.802	12.1733
1.0	0.000346	-7.006673	-5.768180	0.810	12.0655

9 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

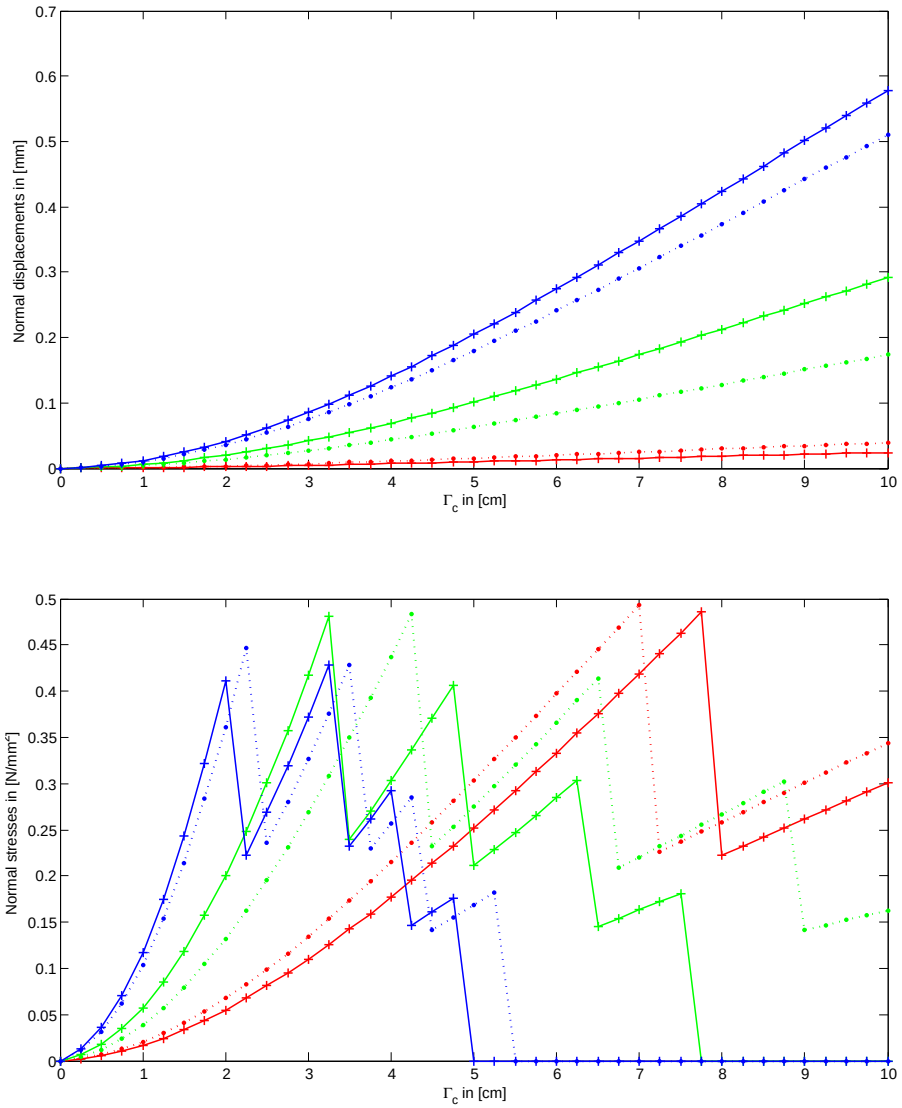

Fig. 5 Comparison of bundle method (*dotted*) and regularization method (*solid*) for 3 different loads $F_2 = 0.2\,\mathrm{N/mm^2}$ (*red*), $0.6\,\mathrm{N/mm^2}$ (*green*), $1.0\,\mathrm{N/mm^2}$ (*blue*) in vertical displacement on Γ_c (*top*) and normal component of the stress vector along Γ_c (*lower*)

method allows to integrate second-order information, if available, and gives a convergence certificate in the 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

Table 4 Comparison of bundle and regularization method. Starting point all 0.1's

$F_2[\mathrm{N/mm^2}]$	Initial Π_h	Optimal Π_h		Running time	
		Bundle	Reg.	Bundle	Reg.
0.2	1934543393.554280	-0.169337	-0.14587	4.998	2.6249
0.4	1934541479.777698	-0.678386	-0.661415	4.959	2.3922
0.6	1934539566.001117	-1.684020	-1.050940	0.555	2.4006
0.8	1934537652.224535	-3.635792	-3.037370	0.829	2.3527
1.0	1934535738.447954	-7.006673	-6.669180	0.429	2.4976

Table 5 Comparison of bundle and regularization method. Starting point all 1.0's

$F_2[\mathrm{N/mm^2}]$	Initial Π_h	Optimal Π_h		Running time	
		Bundle	Reg.	Bundle	Reg.
0.2	193454510602.708070	-0.169337	-0.145998	4.845	2.5282
0.4	193454491464.942230	-0.678386	-0.656331	4.707	2.4290
0.6	193454472327.176420	-1.684020	-1.049730	0.562	2.4911
0.8	193454453189.410610	-3.635792	-3.036380	0.858	2.4273
1.0	193454434051.644810	-7.006673	-6.668540	0.437	2.4023

the material sciences, where upper-C^1 functions have to be minimized. Results obtained by our nonsmooth optimization method were compared to those obtained by the regularization technique of [35,36], and both methods are, in general, in good agreement.

We have observed that practitioners often feel uncomfortable with the regularization approach, considering it *in extremis* as of tampering with the problem. Since our nonsmooth approach deals with the problem in its original form, one may therefore argue that the regularization technique is validated through the nonsmooth method.

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

References

1. Apkarian, P., Noll, D.: Nonsmooth H_∞ synthesis. IEEE Trans. Automat. Control **51**(1), 71–86 (2006)
2. Apkarian, P., Noll, D., Prot, O.: A trust region spectral bundle method for nonconvex eigenvalue optimization. SIAM J. Optim. **19**(1), 281–306 (2008)
3. Aubin, J.P., Frankowska, H.: Set-valued analysis. Mod. Birkhäuser Class. Birkhäuser Boston, Boston (2009)
4. Baniotopoulos, C.C., Haslinger, J., Morávková, Z.: Mathematical modeling of delamination and nonmonotone friction problems by hemivariational inequalities. Appl. Math. **50**(1), 1–25 (2005)
5. Cullum, J., Donath, W.E., Wolfe, P.: The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. In: M.L. Balinski, P. Wolfe (eds.) Nondifferentiable Optimization, *Math. Programming Stud.*, vol. 3, pp. 35–55. North-Holland, Amsterdam (1975)
6. Daniilidis, A., Georgiev, P.: Approximate convexity and submonotonicity. J. Math Anal. Appl. **291**(1), 292–301 (2004)
7. Daniilidis, A., Malick, J.: Filling the gap between lower-C^1 and lower-C^2 functions. J. Convex Anal. **12**(2), 315–329 (2005)

8. Dao, M.N., Noll, D.: Minimizing memory effects of a system. Math. Control Signals Syst. **27**(1), 77–110 (2015)
9. Dao, M.N., Noll, D., Apkarian, P.: Robust eigenstructure clustering by non-smooth optimisation. Int. J. Control **88**(8), 1441–1455 (2015)
10. Glowinski, R.: Numerical methods for nonlinear variational problems. Springer Ser. Comput. Phys. Springer-Verlag, New York (1984)
11. Gwinner, J.: Finite-element convergence for contact problems in plane linear elastostatics. Quart. Appl. Math. **50**(1), 11–25 (1992)
12. Gwinner, J.: *hp*-FEM convergence for unilateral contact problems with Tresca friction in plane linear elastostatics. J. Comput. Appl. Math. **254**, 175–184 (2013)
13. Hare, W., Sagastizábal, C.: Computing proximal points of nonconvex functions. Math. Program., Ser. B **116**(1-2), 221–258 (2009)
14. Hare, W., Sagastizábal, C., Solodov, M.: A proximal bundle method for nonsmooth nonconvex functions with inexact information. Comput. Optim. Appl. **63**(1), 1–28 (2016)
15. Haslinger, J., Miettinen, M., Panagiotopoulos, P.D.: Finite Element Method for Hemivariational Inequalities. Theory, Methods and Applications, *Nonconvex Optim. Appl.*, vol. 35. Kluwer Academic, Dordrecht (1999)
16. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms, Vol. I. Fundamentals, Vol. II. Advanced Theory and Bundle Methods, *Grundlehren Math. Wiss.*, vol. 305-306. Springer-Verlag, Berlin (1993)
17. Kikuchi, N., Oden, J.T.: Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods, *SIAM Stud. Appl. Math.*, vol. 8. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1988)
18. Kiwiel, K.C.: An aggregate subgradient method for nonsmooth convex minimization. Math. Program. **27**(3), 320–341 (1983)
19. Kiwiel, K.C.: Methods of Descent for Nondifferentiable Optimization, *Lecture Notes in Math.*, vol. 1133. Springer-Verlag, Berlin (1985)
20. Kiwiel, K.C.: A proximal bundle method with approximate subgradient linearizations. SIAM J. Optim. **16**(4), 1007–1023 (2006)
21. Kovtunenko, V.A.: A hemivariational inequality in crack problems. Optimization **60**(8-9), 1071–1089 (2011)
22. Lemaréchal, C.: Bundle methods in nonsmooth optimization. In: C. Lemaréchal, R. Mifflin (eds.) Nonsmooth Optimization (Laxenburg, 1977), *IIASA Proc. Ser.*, vol. 3, pp. 79–102. Pergamon Press, Oxford (1978)
23. Lemaréchal, C., Sagastizábal, C.: Variable metric bundle methods: from conceptual to implementable forms. Math. Program., Ser. B **76**(3), 393–410 (1997)
24. Lukšan, L., Vlček, J.: A bundle-Newton method for nonsmooth unconstrained minimization. Math. Program., Ser. A **83**(3), 373–391 (1998)
25. Mäkelä, M.M., Miettinen, M., Lukšan, L., Vlček, J.: Comparing nonsmooth nonconvex bundle methods in solving hemivariational inequalities. J. Global Optim. **14**(2), 117–135 (1999)
26. Mäkelä, M.M., Neittaanmäki, P.: Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control. World Scientific, Singapore (1992)
27. Miettinen, M., Mäkelä, M.M., Haslinger, J.: On numerical solution of hemivariational inequalities by nonsmooth optimization methods. J. Global Optim. **6**(4), 401–425 (1995)
28. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**(6), 959–972 (1977)
29. Mifflin, R.: A modification and extension of Lemarechal's algorithm for nonsmooth minimization. In: D.C. Sorensen, R.J.B. Wets (eds.) Nondifferential and Variational Techniques in Optimization (Lexington, 1980), *Math. Programming Stud.*, vol. 17, pp. 77–90. North-Holland, Amsterdam (1982)
30. Ngai, H.V., Luc, D.T., Théra, M.: Approximate convex functions. J. Nonlinear Convex Anal. **1**(2), 155–176 (2000)
31. Noll, D.: Cutting plane oracles to minimize non-smooth non-convex functions. Set-Valued Var. Anal. **18**(3-4), 531–568 (2010)
32. Noll, D.: Bundle method for non-convex minimization with inexact subgradients and function values. In: D.H.B. et al. (ed.) Computational and Analytical Mathematics, *Springer Proc. Math. Stat.*, vol. 50, pp. 555–592. Springer, New York (2013)

33. Noll, D.: Convergence of non-smooth descent methods using the Kurdyka-Łojasiewicz inequality. J. Optim. Theory Appl. **160**(2), 553–572 (2014)
34. Noll, D., Prot, O., Rondepierre, A.: A proximity control algorithm to minimize nonsmooth and nonconvex functions. Pac. J. Optim. **4**(3), 571–604 (2008)
35. Ovcharova, N.: Regularization Methods and Finite Element Approximation of Hemivariational Inequalities with Applications to Nonmonotone Contact Problems. Cuvillier Verlag, Göttingen (2012). PhD Thesis, Universität der Bundeswehr München
36. Ovcharova, N., Gwinner, J.: A study of regularization techniques of nondifferentiable optimization in view of application to hemivariational inequalities. J. Optim. Theory Appl. **162**(3), 754–778 (2014)
37. Poliquin, R.A., Rockafellar, R.T.: Prox-regular functions in variational analysis. Trans. Amer. Math. Soc. **348**(5), 1805–1838 (1996)
38. Qi, L., Shapiro, A., Ling, C.: Differentiability and semismoothness properties of integral functions and their applications. Math. Program., Ser. A **102**(2), 223–248 (2005)
39. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer-Verlag, Berlin (1998)
40. Ruszczyński, A.: Nonlinear Optimization. Princeton University Press, Princeton (2006)
41. Schramm, H., Zowe, J.: A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. SIAM J. Optim. **2**(1), 121–152 (1992)
42. Spingarn, J.E.: Submonotone subdifferentials of Lipschitz functions. Trans. Amer. Math. Soc. **264**(1), 77–89 (1981)
43. Wetzel, M., Holtmannspötter, J., Gudladt, H.J., Czarnecki, J.V.: Sensitivity of double cantilever beam test to surface contamination and surface pretreatment. Int. J. Adhes. Adhes. **46**, 114–121 (2013)
44. Zowe, J.: The BT-algorithm for minimizing a nonsmooth functional subject to linear constraints. In: F.H. Clarke, V.F. Dem'yanov, F. Giannessi (eds.) Nonsmooth Optimization and Related Topics (Erice, 1988), *Ettore Majorana Internat. Sci. Ser. Phys. Sci.*, vol. 43, pp. 459–480. Plenum Press, New York (1989)