Set-Valued Var. Anal (2016) 24:115–148
DOI 10.1007/s11228-015-0352-5

Nonsmooth Bundle Trust-region Algorithm with Applications to Robust Stability

Pierre Apkarian[1] · Dominikus Noll[2] · Laleh Ravanbod[2]

Received: 21 January 2015 / Accepted: 19 October 2015 / Published online: 30 October 2015

Abstract We propose a bundle trust-region algorithm to minimize locally Lipschitz functions which are potentially nonsmooth and nonconvex. We prove global convergence of our method and show by way of an example that the classical convergence argument in trust-region methods based on the Cauchy point fails in the nonsmooth setting. Our method is tested experimentally on three problems in automatic control.

Keywords Bundle · Cutting plane · Trust-region · Cauchy point · Global convergence · Parametric robustness · Distance to instability · Worst-case H_∞-norm

Mathematics Subject Classification (2010) 49J52 · 90C26 · 93D09 · 93B60 · 93B36

1 Introduction

We consider optimization problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array} \tag{1}$$

✉ Dominikus Noll
dominikus.noll@math.univ-toulouse.fr

Pierre Apkarian
Pierre.Apkarian@onera.fr

Laleh Ravanbod
lalehravanbod@yahoo.fr

[1] ONERA, Control System Department, Toulouse, France

[2] Institut de Mathématiques, Université de Toulouse, Toulouse, France

where $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz, but possibly nonsmooth and nonconvex, and where C is a simply structured closed convex constraint set, such as a polyhedron or a semidefinite defined set. We develop a bundle trust-region algorithm for (1), which uses nonconvex cutting planes in tandem with a suitable trust-region management to assure global convergence. The trust-region management is to be considered as an alternative to proximity control, which is the usual policy in bundle methods. Trust-regions allow a tighter control on the step-size. Our experimental part demonstrates how these features may be exploited algorithmically.

Algorithms where bundle and trust-region elements are combined are rather sparse in the literature. For convex objectives Ruszczyński [42] presents a bundle trust-region method, which can be extended to composite convex functions. An early contribution where bundling and trust-regions are combined is [45, 46], and this is also used in versions of the BT-code [51]. Fuduli et al. [21] use DC-functions to form a non-standard trust-region, which they also use in tandem with cutting planes. These nonsmooth trust-region methods approximate the objective by a polyhedral working model, which is updated by adding cutting planes at unsuccessful trial steps. This feature is of course well-known in bundle methods like Sagastizábal and Hare [22, 43] or [38]. Our main Theorem 1 analyses the interaction of this mechanism with the trust-region management, and assures global convergence under realistic hypotheses.

The trust-region strategy is well-understood in smooth optimization, where global convergence is proved by exploiting properties of the Cauchy point, as pioneered in Powell [40]. For the present work it is therefore of the essence to realize that the Cauchy point fails in the nonsmooth setting. This happens even for polyhedral convex functions, the simplest possible case, as we demonstrate by way of a counterexample. This explains why the convergence proof has to be organized along different lines.

The question is then whether there are more restricted classes of nonsmooth functions, where the Cauchy point argument can be salvaged. In response we show that the classical trust-region strategy with Cauchy point is still valid for upper-C^1 functions, and at least partially, for functions having a strict standard model. It turns out that several problems in control and in contact mechanics are in this class, which justifies the disquisition. Nonetheless, the class of functions where the Cauchy point works remains exceptional in the nonsmooth framework, as is corroborated by the fact that convex functions with a genuine nonsmoothness do not have a strict standard model.

A strong incentive for the present work comes indeed from applications in automatic control. In the experimental part we apply our novel bundle trust-region method to compute locally optimal solutions to three NP-hard problems in the theory of systems with uncertain parameters. This includes (i) computing the worst-case H_∞-norm of a system over a given uncertain parameter range, (ii) checking robust stability of an uncertain system over a given parameter range, and (iii) computing the distance to instability of a nominally stable system with uncertain parameters. In these applications the versatility of the bundle trust-region approach with regard to the choice of the norm is exploited.

Nonsmooth trust-region methods which do *not* include the possibility of bundling are more common, see for instance Yuan [48] for composite convex functions, Dennis et al. [18], where the authors present an axiomatic approach, and [14, Chap. 11], where that idea is further expanded. A recent trust-region method for DC-functions is [32]. For information concerning convex and non-convex bundle methods see e.g. [9, Chap. 10], [21, 22, 28, 29, 36–38, 42].

The structure of the paper is as follows. The algorithm is developed in Section 2, and its global convergence is proved in Section 3. Section 4 gives practical stopping criteria. Applications of the model approach are discussed in Section 5, where we also discuss failure of the Cauchy point, using an example from [28]. Applications to automatic control are discussed in Section 6 and numerical experiments are given in Section 7. Conclusions are stated in Section 8.

Notation For nonsmooth optimization we follow [13]. The Clarke directional derivative of f is $f^\circ(x,d)$, its Clarke subdifferential $\partial f(x)$. For a function ϕ of two variables $\partial_1\phi$ denotes the Clarke subdifferential with respect to the first variable. The normal cone to a closed convex set $C \subset \mathbb{R}^n$ at $x \in C$ is $N_C(x) = \{v \in \mathbb{R}^n : v^T(x-x') \geq 0 \text{ for all } x' \in C\}$. We let $P_C(x)$ denote the orthogonal projection of $x \in \mathbb{R}^n$ onto C. Given $x \in \mathbb{R}^n$, the point $y \in C$ is the projection $P_C(x)$ of x onto C if and only if it satisfies the following normal cone criterion

$$(x'-y)^T(x-y) \leq 0 \quad \text{for every } x' \in C,$$

which we use frequently. For symmetric matrices $Q \preceq 0$ means negative semidefinite, $\text{co}(M)$ is the convex hull of a set M. For linear system theory see [50].

2 Presentation of the Algorithm

In this section we derive our trust-region algorithm to solve program (1) and discuss its building blocks.

2.1 Working Model

We start by explaining how a local approximation of f in the neighborhood of the current serious iterate x, called the *working model* of f, is generated iteratively. We recall the notion of a first-order model of f introduced in [38].

Definition 1 A function $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a *first-order model* of f on a set Ω if $\phi(\cdot, x)$ is convex for every $x \in \Omega$, and the following properties are satisfied:

(M_1) $\phi(x,x) = f(x)$, and $\partial_1\phi(x,x) \subset \partial f(x)$.
(M_2) If $y_k \to x$, then there exist $\epsilon_k \to 0^+$ such that $f(y_k) \leq \phi(y_k, x) + \epsilon_k\|y_k - x\|$.
(M_3) If $x_k \to x$, $y_k \to y$, then $\limsup_{k\to\infty} \phi(y_k, x_k) \leq \phi(y, x)$. □

We may consider $\phi(\cdot, x)$ a non-smooth version of the first-order Taylor expansion of f at x. Every locally Lipschitz function has indeed a first-order model $\phi^\sharp$, which we call the *standard model*, defined as

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

Following [38], a first-order model $\phi(\cdot, x)$ is called *strict* at $x \in \Omega$ if the following strict version of (M_2) is satisfied:

($\widehat{M}_2$) Whenever $y_k \to x$, $x_k \to x$, there exist $\epsilon_k \to 0^+$ such that $f(y_k) \leq \phi(y_k, x_k) + \epsilon_k \|y_k - x_k\|$.

Remark 1 Axiom (M_2) corresponds to the one-sided Taylor type estimate $f(y) \leq \phi(y, x) + o(\|y - x\|)$ as $y \to x$. In contrast, axiom ($\widehat{M}_2$) means $f(y) \leq \phi(y, x) + o(\|y - x\|)$ as $\|y - x\| \to 0$ uniformly on bounded sets. This is analogous to the difference between differentiability and strict differentiability, hence the nomenclature of a strict model.

Remark 2 Note that the standard model $\phi^\sharp$ of f is not always strict [36]. A strict first-order model ϕ is for instance obtained for composite functions $f = h \circ F$ with h convex and F of class C^1, if one defines

$$\phi(y, x) = h\left(F(x) + F'(x)(y - x)\right),$$

where $F'(x)$ is the differential of the mapping F at x. The use of a natural model of this form covers for instance approaches like Powell [40], Yuan [48], or Ruszczyński [42], where composite functions are discussed.

Observe that every convex f is its own strict model $\phi(y, x) = f(y)$ in the sense of definition 1. As a consequence, our algorithmic framework contains the convex cutting plane trust-region method [42] as a special case.

Remark 3 It follows from the previous remark that a function f may have several first-order models. Every model ϕ leads to a different algorithm for (1).

During the following we consider x as the current iterate of our algorithm to be designed, and z as a trial point near x, which is a candidate to become the next serious iterate x^+. The way trial points are generated will be explained in Section 2.2.

Definition 2 Let x be the current serious iterate and z a trial step. Let g be a subgradient of $\phi(\cdot, x)$ at z, for short, $g \in \partial_1\phi(z, x)$. Then the affine function $m(\cdot, x) = \phi(z, x) + g^T(\cdot - z)$ is called a *cutting plane* of f at serious iterate x and trial point z. □

We may always represent a cutting plane at serious iterate x in the form

$$m(\cdot, x) = a + g^T(\cdot - x),$$

where $a = m(x, x) = \phi(z, x) + g^T(x - z) \leq f(x)$ and $g \in \partial_1\phi(z, x)$. We say that the pair (a, g) represents the cutting plane $m(\cdot, x)$. In the terminology of [9, Chap. 10], a is called the linearization error of the cutting plane (a, g).

We also allow cutting planes $m_0(\cdot, x)$ at serious iterate x with trial step $z = x$. We refer to these as *exactness planes* of f at serious iterate x, because $m_0(x, x) = f(x)$. Every (a, g) representing an exactness plane is of the form $(f(x), g_0)$ with $g_0 \in \partial f(x)$.

Remark 4 When f is convex, it may be chosen as its own model $\phi(\cdot, x) = f$. In that case the cutting plane according to Definition 2 coincides with the classical convex cutting plane. The plane $m(\cdot, x)$ of Definition 2 might be termed the *model-based cutting plane*, but since there is no risk of confusion, and since this is consistent with the convex case, we continue to call $m(\cdot, x)$ a cutting plane.

Remark 5 For the standard model $\phi^\sharp$ a cutting plane for trial step z at serious iterate x has the very specific form $m^\sharp(\cdot, x) = f(x) + g_z^T(\cdot - x)$, where $g_z \in \partial f(x)$ attains the maximum $f^\circ(x, z - x) = g_z^T(z - x)$. Here every cutting plane $m^\sharp(\cdot, x)$ is also an exactness plane, a fact which will no longer be true for other models ϕ. If f is strictly differentiable at x, then there is only one cutting plane $m^\sharp(\cdot, x) = f(x) + \nabla f(x)^T(\cdot - x)$, the first-order Taylor polynomial.

Definition 3 Let $\mathcal{G}_k$ be a set of pairs (a, g) all representing cutting planes of f at trial steps around the serious iterate x. Suppose $\mathcal{G}_k$ contains at least one exactness plane at x. Then $\phi_k(\cdot, x) = \sup_{(a,g)\in\mathcal{G}_k} a + g^T(\cdot - x)$ is called a *working model* of f at x. $\square$

Remark 6 We index working models ϕ_k by the inner loop counter k to highlight that they are updated in the inner loop by adding tangent planes of the ideal model ϕ at the null steps y^k.

Usually the ϕ_k are rough polyhedral approximation of ϕ, but we do not exclude cases where the ϕ_k are generated by infinite sets $\mathcal{G}_k$. This is for instance the case in the spectral bundle method [23–25], see also [6], which we discuss in Section 5.3.

Remark 7 Note that even the choice $\phi_k = \phi$ is allowed in definition 1 and in Algorithm 1. This corresponds to $\mathcal{G} = \{(a, g) : g \in \partial f(z), a = \phi(z, x) + g^T(x - z)\}$, which is the largest possible set of cuts, or the set of all cuts obtained from ϕ. We discuss this case in Section 5.1. If $\phi^\sharp$ is used, then the corresponding working models are denoted $\phi_k^\sharp$. Their case is analyzed in Section 5.4.

The properties of a working model may be summarized as follows

Proposition 1 *Let $\phi_k(\cdot, x)$ be a working model of f at x built from $\mathcal{G}_k$ and based on the ideal model ϕ. Then*

(i) $\phi_k(\cdot, x) \le \phi(\cdot, x)$.
(ii) $\phi_k(x, x) = \phi(x, x) = f(x)$.
(iii) $\partial_1\phi_k(x, x) \subset \partial_1\phi(x, x) \subset \partial f(x)$.
(iv) *If $(a, g) \in \mathcal{G}_k$ contributes to ϕ_k and stems from the trial step z at serious iterate x, then $\phi_k(z, x) = \phi(z, x)$.*

Proof By construction ϕ_k is a supremum of affine minorants of ϕ, which proves (i). Since at least one plane in $\mathcal{G}_k$ is of the form $m_0(\cdot, x) = \phi(x, x) + g^T(\cdot - x)$ with $g \in \partial_1\phi(x, x)$, we have $\phi_1(x, x) \ge m_0(x, x) = \phi(x, x) = f(x)$, which proves (ii). To prove (iii), observe that since $\phi_k(\cdot, x)$ is convex, every $g \in \partial_1\phi_k(x, x)$ gives an affine minorant $m(\cdot, x) = \phi_k(x, x) + g^T(\cdot - x)$ of $\phi_k(\cdot, x)$. Then $m(\cdot, x) \le \phi(\cdot, x)$ with equality at x. By convexity $g \in \partial_1\phi(x, x)$, and by axiom (M_1) we have $g \in \partial f(x)$. As for (iv), observe that every cutting plane $m(\cdot, x)$ at z satisfies $m(z, x) = \phi(z, x)$, hence also $\phi_k(z, x) = \phi(z, x)$. $\square$

2.2 Tangent Program

In this section we discuss how trial steps z^k are generated. Given the current working model $\phi_k(\cdot, x) = \sup\{a + g^T(\cdot - x) : (a, g) \in \mathcal{G}_k\}$, and the current trust-region radius R_k, the *tangent program* is the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \phi_k(y, x) \\ \text{subject to} & y \in C \\ & \|y - x\| \leq R_k \end{array} \tag{2}$$

where $\|\cdot\|$ could be any norm on $\mathbb{R}^n$. Let y^k be an optimal solution of (2). By the necessary optimality condition there exists a subgradient $g_k^* \in \partial\left(\phi_k(\cdot, x) + i_C\right)(y^k)$ and a vector v_k in the normal cone to $B(x, R_k)$ at $y^k \in B(x, R_k)$ such that $0 = g_k^* + v_k$, where i_C is the indicator function of C [13]. We call g_k^* the *aggregate subgradient* at y^k. The aggregate plane is defined as the affine function $m_k^*(\cdot, x) = a_k^* + g_k^{*T}(\cdot - x)$, where $a_k^* = \phi_k(y^k, x) + g_k^{*T}(x - y^k)$. The aggregate plane satisfies $m_k^*(y^k, x) = \phi_k(y^k, x)$. This terminology stems from the classical bundle method, when a polyhedral working model is used, see Ruszczyński [42], Kiwiel [29].

Remark 8 Consider the case of a polyhedral $\phi_k(\cdot, x) = \max_{i=1,\ldots,k} a_k + g_k^T(\cdot - x)$ with $C = \mathbb{R}^n$ and $\|\cdot\| = |\cdot|$ the Euclidean norm. Here (2) may be written as

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i + g_i^T(y - x) - t \leq 0, i = 1, \ldots, k, \\ & \frac{1}{2}|y - x|^2 - \frac{1}{2}R_k^2 \leq 0, \end{array}$$

with decision variable (t, y). The necessary optimality conditions are

$$\sum_{i=1}^{k} \lambda_i = 1, \quad \sum_{i=1}^{k} \lambda_i g_i + \mu(y - x) = 0, \quad \lambda_i \geq 0, \mu \geq 0$$

along with complementarity and satisfaction of the constraints. We derive the formula $y^k = x - \mu^{-1}\sum_{i=1}^{k} \lambda_i g_i$. In this case the aggregate subgradient g_k^* introduced in more abstract terms above takes the concrete form $g_k^* = \sum_{i=1}^{k} \lambda_i g_i$ of a convex combination of older subgradients arising from cuts, $a_k^* = \sum_{i=1}^{k} \lambda_i a_i$, and $v_k = \mu(y^k - x)$ is the normal to the Euclidean ball $B(x, R_k)$ at y^k. This is analogue to the update formula for the bundle method [9, 29, 38], and explains in which way g_k^* aggregates information from previous cuts. This justifies the use of the term aggregate subgradient in the more general situation of (2).

Solutions y^k of (2) are candidates to become the next serious iterate x^+. For practical reasons we now enlarge the set of possible candidates. Fix $0 < \theta \ll 1$ and $M \geq 1$ once and for all, then every $z^k \in C \cap B(x, M\|x - y^k\|)$ satisfying

$$f(x) - \phi_k(z^k, x) \geq \theta\left(f(x) - \phi_k(y^k, x)\right) \tag{3}$$

is called a *trial step*. Note that y^k itself is of course a trial step, because $f(x) \geq \phi_k(y^k, x)$ by the definition of the tangent program. But due to $\theta \in (0, 1)$, there exists an entire neighborhood U of y^k such that every $z^k \in U \cap C$ is a trial step.

Remark 9 The role of y^k here is not unlike that of the Cauchy point in classical trust-region methods. Suppose we use a standard working model $\phi_k^\sharp$ and f is strictly differentiable at x. Then $\phi_k^\sharp(\cdot, x) = \phi^\sharp(\cdot, x) = f(x) + \nabla f(x)^T(\cdot - x)$. In the unconstrained case $C = \mathbb{R}^n$ the solution y^k has then the explicit form $y^k = x - R_k \frac{\nabla f(x)}{\|\nabla f(x)\|}$, which is indeed the Cauchy point as considered in [44], see also [42, (5.108)]. Condition (3) then takes the familiar form $f(x) - \phi_k^\sharp(z^k, x) \geq \sigma \|\nabla f(x)\| R_k$, see [42, (5.110)].

2.3 Acceptance Test

In order to decide whether a trial step z^k will become the next serious iterate x^+, we compute the test quotient

$$\rho_k = \frac{f(x) - f(z^k)}{f(x) - \phi_k(z^k, x)}, \tag{4}$$

which compares as usual actual progress and model predicted progress. For a fixed parameter $0 < \gamma < 1$, the decision is as follows. If $\rho_k \geq \gamma$, then the trial step z^k is accepted as the new iterate $x^+ = z^k$, and we call this a *serious step*. On the other hand, if $\rho_k < \gamma$, then z^k is rejected and referred to as a *null step*. In that case we compute a cutting plane $m_k(\cdot, x)$ at z^k, and add it to the new set $\mathcal{G}_{k+1}$ in order to improve our working model. In other words, a pair (a_k, g_k) is added, where $g_k \in \partial_1 \phi(z^k, x)$ and $a_k = \phi(z^k, x) + g_k^T(x - z^k)$.

Remark 10 Adding one cutting plane at the null step z^k is mandatory, but we may at leisure add several other tangent planes of $\phi(\cdot, x)$ to further improve the working model ϕ_k, for instance, the one at y^k if $y^k \neq z^k$. A case of practical importance, where the ϕ_k are generated by infinite sets $\mathcal{G}_k$ of cuts, is presented in Section 5.3.

Remark 11 In most applications ϕ_k is a polyhedral convex function. If C is also polyhedral, then it is attractive to choose a polyhedral trust-region norm $\|\cdot\|$, because this makes (2) a linear program. For related ideas in bundle methods see [9, 20, 42].

Remark 12 In the situation of remark 8 it is instructive to compare the bundle and trust-region method when both operate with the Euclidean norm. Namely, if the trust-region constraint is active at the trial step y^k, then the multiplier μ_k of the trust-region constraint plays the role of the proximity control parameter τ_k in e.g. [9, 22, 38]. Increasing τ_k corresponds to decreasing R_k, and decreasing τ_k corresponds to increasing R_k. A new case arises when the solution y^k of the trust region tangent program is in the interior of the trust-region, because here $\mu_k = 0$, and this case has no analogue in the bundle method. However, from a practical point of view the most significant difference between the two method is that trust-regions allow a more direct control of the stepsize.

2.4 Management of the Trust-region Radius

Let x be the current serious iterate and suppose z^k is a trial step that is rejected, the corresponding solution of the tangent program (2) being y^k. Then a cutting plane $m_k(\cdot, x)$ cutting away the unsuccessful trial z^k is added to the working model ϕ_k with the goal to have a better model ϕ_{k+1} at the next sweep. However, it is also necessary to decide whether at the next iteration $k+1$ the trust-region radius should be decreased.

This is where a major difference between the classical trust-region management and the bundle trust-region management occurs. In classical trust-regions the radius R_k is always reduced in the case of a null step. Here we need a different strategy, which has the following rationale. We compute the test quotient

$$\widetilde{\rho}_k = \frac{f(x) - \phi(z^k, x)}{f(x) - \phi_k(z^k, x)}$$

which compares the model predicted progress $f(x) - \phi_k(z^k, x)$ at z^k to the progress $f(x) - \phi(z^k, x) = f(x) - \phi_{k+1}(z^k, x)$ we could have achieved at z^k had we already known the cutting plane $m_k(\cdot, x)$. When $\widetilde{\rho}_k \approx 1$, then adding the cutting plane has little to no effect, and we should reduce the trust-region radius. On the other hand, for $\widetilde{\rho}_k \ll 1$ we can still rely on the effect of adding a cutting plane and keep the trust-region radius invariant. Fixing a constant $\widetilde{\gamma}$ with $0 < \gamma < \widetilde{\gamma} < 1$, this management is put to work as follows.

$$R_{k+1} = \begin{cases} R_k & \text{if } \widetilde{\rho}_k < \widetilde{\gamma},\ \rho_k < \gamma \\ \frac{1}{2} R_k & \text{if } \widetilde{\rho}_k \geq \widetilde{\gamma},\ \rho_k < \gamma \end{cases}. \tag{5}$$

The corresponding rule is applied in step 7 of the algorithm.

Remark 13 Recall that the full-model case $\phi_k = \phi$ is covered by theory. Here the test quotient $\widetilde{\rho}_k$ equals 1, as no bundling is applied. The test (5) becomes redundant (because of $\widetilde{\gamma} < 1$), and the trust-region radius is always reduced in case of a null step. This means that our novel management encompasses the classical situation without bundling as a special case.

2.5 Nonsmooth Solver

We are now ready to present our nonsmooth trust-region algorithm for program (1), given on the next page.

3 Convergence

In this section we analyze the convergence properties of the nonsmooth trust-region algorithm.

3.1 Convergence of the Inner Loop

We start by proving finiteness of the inner loop with counter k. Since the outer loop counter j is fixed, we simplify notation and write $x = x^j$ for the current serious iterate, and $x^+ = x^{j+1}$ for the next serious iterate, which is the result of the inner loop.

Lemma 1 *There exists a constant $\sigma > 0$ depending only on $\theta \in (0, 1)$, $M > 0$, and the norm $\|\cdot\|$, such that for every trial point z^k at inner loop instant k, associated with the solution y^k of the tangent program, and for the corresponding aggregate subgradient g_k^*, we have*

$$f(x) - \phi_k(z^k, x) \geq \sigma \|g_k^*\| \|x - z^k\|. \tag{6}$$

Algorithm 1 Nonsmooth trust-region method

Parameters: $0 < \gamma < \widetilde{\gamma} < 1, 0 < \gamma < \Gamma \leq 1, 0 < \theta \ll 1, M \geq 1$.

▷ **Step 1** (`Initialize outer loop`). Choose initial iterate $x^1 \in C$. Initialize memory trust-region radius as $R_1^\sharp > 0$. Put $j = 1$.

◇ **Step 2** (`Stopping test`). At outer loop counter j, stop if x^j is a critical point of (1). Otherwise, goto inner loop.

▷ **Step 3** (`Initialize inner loop`). Put inner loop counter $k = 1$ and initialize trust-region radius as $R_1 = R_j^\sharp$. Build initial working model $\phi_1(\cdot, x^j)$ based on $\mathcal{G}_1$, where at least $(f(x^j), g_{0j}) \in \mathcal{G}_1$ for some $g_{0j} \in \partial f(x^j)$. Possibly enrich $\mathcal{G}_1$ by recycling some of the planes from the previous serious step.

▷ **Step 4** (`Trial step generation`). At inner loop counter k find solution y^k of the tangent program

$$\begin{array}{ll} \text{minimize} & \phi_k(y, x^j) \\ \text{subject to} & y \in C \\ & \|y - x^j\| \leq R_k \end{array}$$

Then compute any trial step $z^k \in C \cap B(x^j, M\|x^j - y^k\|)$ satisfying $f(x^j) - \phi_k(z^k, x^j) \geq \theta\left(f(x^j) - \phi_k(y^k, x^j)\right)$.

◇ **Step 5** (`Acceptance test`). If

$$\rho_k = \frac{f(x^j) - f(z^k)}{f(x^j) - \phi_k(z^k, x^j)} \geqslant \gamma,$$

put $x^{j+1} = z^k$ (serious step), quit inner loop and goto step 8. Otherwise (null step), continue inner loop with step 6.

▷ **Step 6** (`Update working model`). Generate a cutting plane $m_k(\cdot, x^j) = a_k + g_k^T(\cdot - x^j)$ of f at the null step z^k at counter k belonging to the current serious step x^j. Add (a_k, g_k) to $\mathcal{G}_{k+1}$. Possibly taper out $\mathcal{G}_{k+1}$ by removing some of the older inactive planes in $\mathcal{G}_k$. Build ϕ_{k+1} based on $\mathcal{G}_{k+1}$.

◇ **Step 7** (`Update trust-region radius`). Compute secondary control parameter

$$\widetilde{\rho}_k = \frac{f(x^j) - \phi(z^k, x^j)}{f(x^j) - \phi_k(z^k, x^j)}$$

and put

$$R_{k+1} = \begin{cases} R_k & \text{if } \widetilde{\rho}_k < \widetilde{\gamma}, \\ \frac{1}{2} R_k & \text{if } \widetilde{\rho}_k \geqslant \widetilde{\gamma}. \end{cases}$$

Increase inner loop counter k and loop back to step 4.

◇ **Step 8** (`Update memory radius`). Store new memory radius

$$R_{j+1}^\sharp = \begin{cases} R_k & \text{if } \rho_k < \Gamma, \\ 2R_k & \text{if } \rho_k \geqslant \Gamma. \end{cases}$$

Increase outer loop counter j and loop back to step 2.

Proof 1) Let $\|\cdot\|$ be the norm used in the trust-region tangent program, $|\cdot|$ the standard Euclidean norm. There exists $\epsilon > 0$ such that $|u| \leq \epsilon$ implies $\|u\| \leq 1$. Now if $\|u\| = 1$ and if v is in the normal cone to the $\|\cdot\|$-unit ball at u, we have $v^T(u - u') \geq 0$ for every $\|u'\| \leq 1$ by the normal cone criterion. Hence $v^T(u - u') \geq 0$ for every $|u'| \leq \epsilon$ by the above, and using $u' = \epsilon v/|v|$ that implies $v^T u \geq \epsilon |v|$.

2) Since y^k is an optimal solution of (2), we have $0 = g_k^* + v_k$, where $g_k^* \in \partial\left(\phi_k(\cdot, x) + i_C\right)(y^k)$ and v_k a normal vector to the $\|\cdot\|$-norm ball $B(x, R_k)$ at y^k. By the

subgradient inequality,

$$g_k^{*T}(x-y^k) \le \phi_k(x,x) - \phi_k(y^k,x) = f(x) - \phi_k(y^k,x).$$

Now by part 1), on putting $u_k = (y^k - x)/\|y^k - x\|$, we have $v_k^T u_k \ge \epsilon|v_k|$ independently of k, because v_k, being normal to the $\|\cdot\|$-ball of radius $\|y^k - x\|$ and center 0 at $y^k - x$, is also normal to the $\|\cdot\|$-unit ball at u_k. But then $g_k^{*T}(x-y^k) = v_k^T(y^k - x) \ge \epsilon|v_k|\|y^k - x\| \ge \epsilon^2\|v_k\|\|y^k - x\| = \epsilon^2\|g_k^*\|\|y^k - x\|$. Invoking (3) for the trial point z^k, and using $\|x - z^k\| \le M\|x - y^k\|$, we get (6) with $\sigma = \epsilon^2\theta M^{-1}$. □

Lemma 2 *Suppose the inner loop at x with trial point z^k at inner loop counter k and solution y^k of the tangent program (2) turns infinitely, and the trust-region radius R_k stays bounded away from 0. Then x is a critical point of (1).*

Proof We have $\rho_k < \gamma$ for all k. Since $\liminf_{k\to\infty} R_k > 0$, and since according to (5) the trust-region radius is reduced when $\widetilde{\rho}_k \ge \widetilde{\gamma}$, and is never increased during the inner loop, we conclude that there exists k_0 such that $\widetilde{\rho}_k < \widetilde{\gamma}$ for all $k \ge k_0$, and $R_k = R_{k_0} > 0$ for all $k \ge k_0$.

As $z^k, y^k \in B(x, R_{k_0})$, we can extract an infinite subsequence $k \in \mathcal{K}$ such that $z^k \to z$, $y^k \to y$, $k \in \mathcal{K}$. Now consider $k \in \mathcal{K}$ and its predecessor $k' \in \mathcal{K}$, $k' < k$. Since the cutting plane drawn at $z^{k'}$ contributes to ϕ_k, we have $\phi_k(z^{k'}, x) = \phi(z^{k'}, x) \to \phi(z, x)$. Since the working models are minorants of the ideal model ϕ, they have a common Lipschitz constant $L > 0$ on the compact set $B(x, R_{k_0})$, i.e., $|\phi_k(z^k, x) - \phi_k(z^{k'}, x)| \le L\|z^k - z^{k'}\|$ for all $k', k \in \mathcal{K}$. Since $z^k - z^{k'} \to 0$ and $\phi_k(z^{k'}, x) \to \phi(z, x)$ by what was observed above, we deduce $\phi_k(z^k, x) \to \phi(z, x)$. Therefore the numerator and denominator in the quotient $\widetilde{\rho}_k$ both converge to $\phi(x,x) - \phi(z,x)$, $k \in \mathcal{K}$. Since $\widetilde{\rho}_k < \widetilde{\gamma} < 1$ for all k, this could only mean $\phi(x,x) - \phi(z,x) = 0$.

Now by condition (3) we have

$$\phi(x,x) - \phi_k(y^k,x) \le \theta^{-1}\left(\phi(x,x) - \phi_k(z^k,x)\right) \to 0,$$

hence $\limsup_{k\in\mathcal{K}} \phi(x,x) - \phi_k(y^k,x) \le 0$. On the other hand, $\phi_k(y^k,x) \le \phi(x,x)$ since y^k solves the tangent program, hence $\phi_k(y^k,x) \to \phi(x,x)$, too.

By the necessary optimality condition for the tangent program (2) there exist $p_k \in \partial_1\phi_k(y^k,x)$ and a normal vector q_k to $C \cap B(x, R_{k_0})$ at y^k such that $0 = p_k + q_k$. By boundedness of the y^k and local boundedness of the subdifferential, see e.g. [13, Prop. 2.1.2] or [41], the sequence p_k is bounded, and hence so is the sequence q_k. Passing to yet another subsequence $k \in \mathcal{K}' \subset \mathcal{K}$, we may assume $p_k \to p$, $q_k \to q$, and by upper semi-continuity of the subdifferential, $p \in \partial_1\phi(y,x)$, while q is in the normal cone to $C \cap B(x, R_{k_0})$ at y. Since $0 = p + q$, we deduce that y is a critical point of the optimization program $\min\{\phi(y,x) : y \in C \cap B(x, R_{k_0})\}$, and since this is a convex program, y is a minimum. But from the previous argument we have seen that $\phi(y,x) = \phi(x,x)$, and since x is admissible for that program, it is also a minimum. A simple convexity argument now shows that x is a minimum of (2), and by axiom (M_1) x is then a critical point of (1). □

Remark 14 Note that the argument in Lemma 2 remains valid if we take the cutting plane at y^k instead of z^k. That gives our method additional flexibility.

Lemma 3 *Suppose the inner loop at x with trial point z^k and solution y^k of the tangent program at inner loop counter k turns forever, and* $\liminf_{k\to\infty} R_k = 0$. *Then x is a critical point of (1).*

Proof This proof uses (6) obtained in Lemma 1. We are in the case where $\widetilde{\rho}_k \geq \widetilde{\gamma}$ for infinitely many $k \in \mathcal{N}$. Since R_k is never increased in the inner loop, we have $R_k \to 0$ by rule (5). Hence $y^k, z^k \to x$ as $k \to \infty$.

We claim that $\phi_k(z^k, x) \to f(x)$. Indeed, we clearly have $\limsup_{k\to\infty} \phi_k(z^k, x) \leq \limsup_{k\to\infty} \phi(z^k, x) = \lim_{k\to\infty} \phi(z^k, x) = f(x)$. On the other hand, the exactness plane $m_0(\cdot, x) = f(x) + g_0^T(\cdot - x)$ is an affine minorant of $\phi_k(\cdot, x)$ at all times k, hence $f(x) = \lim_{k\to\infty} m_0(z^k, x) \leq \liminf_{k\to\infty} \phi_k(z^k, x)$, and the two together show $\phi_k(z^k, x) \to f(x)$.

By condition (6) we have $f(x) - \phi_k(z^k, x) \geq \sigma\|g_k^*\|\|x - z^k\|$, where $g_k^* \in \partial\left(\phi_k(\cdot, x) + i_C\right)(y^k)$ is the aggregate subgradient, and where σ is independent of k. Now assume that $\|g_k^*\| \geq \eta > 0$ for all k. Then $f(x) - \phi_k(z^k, x) \geq \sigma\eta\|x - z^k\|$.

Since $z^k \to x$, using axiom (M_2) there exist $\epsilon_k \to 0^+$ such that $f(z^k) - \phi(z^k, x) \leq \epsilon_k\|x - z^k\|$. But then

$$\widetilde{\rho}_k = \rho_k + \frac{f(z^k) - \phi(z^k, x)}{f(x) - \phi_k(z^k, x)} \leq \rho_k + \frac{\epsilon_k\|x - z^k\|}{\sigma\eta\|x - z^k\|} = \rho_k + \epsilon_k/(\sigma\eta).$$

Since $\epsilon_k \to 0$, $\rho_k < \gamma$, we have $\limsup_{k\to\infty} \widetilde{\rho}_k \leq \gamma < \widetilde{\gamma}$, contradicting the fact that $\widetilde{\rho}_k > \widetilde{\gamma}$ for infinitely many k. Hence $\|g_k^*\| \geq \eta > 0$ was impossible.

Select $k \in \mathcal{K}$ such that $g_k^* \to 0$. Write $g_k^* = p_k + q_k$ with $p_k \in \partial_1\phi_k(y^k, x)$ and $q_k \in N_C(y^k)$. Using boundedness of the y^k, and hence boundedness of the p_k, we extract another subsequence $k \in \mathcal{K}'$ such that $p_k \to p$, $q_k \to q$. Since $y^k \to x$, we have $q \in N_C(x)$. We argue that $p \in \partial f(x)$. Indeed, for any test vector h the subgradient inequality gives

$$p_k^T h \leq \phi_k(y^k + h, x) - \phi_k(y^k, x) \leq \phi(y^k + h, x) - \phi_k(y^k, x).$$

Since $\phi_k(y^k, x) \to f(x) = \phi(x, x)$, passing to the limit gives

$$p^T h \leq \phi(x + h, x) - \phi(x, x),$$

proving $p \in \partial_1\phi(x, x) \subset \partial f(x)$ by axiom (M_1). Since $p + q = 0$, this proves that x is a critical point of (1). $\square$

Remark 15 For polyhedral ϕ_k one can limit the size of the sets $\mathcal{G}_k$ to $|\mathcal{G}_k| \leq n + 2$. Namely, if (a_k, g_k) represents the cutting plane at null step z^k and (a_k^*, g_k^*) the aggregate plane at the corresponding solution y^k of the tangent program, then by Carathéodory's theorem we can find $\mathcal{G}_{k+1}$ of size at most $n + 2$ such that the convex hull of $\mathcal{G}_{k+1}$ coincides with the convex hull of $\mathcal{G}_k \cup \{(a_k, g_k), (a_k^*, g_k^*)\}$. As Lemma 4 below shows, finiteness of the inner loop can then still be guaranteed.

This estimate $n + 2$ is pessimistic, an efficient heuristic method is to remove from $\mathcal{G}_k$ inactive cuts as well as a certain number of active cuts and represent those by the aggregate plane, which is added to $\mathcal{G}_{k+1}$.

Remark 16 In the bundle method with proximity control, Kiwiel's aggregate subgradient [29] allows a rigorous theoretical limit $|\mathcal{G}_k| \leq 3$, even though in practice one keeps more cuts in $\mathcal{G}_k$. It is not known whether Kiwiel's argument can be extended to the trust-region case, see also [42, Ch. 7.5] for a discussion.

Lemma 4 *Suppose $|\mathcal{G}_k| \leq n+2$, and let z^k be a null step with associated solution y^k of the tangent program. Let (a_k, g_k) represent the cutting plane at z^k and (a_k^*, g_k^*) the aggregate plane at y^k. Then we can build a set of cuts $\mathcal{G}_{k+1}$ such that $\mathrm{co}(\mathcal{G}_{k+1}) = \mathrm{co}\left(\mathcal{G}_k \cup \{(a_k, g_k), (a_k^*, g_k^*)\}\right)$, $|\mathcal{G}_{k+1}| \leq n+2$, and such that the conclusions of Lemmas 2 and 3 remain valid for the working model based on $\mathcal{G}_{k+1}$.*

Proof From Carathéodory's theorem we get $\mathcal{G}_{k+1}$ of size at most $n+2$ such that the convex hull of $\mathcal{G}_{k+1}$ coincides with that of $\mathcal{G}_k \cup \{(a_k^*, g_k^*), (a_k, g_k)\}$. Since the planes in $\mathcal{G}_k$ are affine minorants of ϕ, the same remains true in $\mathcal{G}_{k+1}$, because $(a_k, g_k), (a_k^*, g_k^*)$ are also affine minorant of $\phi(\cdot, x)$. Now build ϕ_{k+1} from $\mathcal{G}_{k+1}$, then what is needed in the proofs of Lemmas 2, 3 is that $\phi_{k+1}(y^k, x) \geq \phi_k(y^k, x)$ and $\phi_{k+1}(z^k, x) = \phi(z^k, x)$, which we now check.

Since the aggregate plane belongs to the set $\mathcal{G}_k \cup \{(a_k, g_k), (a_k^*, g_k^*)\}$, there exists a convex combination $(a_k^*, g_k^*) = \sum_{i=1}^{n+2} \lambda_i (a_i, g_i)$ with $(a_i, g_i) \in \mathcal{G}_{k+1}$. Then $\phi_k(y^k, x) = m_k^*(y^k, x) = a_k^* + g_k^{*T}(y^k - x) = \sum_{i=1}^{n+2} \lambda_i \left(a_i + g_i^T(y^k - x)\right) \leq \sum_{i=1}^{n+2} \lambda_i \phi_{k+1}(y^k, x) = \phi_{k+1}(y^k, x)$ proving the first inequality. A similar argument showing $\phi_{k+1}(z^k, x) = \phi(z^k, x)$ applies to the cutting plane. □

3.2 Convergence of the Outer Loop

In this section we prove convergence of the outer loop. This is where axiom $(\widehat{M}_2)$ will be required.

Theorem 1 *Suppose that f has a strict first-order model ϕ. Let $x^1 \in C$ be such that $\{x \in C : f(x) \leq f(x^1)\}$ is bounded. Let $x^j \in C$ be the sequence of iterates generated by Algorithm 1 based on ϕ. Then every accumulation point x^* of the x^j is a critical point of (1).*

Proof 1) Without loss we consider the case where the algorithm generates an infinite sequence $x^j \in C$ of serious iterates. Suppose that at outer loop counter j the inner loop finds a successful trial step at inner loop counter k_j, that is, $z^{k_j} = x^{j+1}$, where the corresponding solution of the tangent program is $\tilde{x}^{j+1} = y^{k_j}$. Then $\rho_{k_j} \geq \gamma$, which means

$$f(x^j) - f(x^{j+1}) \geq \gamma \left(f(x^j) - \phi_{k_j}(x^{j+1}, x^j)\right). \tag{7}$$

Moreover, by condition (3) we have $\|\tilde{x}^{j+1} - x^j\| \leq M\|x^{j+1} - x^j\|$ and

$$f(x^j) - \phi_{k_j}(x^{j+1}, x^j) \geq \theta \left(f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j)\right), \tag{8}$$

and combining (7) and (8) gives

$$f(x^j) - f(x^{j+1}) \geq \gamma\theta \left(f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j)\right). \tag{9}$$

Since $y^{k_j} = \tilde{x}^{j+1}$ is a solution of the k_jth tangent program (2) of the jth inner loop, there exist $g_j^* \in \partial \left(\phi_{k_j}(\cdot, x^j) + i_C\right)(\tilde{x}^{j+1})$ and a unit normal vector v_j to the ball $B(x^j, R_{k_j})$ at $\tilde{x}^{j+1}$ such that

$$g_j^* + \|g_j^*\| v_j = 0. \tag{10}$$

Consider an accumulation point x^* of the sequence of serious iterates x^j, and a subsequence $j \in J$ such that $x^j \to x^*$. We have to show that x^* is critical. We shall now analyze two types of infinite subsequences $j \in J$, those where the trust-region constraint is active at $\tilde{x}^{j+1}$ and the Lagrange multiplier of the trust-region constraint is nonzero, i.e. $g_j^* \neq 0$ in (10), and those where the Lagrange multiplier of the trust-region constraint vanishes, i.e., $g_j^* = 0$ in (10).

2) Let us start with the simpler case of an infinite subsequence x^j, $j \in J$, where the Lagrange multiplier of the trust-region constraint vanishes, i.e., $g_j^* = 0$ in (10). That occurs either when $\|x^j - \tilde{x}^{j+1}\| < R_{k_j}$, i.e., where the trust-region constraint is inactive, or when it is active but with vanishing multiplier. Now there exist $p_j \in \partial_1 \phi_{k_j}(\tilde{x}^{j+1}, x^j)$ and $q_j \in N_C(\tilde{x}^{j+1})$ such that

$$0 = g_j^* = p_j + q_j.$$

By the subgradient inequality, applied to $p_j \in \partial \phi_{k_j}(\cdot, x^j)(\tilde{x}^{j+1})$, we have

$$\begin{aligned}
-q_j^T(x^j - \tilde{x}^{j+1}) = p_j^T(x^j - \tilde{x}^{j+1}) &\leq \phi_{k_j}(x^j, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&= f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&\leq \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right),
\end{aligned}$$

using (9). Since $p_j^T(x^j - \tilde{x}^{j+1}) = q_j^T(\tilde{x}^{j+1} - x^j) \geq 0$ by the normal cone criterion, we deduce summability $\sum_{j \in J} p_j^T(x^j - \tilde{x}^{j+1}) < \infty$ from telescoping of the last term above, hence $p_j^T(x^j - \tilde{x}^{j+1}) \to 0$, $j \in J$, and then also $q_j^T(x^j - \tilde{x}^{j+1}) \to 0$. Passing to a subsequence, we may assume $p_j \to p$, $q_j \to q$, and $\tilde{x}^{j+1} \to \tilde{x}$.

Let h be any test vector, then from the subgradient inequality,

$$\begin{aligned}
p_j^T h &\leq \phi_{k_j}(\tilde{x}^{j+1} + h, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&\leq \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) + f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\
&\leq \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).
\end{aligned}$$

Now let h' be another test vector and put $h = x^j - \tilde{x}^{j+1} + h'$. On substituting this expression we obtain

$$p_j^T(x^j - \tilde{x}^{j+1}) + p_j^T h' \leq \phi(x^j + h', x^j) - f(x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Passing to the limit in suitable convergent subsequences, we have $p_j^T(x^j - \tilde{x}^{j+1}) \to 0$ by the above, and $f(x^j) - f(x^{j+1}) \to 0$ by the construction of the descent method. Moreover, $\limsup_{j \in J} \phi(x^j + h', x^j) \leq \phi(x^* + h', x^*)$ by $x^j \to x^*$, axiom (M_3), and $p_j \to p$. That shows

$$p^T h' \leq \phi(x^* + h', x^*) - f(x^*) = \phi(x^* + h', x^*) - \phi(x^*, x^*).$$

Since h' was arbitrary and $\phi(\cdot, x^*)$ is convex, we deduce $p \in \partial_1\phi(x^*, x^*)$, hence $p \in \partial f(x^*)$ by axiom (M_1).

Now we have to show that $q \in N_C(x^*)$. Since $q_j^T(x^j - \tilde{x}^{j+1}) \to 0$, we have $q^T(x^* - \tilde{x}) = 0$. Now for any element $x \in C$ we have $q^T(\tilde{x} - x) \geq 0$ by the normal cone criterion. Hence $q^T(x^* - x) = q^T(\tilde{x} - x) + q^T(x^* - \tilde{x}) = q^T(\tilde{x} - x) \geq 0$, so the normal cone criterion holds also at x^*, proving $q \in N_C(x^*)$. We have shown that $0 = p + q \in \partial\left(\phi(\cdot, x^*) + i_C\right)(x^*)$, hence x^* is a critical point of (1).

3) Let us now consider the more complicated case of an infinite subsequence, where $\|x^j - \tilde{x}^{j+1}\| = R_{k_j}$ with $g_j^* \neq 0$, corresponding to the case of a non-vanishing multilplier in (10). Recall that $x^j \to x^*$, $j \in J$, and that we have to show that x^* is critical.

As a consequence of Lemma 1 we have

$$f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \geq \sigma\|g_j^*\|\|x^j - \tilde{x}^{j+1}\| \tag{11}$$

for a constant $\sigma > 0$ depending only on the norm $\|\cdot\|$, and therefore independent of j. Combining this with (9) gives

$$\|g_j^*\|\|x^j - \tilde{x}^{j+1}\| \leq \sigma^{-1}\gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Summing both sides from $j = 1$ to $j = J$ gives

$$\sum_{j=1}^{J}\|g_j^*\|\|x^j - \tilde{x}^{j+1}\| \leq \sigma^{-1}\gamma^{-1}\theta^{-1}\left(f(x^1) - f(x^{J+1})\right).$$

Since the values $f(x^j)$ are decreasing and $\{x \in C : f(x) \leq f(x^1)\}$ is bounded, the sequence x^j must be bounded. We deduce that the right hand side is bounded, hence the series on the left converges:

$$\sum_{j=1}^{\infty}\|g_j^*\|\|x^j - \tilde{x}^{j+1}\| < \infty. \tag{12}$$

In particular, this implies $\|g_j^*\|\|x^j - \tilde{x}^{j+1}\| \to 0$. Using $\|x^j - x^{j+1}\| \leq M\|x^j - \tilde{x}^{j+1}\|$, we also have $\|g_j^*\|\|x^j - x^{j+1}\| \to 0$.

We shall now have to distinguish two subcases. Either $R_{k_j} \geq R_0 > 0$ for some $R_0 > 0$ and all $j \in J$, or there exists a subsequence $J' \subset J$ such that $R_{k_j} \to 0$ as $j \in J'$. The first case is discussed in 4), the second case will be handled in 5) - 6).

4) Let us consider the sub-case of an infinite subsequence $j \in J$ where $\|x^j - \tilde{x}^{j+1}\| = R_{k_j} \geq R_0 > 0$ for every $j \in J$. Going back to (12), we see that we now must have $g_j^* \to 0$, as $x^j - \tilde{x}^{j+1} \not\to 0$. Let us write $g_j^* = p_j + q_j$, where $p_j \in \partial_1\phi_{k_j}(\tilde{x}^{j+1}, x^j)$ and $q_j \in N_C(\tilde{x}^{j+1})$. Then by the subgradient inequality and (9) we have

$$p_j^T(x^j - \tilde{x}^{j+1}) \leq \phi_{k_j}(x^j, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \leq \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Now $g_j^{*T}(x^j - \tilde{x}^{j+1}) = p_j^T(x^j - \tilde{x}^{j+1}) + q_j^T(x^j - \tilde{x}^{j+1}) \leq p_j^T(x^j - \tilde{x}^{j+1})$, because the normal cone criterion for $\tilde{x}^{j+1} \in C$ and $q_j \in N_C(\tilde{x}^{j+1})$ gives $q_j^T(\tilde{x}^{j+1} - x^j) \geq 0$. Hence we have

$$g_j^{*T}(x^j - \tilde{x}^{j+1}) \leq p_j^T(x^j - \tilde{x}^{j+1}) \leq \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right),$$

so $p_j^T(x^j - \tilde{x}^{j+1}) \to 0$, because the lefthand term and the righthand term both converge to 0. As a consequence, we also have $q_j^T(x^j - \tilde{x}^{j+1}) \to 0$.

Now observe that the sequence $x^j \in C$ is bounded, because $\{x \in C : f(x) \leq f(x^1)\}$ is bounded and the x^j form a descent sequence for f. Let us say $\|x^1 - x^j\| \leq K$ for all j. We argue that the sequence p_j is then also bounded. This can be shown as follows. Let h be a test vector with $\|h\| = 1$. Then

$$\begin{aligned} p_j^T h &\leq \phi_{k_j}(\tilde{x}^{j+1} + h, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\ &\leq \phi(\tilde{x}^{j+1} + h, x^j) - m_{0j}(\tilde{x}^{j+1}, x^j) \\ &= \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) - g_{0j}^T(\tilde{x}^{j+1} - x^j) \\ &\leq C_1 + C_2 + \|g_{0j}\| \|x^j - \tilde{x}^{j+1}\|, \end{aligned}$$

where $C_1 := \max\{\phi(u, v) : \|u - x^1\| \leq MK + 1, \|v - x^1\| \leq K\} < \infty$ and $C_2 = \max\{|f(x^j)| : j \in \mathbb{N}\}$, and where $g_{0j} \in \partial f(x^j)$ by the definition of the exactness plane at x^j. But observe that ∂f is locally bounded by [13, Prop. 2.1.2], [41], so $\|g_{0j}\| \leq K' < \infty$. We deduce $\|p_j\| \leq C_1 + C_2 + K'(2K + M) < \infty$. Hence the sequence p_j is bounded, and since $g_j^* = p_j + q_j \to 0$ by the above, the sequence q_j is also bounded.

Therefore, on passing to a subsequence $j \in J'$, we may along with the standing $x^j \to x^*$ also assume that $\tilde{x}^{j+1} \to \tilde{x}$, $p_j \to p$, $q_j \to q$. Then $q \in N_C(\tilde{x})$. Now from the subgradient inequality

$$\begin{aligned} p_j^T h &\leq \phi_{k_j}(\tilde{x}^{j+1} + h, x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\ &\leq \phi(\tilde{x}^{j+1} + h, x^j) - f(x^j) + f(x^j) - \phi_{k_j}(\tilde{x}^{j+1}, x^j) \\ &\leq \phi(\tilde{x}^{j+1} + h, x^j) - \phi(x^j, x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right), \end{aligned}$$

where we use (9), $\phi_{k_j} \leq \phi$, and acceptance $\rho_{k_j} \geq \gamma$, and where the test vector h is arbitrary. Let h' another test vector and put $h = x^j - \tilde{x}^{j+1} + h'$. Substituting this gives

$$p_j^T(x^j - \tilde{x}^{j+1}) + p_j^T h' \leq \phi(x^j + h', x^j) - \phi(x^j, x^j) + \gamma^{-1}\theta^{-1}\left(f(x^j) - f(x^{j+1})\right). \quad (13)$$

Now $p_j^T(x^j - \tilde{x}^{j+1}) = (p_j + q_j)^T(x^j - \tilde{x}^{j+1}) + q_j^T(\tilde{x}^{j+1} - x^j) \geq (p_j + q_j)^T(x^j - \tilde{x}^{j+1})$ using the normal cone criterion for $q_j \in N_C(\tilde{x}^{j+1})$. Therefore, on passing to the limit in (13), using $(p_j + q_j)^T(x^j - \tilde{x}^{j+1}) \to 0$, $f(x^j) - f(x^{j+1}) \to 0$, $p_j \to p$ and $\limsup_{j \in J'} \phi(x^j + h', x^j) \leq \phi(x^* + h', x^*)$, which follows from axiom (M_3), we find

$$p^T h' \leq \phi(x^* + h', x^*) - \phi(x^*, x^*).$$

Since h' was arbitrary and $\phi(\cdot, x^*)$ is convex, we deduce $p \in \partial_1\phi(x^*, x^*)$, and by axiom (M_1), $p \in \partial f(x^*)$.

It remains to show $q \in N_C(x^*)$. Now recall that $q_j^T(x^j - \tilde{x}^{j+1}) \to 0$ was shown at the beginning of part 4), so $q^T(x^* - \tilde{x}) = 0$. Given any test element $x \in C$, the normal cone criterion for $q \in N_C(\tilde{x})$ gives $q^T(\tilde{x} - x) \geq 0$. But then $q^T(x^* - x) = q^T(\tilde{x} - x) + q^T(x^* - \tilde{x}) = q^T(\tilde{x} - x) \geq 0$, so the normal cone criterion also holds for q at x^*, proving $q \in N_C(x^*)$.

With $q \in N_C(x^*)$ and $p + q = 0$, we have shown that x^* is a critical point of (1). That settles the case where the trust-region radius is active and bounded away from 0.

5) It remains to discuss the most complicated sub-case of an infinite subsequence $j \in J$, where the trust-region constraint is active with non-vanishing multiplier, and $R_{k_j} \to 0$. This needs two sub-sub-cases. The first of these is a sequence $j \in J$ where in each jth outer loop the trust-region radius was reduced at least once. The second sub-sub-case are infinite subsequences where the trust-region radius stayed frozen ($R_j^\sharp = R_{k_j}$) throughout the jth inner loop for every $j \in J$. This is discussed in 6) below.

Let us first consider the case of an infinite sequence $j \in J$ where R_{k_j} is active at $\tilde{x}^{j+1}$, and $R_{k_j} \to 0$, $j \in J$, and during the jth inner loop the trust-region radius was reduced at least once. Suppose this happened the last time before acceptance at inner loop counter $k_j - \nu_j$ for some $\nu_j \geq 1$. Then for $j \in J$,

$$R_{k_j} = R_{k_j-1} = \cdots = R_{k_j-\nu_j+1} = \tfrac{1}{2} R_{k_j-\nu_j}.$$

By step 7 of the algorithm, that implies

$$\widetilde{\rho}_{k_j-\nu_j} \geq \widetilde{\gamma}, \quad \rho_{k_j-\nu_j} < \gamma.$$

Now $\|x^{j+1} - x^j\| \leq R_{k_j}$ and $\|z^{k_j-\nu_j} - x^j\| \leq M R_{k_j-\nu_j-1} = 2MR_{k_j}$, hence $x^{j+1} - z^{k_j-\nu_j} \to 0$, $x^j - z^{k_j-\nu_j} \to 0$, $j \in J''$. From axiom $(\widehat{M}_2)$ we deduce that there exists a sequence $\epsilon_j \to 0^+$ such that

$$f(z^{k_j-\nu_j}) \leq \phi(z^{k_j-\nu_j}, x^j) + \epsilon_j \|z^{k_j-\nu_j} - x^j\|.$$

By the definition of the aggregate subgradient $\widetilde{g}_j \in \partial \left(\phi_{k_j-\nu_j}(\cdot, x^j) + i_C\right)(y^{k_j-\nu_j})$ at $y^{k_j-\nu_j}$ and by Lemma 1 we have $f(x^j) - \phi_{k_j-\nu_j}(z^{k_j-\nu_j}, x^j) \geq \sigma \|\widetilde{g}_j\| \|x^j - z^{k_j-\nu_j}\|$ for a constant σ independent of j. Now recall that $x^j \to x^*$ and that we have to show that x^* is critical. It suffices to show that there is a subsequence $j \in J'$ with $\widetilde{g}_j \to 0$. This argument uses the fact that $z^{k_j-\nu_j} - x^j \to 0$.

Assume on the contrary that $\|\widetilde{g}_j\| \geq \eta > 0$ for every $j \in J$. Then

$$f(x^j) - \phi_{k_j-\nu_j}(z^{k_j-\nu_j}, x^j) \geq \eta\sigma \|z^{k_j-\nu_j} - x^j\|.$$

Now

$$\widetilde{\rho}_{k_j-\nu_j} = \rho_{k_j-\nu_j} + \frac{f(z^{k_j-\nu_j}) - \phi(z^{k_j-\nu_j}, x^j)}{f(x^j) - \phi_{k_j-\nu_j}(z^{k_j-\nu_j}, x^j)} \leq \rho_{k_j-\nu_j} + \frac{\epsilon_j \|z^{k_j-\nu_j} - x^j\|}{\eta\sigma \|z^{k_j-\nu_j} - x^j\|} < \widetilde{\gamma}$$

for $j \in J$ sufficiently large, contradicting $\widetilde{\rho}_{k_j-\nu_j} \geq \widetilde{\gamma}$. This shows that there must exist a subsequence J' such that $\widetilde{g}_j \to 0$, $j \in J'$. Passing to the limit $j \in J'$, this shows $0 \in \partial\left(\phi(\cdot, x^*) + i_C\right)(x^*)$, hence x^* is critical for (1).

6) Now consider an infinite subsequence $j \in J$ where $x^j \to x^*$, the trust-region radius R_{k_j} was active at $\tilde{x}^{j+1}$ with non-zero multiplier when x^{j+1} was accepted, $R_{k_j} \to 0$, but during the jth inner loop the trust-region radius was never reduced. In the classical case this can only happen when x^{j+1} at j is immediately accepted, but with bundling this could also happen when the inner loop adds cutting planes for a time, while the test in step 7 keeps $R_{k+1} = R_k$ in the inner loop. Since $R_{k_j} \to 0$, the work to bring the radius to 0 must be put about somewhere else. For every $j \in J$ define $j' \in \mathbb{N}$ to be the largest index $j' < j$ such that in the j'th inner loop, the trust-region radius was reduced at least once. Let $J' = \{j' : j \in J\}$, where we understand $j \mapsto j'$ as a function. Passing to a subsequence of J, J', we may assume that $x^{j'} \to x'$ and $g^*_{j'} \to 0$, because the sequence J' corresponds to one of the cases discussed in parts 2) - 5). Passing to yet another subsequence, we may arrange that the

sequences J, J' are interlaced. That is, $j' < j < j'^{+} < j^{+} < j'^{++} < j^{++} < \cdots \to \infty$. This is because j' tends to ∞ as a function of j.

Now assume that there exists $\eta > 0$ such that $\|g_j^*\| \geq \eta$ for all $j \in J$. Then since $x^j \to x^*$, we also have $x^{j+1} \to x^*$ due to (12). Fix $\epsilon > 0$ with $\epsilon < \eta$. For $j \in J$ large enough we have $\|g_{j'}^*\| < \epsilon$, because $g_{j'}^* \to 0$, $j' \in J'$, and as j gets larger, so does j'. That means in the interval $[j', j)$ there exists an index $j'' \in \mathbb{N}$ such that

$$\|g_{j''}^*\| < \epsilon, \quad \|g_i^*\| \geq \epsilon \text{ for all } i = j''+1, \ldots, j.$$

The index j'' may coincide with j', it might also be larger, but it precedes j. In any case, $j \mapsto j''$ is again a function on J and defines another infinite index set J'' still interlaced with J.

Now recall from part 3), estimate (12), and $\|x^j - x^{j+1}\| \leq M\|x^j - \tilde{x}^{j+1}\|$, that for some constant $c > 0$

$$\sum_{i=j''+1}^{j} \|g_i^*\|\|x^i - x^{i+1}\| \leq c\left(f(x^{j''+1}) - f(x^{j+1})\right) \to 0 \qquad (j \in J, j \to \infty, j \mapsto j'').$$

Since by construction $\|g_i^*\| \geq \epsilon$ for all $i \in [j''+1, \ldots, j]$, and that for all $j \in J$, the sequence $\sum_{i=j''+1}^{j} \|x^i - x^{i+1}\| \to 0$ converges as $j \in J, j \to \infty$, and by the triangle inequality, $x^{j''+1} - x^{j+1} \to 0$. Therefore $x^{j''+1} \to x^*$. Since $g_{j''}^* \in \partial(f + i_C)(x^{j''+1})$, passing to yet another subsequence and using upper semi-continuity of the subdifferential, we get $g_{j''}^* \to g^* \in \partial(f + i_C)(x^*)$. Since $\|g_{j''}^*\| < \epsilon$, we have $\|g^*\| \leq \epsilon$. It follows that $\partial(f + i_C)(x^*)$ contains an element g^* of norm less than or equal ϵ. As $\epsilon < \eta$ was arbitrary, we conclude that $0 \in \partial(f + i_C)(x^*)$. That settles the remaining case. □

4 Stopping Test

A closer look at the convergence proof indicates stopping criteria for Algorithm 1. As is standard in bundle methods, step 2 is not executed as such but delegated to the inner loop. When a serious step x^{j+1} is accepted, we apply the tests

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

in tandem with

$$\frac{\min\{\|P_C(-g_j^*)\|, \|P_C(-g_{j'}^*)\|, \|P_C(-\widetilde{g}_j)\|\}}{1 + |f(x^j)|} < \text{tol}_3.$$

Here g_j^* is the aggregate subgradient at acceptance k_j. In the case treated in part 6) of the proof we had to consider the largest index $j' < j$, where the trust-region radius was reduced for the last time, and $g_{j'}^*$ was the aggregate subgradient at that index $j' < j$. This explains the second projected gradient.

The third projected aggregate concerns the case discussed in part 5) of the proof. This is a subsequence J such that for every $j \in J$ the trust-region radius was reduced at least once and $R_j \to 0$. Here we have to take the last aggregate $\widetilde{g}_j \in \partial_1\left(\phi_{k_j - \nu_j}(\cdot, x^j) + i_C\right)(y^{k_j - \nu_j})$

before reduction into account, hence the third term. If the three criteria are satisfied, then we return x^{j+1} as our candidate for the optimal solution.

On the other hand, when the inner loop has difficulties finding a new serious iterate, and if a maximum number $k_{\max}$ is exceeded, or if for $\nu_{\max}$ consecutive steps

$$\frac{\|x^j - z^k\|}{1 + \|x^j\|} < \mathrm{tol}_1, \qquad \frac{f(x^j) - f(z^k)}{1 + |f(x^j)|} < \mathrm{tol}_2$$

in tandem with

$$\frac{\|P_C(-g_k^*)\|}{1 + |f(x^j)|} < \mathrm{tol}_3$$

are satisfied, where g_k^* is the aggregate subgradient at y^k, then the inner loop is stopped and x^j is returned as optimal. In our tests we use $k_{\max} = 50$, $\nu_{\max} = 5$, $\mathrm{tol}_1 = \mathrm{tol}_2 = 10^{-5}$, $\mathrm{tol}_3 = 10^{-6}$. Typical values in Algorithm 1 are $\gamma = 0.0001$, $\widetilde{\gamma} = 0.0002$, $\Gamma = 0.1$.

5 Applications

In this section we highlight the potential of the model-based trust-region approach by presenting several applications.

5.1 Full Model Versus Working Model

Our convergence theory covers the specific case $\phi_k = \phi$, which we call the *full model case*. Here the algorithm simplifies, because cutting planes are redundant, so that step 6 becomes obsolete. Moreover, in step 7 the quotient $\widetilde{\rho}_k$ always equals 1, so the only action taken is reduction of the trust-region radius. This is now close to the rationale of the classical trust-region method.

5.2 Natural Model

For a composite function $f = g \circ F$ with g convex and F of class C^1 the *natural model* is $\phi(y, x) = g\left(F(x) + F'(x)(y - x)\right)$, because ϕ is strict and can be used in Algorithm 1. In the full model case $\phi_k = \phi$, our algorithm reduces to the algorithm of Ruszczyński [42, Chap. 7.5] for composite nonsmooth functions.

5.3 Spectral Model

An important field of applications, where the natural model often comes into action, is eigenvalue optimization

$$\begin{array}{ll} \text{minimize} & \lambda_1\left(\mathcal{F}(x)\right) \\ \text{subject to} & x \in C \end{array} \tag{14}$$

where $\mathcal{F} : \mathbb{R}^n \to \mathbb{S}^m$ is a class C^1-mapping into the space of $m \times m$ symmetric or Hermitian matrices $\mathbb{S}^m$, and $\lambda_1(\cdot)$ the maximum eigenvalue function on $\mathbb{S}^m$, which is convex but nonsmooth. Here the natural model is $\phi(y, x) = \lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right)$, where $\mathcal{F}'$ is the differential of $\mathcal{F}$. Every nonlinear semidefinite program

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \mathcal{F}(x) \preceq 0 \\ & x \in C \end{array} \tag{15}$$

can be cast as a special cases of (14) if exact penalization is used. We write (15) in the form

$$\begin{array}{ll} \text{minimize} & f(x) + c \max\{0, \lambda_1(\mathcal{F}(x))\} \\ \text{subject to} & x \in C \end{array}$$

with a suitable $c > 0$. Namely, this new objective may be written as the maximum eigenvalue of the mapping

$$\mathcal{F}^{\sharp}(x) = \begin{bmatrix} f(x) & 0 \\ 0 & f(x)I_m + c\mathcal{F}(x) \end{bmatrix} \in \mathbb{S}^{1+m}.$$

Let us apply the bundling idea to (14) using the natural model ϕ. Here we may build working models ϕ_k generated by infinite sets $\mathcal{G}_k$ of cuts (a, g) from ϕ, and still arrive at a computable tangent program. Indeed, suppose for simplicity that $y^k = z^k$ is a null step at serious iterate x. According to step 6 of Algorithm 1 we have to generate one or several cutting planes at y^k. This means we have to compute $g_k \in \partial\lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(\cdot - x)\right)(y^k)$. Now by the generalized chain rule the subdifferential of the composite function $y \mapsto \lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right)$ at y is $\mathcal{F}'(x)^*\partial\lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right)$, where $\partial\lambda_1$ is now the convex subdifferential of λ_1 in matrix space $\mathbb{S}^m$, i.e.,

$$\partial\lambda_1(X) = \{G \in \mathbb{S}^m : G \succeq 0, \operatorname{tr}(G) = 1, G \bullet X = \lambda_1(X)\}$$

with $X \bullet Y = \operatorname{tr}(XY)$ the scalar product in $\mathbb{S}^m$. Here $\mathcal{F}'(x)^* : \mathbb{S}^m \to \mathbb{R}^n$ is the adjoint of the linear operator $\mathcal{F}'(x)$. It follows that every subgradient g of the composite function is of the form

$$g = \mathcal{F}'(x)^* G, \quad G \in \partial\lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right). \tag{16}$$

The corresponding a is $a = \lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y - x)\right) + g^T(x - y)$. As soon as the maximum eigenvalue $\lambda_1(X)$ has multiplicity strictly larger than one, the set $\partial\lambda_1(X)$ is not singleton. This is where we may include infinitely many subgradients into the new set $\mathcal{G}_{k+1}$, as we indicate below.

Let y^k be a null step, and let Q_r be an $m \times t_k$ matrix whose t_k columns form an orthogonal basis of the maximum eigenspace of $\mathcal{F}(x) + \mathcal{F}'(x)(y^k - x)$. Let Y_k be a $t_k \times t_k$-matrix with $Y_k = Y_k^T$, $Y_k \succeq 0$, $\operatorname{tr}(Y_k) = 1$, then subgradients (16) are of the form $G_k = Q_k Y_k Q_k^T$. Therefore all pairs $(a_r, g(Y_r)) \in \mathcal{G}_k$ are of the form

$$a_r = \lambda_1\left(\mathcal{F}(x) + \mathcal{F}'(x)(y^r - x)\right) + g(Y_r)^T(x - y^r),$$
$$g(Y_r) = \mathcal{F}'(x)^* G_r, \quad G_r = Q_r Y_r Q_r^T,$$

indexed by $Y_r \succeq 0$, $\operatorname{tr}(Y_r) = 1$, $Y_r \in \mathbb{S}^{t_r}$ stemming from older null steps $r = 1, \ldots, k$. The trust-region tangent program is then

$$\begin{array}{ll} \text{minimize} & \max\limits_{r=1,\ldots,k} a_r + \lambda_1\left(Q_r\mathcal{F}'(x)(y - x)Q_r^T\right) \\ \text{subject to} & y \in C, \ \|y - x\| \leq R_k. \end{array} \tag{17}$$

This is a linear semidefinite program if a polyhedral or a conical norm is used, and if C is a convex semidefinite constraint set. For large scale problems Helmberg and Rendl [24] and

Helmberg and Oustry [25] show how the tangent program (17) can be limited to a practical size. See Helmberg and Kiwiel [23] for additional information on spectral bundle methods.

We can go one step further and consider semi-infinite maximum eigenvalue problems as in [6], as this has scope for applications in automatic control. It allows us for instance to optimize the H_∞-norm, or more general IQC-constrained programs, see [5].

5.4 Standard Model

The most straightforward choice of a model is the *standard model*

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

as it gives a direct substitute for the first-order Taylor expansion of f at x. Here the full model tangent program (2) has the specific form

$$\begin{aligned} \text{minimize} \ \ & f(x) + f^\circ(x, y-x) \\ \text{subject to} \ \ & y \in C \\ & \|y-x\| \le R_k \end{aligned} \tag{18}$$

and if a polyhedral working model $\phi_k^\sharp$ is used to approximate $\phi^\sharp$ via bundling, then we get an even simpler tangent program of the form

$$\begin{aligned} \text{minimize} \ \ & f(x) + \max_{i=1,\dots,k} g_i^T(y-x) \\ \text{subject to} \ \ & y \in C \\ & \|y-x\| \le R_k \end{aligned} \tag{19}$$

where $g_i \in \partial f(x)$. If a polyhedral norm is used and C is a polyhedron, then (19) is just a linear program, which makes this computationally attractive.

Remark 17 Consider the unconstrained case $C = \mathbb{R}^n$ with $\phi_k^\sharp = \phi^\sharp$, then $y^k = x - R_k g(x)/\|g(x)\|$, where $g(x) = \operatorname*{argmin}_{g\in\partial f(x)} \{\|g\| : g \in \partial f(x)\}$, and this is the nonsmooth steepest descent step of length R_k at x. In classical trust-region algorithms the steepest descent step of length R_k is often chosen as the Cauchy step.

This raises the following question. Can we use the solution of y^k of (18), or (19), as a nonsmooth Cauchy point? In general the answer is in the negative, because according to Theorem 1 the use of the standard model $\phi^\sharp$ in Algorithm 1 is only authorized when $\phi^\sharp$ is strict. A sufficient condition for strictness of $\phi^\sharp$ is given in [37]. To discuss it, we need the following definition.

Definition 4 (Spingarn [47], Rockafellar-Wets [41]) A locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ is lower-C^1 at $x_0 \in \mathbb{R}^n$ if there exist a compact space $\mathbb{K}$, a neighborhood U of x_0, and a mapping $F : \mathbb{R}^n \times \mathbb{K} \to \mathbb{R}$ such that

$$f(x) = \max_{y\in\mathbb{K}} F(x,y) \tag{20}$$

for all $x \in U$, and F and $\partial F/\partial x$ are jointly continuous. The function f is said to be upper-C^1 at x_0 if $-f$ is lower-C^1 at x_0. □

Lemma 5 *(See [37]). Suppose f is locally Lipschitz and upper-C^1. Then the standard model $\phi^\sharp$ of f is strict.* □

Example 1 The lightning function $f : \mathbb{R} \to \mathbb{R}$ in [30] is an example where $\phi^\sharp$ is strict, but f is not upper-C^1. It is Lipschitz with constant 1 and has $\partial f(x) = [-1, 1]$ for every x. The standard model of f is strict, because for all x, y there exists $\rho = \rho(x, y) \in [-1, 1]$ such that

$$\begin{aligned} f(y) = f(x) + \rho|y - x| &\leq f(x) + \text{sign}(y - x)(y - x) \\ &\leq f(x) + f^\circ(x, y - x) = \phi^\sharp(x, y - x), \end{aligned}$$

using the fact that $\text{sign}(y - x) \in \partial f(x)$. At the same time f is certainly not upper-C^1, because it is not semi-smooth in the sense of [34].

When using the standard model $\phi^\sharp$ in Algorithm 1, we expect the trust-region method to coincide with its classical antecedent, or at least, to be very similar to it. But we expect more. Let $\mathscr{S}$ be the class of nonsmooth locally Lipschitz functions f which have a strict standard model $\phi^\sharp$. Suppose a subclass $\mathscr{S}'$ of $\mathscr{S}$ leads to simplifications of Algorithm 1 which reduce it to its classical counterpart. Then we have a theoretical justification to say that functions $f \in \mathscr{S}'$, even though nonsmooth, can be optimized *as if they were smooth*.

As we shall see in proposition 2 below, such simplifications occur for functions which are densely strictly differentiable. Criteria for dense strict differentiability are known in the literature. Following Borwein and Moors [10], a function f is called *essentially smooth* if it is locally Lipschitz and strictly differentiable almost everywhere. Nonsmooth functions arising in practice are essentially smooth as a rule, cf. [10]. Sufficient conditions to guarantee this are for instance semi-smooth functions in the sense of [34], arc-wise essentially smooth functions, or pseudo-regular functions in the sense of [10].

Nonetheless, there exist locally Lipschitz functions which are nowhere strictly differentiable. The lightning function of example 1 is a pathological case, which is differentiable almost everywhere, but nowhere strictly differentiable.

Proposition 2 *Let f be essentially smooth and suppose C has nonempty interior. Let $x^1 \in C$ be such that $\{x \in C : f(x) \leq f(x^1)\}$ is bounded. Suppose the standard model $\phi^\sharp$ is used in Algorithm 1. Then trial points $z^k \in C$ satisfying (3) in step 4 may be chosen as points of strict differentiability of f. This makes the steps of the algorithm identical with the steps of the classical first-order trust-region algorithm. In addition, if $\phi^\sharp$ is strict, then every accumulation point of the sequence x^j is critical.*

Proof Since there exists a full neighborhood U of y^k such that every $z^k \in U \cap C$ is a valid trial point, and since the points of strict differentiability of f are dense in $U \cap C$, we can assure that z^k is chosen as a point of strict differentiability. That guarantees that the entire sequence x^j consists of points of strict differentiability. In consequence, the standard model at x^j is $\phi^\sharp(\cdot, x^j) = f(x^j) + \nabla f(x^j)^T(\cdot - x^j)$. That means cutting planes are redundant, as is the secondary test in step 7 of the algorithm. The procedure then reduces to the classical first-order trust-region method. Naturally, convergence is only guaranteed when $\phi^\sharp$ is strict. □

Note that we should not expect the y^k themselves to be points of differentiability, let alone strict differentiability. In fact the y^k will typically lie in a set of measure 0. For instance, if C is a polyhedron, then y^k is typically a vertex of C, or a vertex of the polyhedron of the linear program (19).

Proposition 2 applies in particular when f is upper-C^1, because upper-C^1-functions are essentially smooth. However, for upper-C^1 functions we have the following stronger result. A similar observation in the context of bundle methods was first made in [16].

Lemma 6 *Suppose f is locally Lipschitz and upper-C^1 and the standard model $\phi^\sharp$ is used in Algorithm 1. Then we can choose the cutting plane $m_k(\cdot, x) = f(x) + g_k^T(\cdot - x)$ in step 6 with $g_k \in \partial f(x)$ arbitrarily, because $f^\circ(x, z^k - x) - g_k^T(z^k - x) \leq \epsilon_k \|z^k - x\|$ holds automatically for certain $\epsilon_k \to 0^+$ in the inner loop at x, and $f^\circ(x^j, x^{j+1} - x^j) - g_j^T(x^{j+1} - x^j) \leq \epsilon_j \|x^{j+1} - x^j\|$ holds automatically for certain $\epsilon_j \to 0^+$ in the outer loop.*

Proof Daniilidis and Georgiev [15, Thm. 2] prove that an upper-C^1 function is super-monotone at x in the following sense. For every $\epsilon > 0$ there exists $\delta > 0$ such that $(g_1 - g_2)^T(x_1 - x_2) \leq \epsilon \|x_1 - x_2\|$ for all $x_i \in U$ and $g_i \in \partial f(x_i)$. Hence for sequences $x^j, y^j \to x$ we find $\epsilon_j \to 0^+$ such that $(g_j^* - g_j)^T(x^j - y^k) \leq \epsilon_j \|y^j - x^j\|$ for all $g_j^* \in \partial f(y^j)$, $g_j \in \partial f(x^j)$. Choosing g_j^* such that $f^\circ(x^j, y^j - x^j) = g_j^{*T}(y^j - x^j)$ then gives the result. □

For the following result recall from [8] that a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies a Kurdyka-Łojasiewicz inequality at $x_0 \in \mathbb{R}^n$ if there exist $\eta > 0$, a neighborhood U of x_0, and a concave function $\kappa : [0, \eta] \to [0, \infty)$ which is of class C^1 on $(0, \eta)$ such that the following conditions are satisfied.

(i) $\kappa(0) = 0$ and $\kappa' > 0$ on $(0, \eta)$.

(ii) For every $x \in U$ with $f(x_0) < f(x) < f(x_0) + \eta$ we have

$$\kappa'(f(x) - f(x_0)) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

This inequality is satisfied as soon as a function f is defined in a natural way, see [8] for details.

Theorem 2 *Suppose f is upper-C^1, $x^1 \in C$, and $\{x \in C : f(x) \leq f(x^1)\}$ is bounded. Suppose the classical trust-region algorithm is used in the following sense. The only plane in step 6 chosen at x^j is an arbitrarily fixed exactness plane, and in step 7 the trust-region radius is reduced whenever a null step occurs. Then every accumulation point of the sequence of serious iterates x^j is a critical point of (1). Moreover, if f satisfies a Kurdyka-Łojasiewicz inequality, then the x^j converge to a single critical point x^* of f.*

Proof By Lemma 6 the proof of Theorem 1 applies regardless how we choose cutting planes from $\phi^\sharp$. In particular, the present choice of taking an arbitrary exactness plane and keeping it all the time, is covered by Lemma 6. This makes step 6 redundant and reduces step 7 to the usual modification of the trust-region radius. And this is now just the classical trust-region strategy, for which we then have subsequence convergence by Theorem 1.

It remains to show that under the Kurdyka-Łojasiewicz inequality the x^j converge even to a single limit. This can be based on the technique of [1, 7, 37]. □

Remark 18 An axiomatic approach to trust-region methods is Dennis et al. [18], and the idea is adopted in [14, Chap. 11]. The difference with our approach is that ϕ in [14, 18] has to be jointly continuous, while we use the weaker axiom (M_3), and that their f has to be regular, which precludes the use of the standard model $\phi^\sharp$, hence makes it impossible to use the Cauchy point. Bundling is not discussed in these approaches.

On the other hand, the authors of [18], [14] allow non-convex models, while in our approach $\phi(\cdot, x)$ is convex because we want to assure a computable tangent program, and be able to draw cutting planes. Convexity of $\phi(\cdot, x)$ could be relaxed to $\phi(\cdot, x)$ being lower-C^1. For that the downshift idea [34, 36] would have to be used.

5.5 Failure of the Cauchy Point

We will show by way of an example that the classical trust-region approach based on the Cauchy point fails in the nonsmooth case. We operate Algorithm 1 with the full standard model $\phi^\sharp = \phi_k^\sharp$, compute the Cauchy point y^k via (18) based on the Euclidian norm, and use $z^k = y^k$ as the trial step. This corresponds essentially to a classical first-order trust-region method.

The following example adapted from [28] can be used to show the difficulties with this classical scheme. We define a convex piecewise affine function $f : \mathbb{R}^2 \to \mathbb{R}$ as

$$f(x) = \max\{f_0(x), f_{\pm 1}(x), f_{\pm 2}(x)\} \tag{21}$$

where $x = (x_1, x_2)$ and

$$f_0(x) = -100,\ f_{\pm 1}(x) = \pm 2x_1 + 3x_2,\ f_{\pm 2}(x) = \pm 5x_1 + 2x_2.$$

The plot in Fig. 1 shows that part of the level curve $\{x : f(x) = a\}$ for $a > 0$, which lies in the upper half plane $x_2 \geq 0$. It consists of the polygon connecting the five points $(-\frac{a}{5}, 0), (-\frac{a}{11}, \frac{3a}{11}), (0, \frac{a}{3}), (\frac{a}{11}, \frac{3a}{11}), (\frac{a}{5}, 0)$. We are interested in that part of the lower level set $\{x : f(x) \leq a\}$, which lies within the gray-shaded diamond-shaped area inside the polygon $\{x : f(x) = a\}$, and above the x_1-axis.

Consider the exceptional set $N = \cup_{i \neq j}\{x : f_i(x) = f_j(x) = f(x)\}$, whose intersection with the upper half-plane $x_2 \geq 0$ consists of the three lines $x_1 = 0$, $x_2 = \pm 3x_1$. Then for $x \notin N$ the gradient $\nabla f(x)$ is unique. We will generate a sequence x^j of iterates which never meets N, so that $\phi^\sharp(y, x) = f(x) + \nabla f(x)^T(y - x)$ with $\nabla f(x) \in \{\pm(2, 3), \pm(5, 2)\}$ at all iterates x^j. It will turn out that serious iterates x^j never leave the diamond area, only trial points may.

Assume that our current iterate x has $f(x) = a$ and is situated on the right upper part of the a-diamond, shown as the blue x in the figure. That means

$$x = (x_1, -\tfrac{2}{3}x_1 + \tfrac{a}{3}), \quad f(x) = a, \quad 0 < x_1 \leq \tfrac{a}{11}.$$

Then $\phi^\sharp(y, x) = f_{1+}(y) = 2y_1 + 3y_2$. If the current trust-region radius is $R = \sqrt{13}r$, then the solution of (2) is $y = x + r(-2, -3) = (x_1 - 2r, -\frac{2}{3}x_1 + \frac{a}{3} - 3r)$. If we follow the point y as a function of r along the steepest descent line shown in blue, we will reach the points A, B in increasing order at $0 < r_A < r_B$. Here A is the intersection of the steepest descent line with the x_2 axis, reached at $r_A = x_1/2$. The point B is when the ray meets the boundary of the a-diamond, which is the line $x_2 = -3x_1$ on the left, reached at

$$r_B = \tfrac{7}{27}x_1 + \tfrac{a}{27}.$$

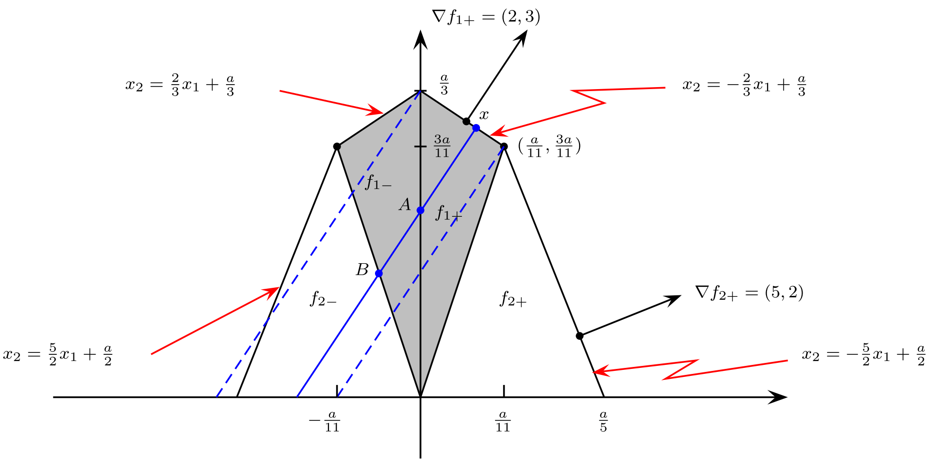

Fig. 1 Curve of level $a > 0$ of (21). Cauchy-step based trust-region iterates do not leave the diamond-shaped area and get stalled at the origin

We have $f(A) = f_{1+}(A) = a - \frac{17}{4}x_1$ and $f(B) = f_{1-}(B) = -\frac{143}{27}x_1 + \frac{22}{2}a$, and from here on f increases along the ray. The test quotient ρ for trial points y of this form behaves as follows

$$\rho = \frac{f(x_a) - f(y)}{f(x_a) - \phi^\sharp(y, x_a)} = \begin{cases} 1 & \text{if } 0 < r \le r_A \\ \frac{4x_1+5r}{13r} & \text{if } r_A \le r \le r_B \\ \frac{a-12r+19x_1}{39r} & \text{if } r_B \le r < \infty \end{cases}$$

The quotient is therefore constant $= 1$ on $[0, r_A]$, and decreasing on $[r_A, \infty)$. If we trace the quotient at the point B as a function of x_1, we see that $\rho = \frac{5}{13}$ at $x_1 = 0$, and $\rho = \frac{198}{234}$ at $x_1 = \frac{a}{11}$. That means if we take the Armijo constant as $\gamma \in (\frac{198}{234}, 1)$, then none of the points in $[B, \infty)$ is accepted, whatever $x_1 \in (0, \frac{a}{11}]$. Let the value r where the quotient ρ equals γ be called r_γ. Then $r_A < r_\gamma < r_B$, and we have $r_\gamma = \frac{4x_1}{13\gamma-5}$.

Let us for simplicity put $\Gamma = 1$. That means good steps where the trust-region radius is doubled are exactly those in $(x, A]$, that is, $0 < r \le r_A$. Such a step is immediately accepted, and we stay on the right upper half of the a^+-diamond, where $a^+ < a$, except for the point A, which we will exclude later. We find for $0 < r < r_A = x_1/2$:

$$a^+ = a - 13r > 0, \quad x^+ = (x_1 - 2r, -\tfrac{2}{3}x_1 + \tfrac{a}{3} - 3r) = (x_1^+, -\tfrac{2}{3}x_1^+ + \tfrac{a^+}{3}).$$

Note that $a = a^+$ for the limiting case $x_1 = 0$, and $a^+ = \frac{9}{22}a$ for the limiting case $x_1 = \frac{a}{11}$. According to step 8 of the algorithm the trust-region radius is doubled ($R^+ = 2R$) for $0 < r < r_A$, because $\rho = 1 \ge \Gamma = 1$.

The second case is when from the current x with $f(x) = a$ a step with $R = \sqrt{13}r$ and $r \in (r_A, r_\gamma)$ is taken. Then we end up on the left hand side of the diamond with the new situation

$$x^+ = (x_1 - 2r, -\tfrac{2}{3}x_1 + \tfrac{a}{3} - 3r), \quad f(x^+) = f_{1-}(x^+) = -4x_1 + a - 5r = a^+.$$

By symmetry, this case is analogous to the initial situation, the model at x^+ now being f_{1-}. We are now on the upper left side of the smaller a^+-diamond. Since $\gamma \le \rho < \Gamma$, the trust-region radius remains unchanged.

The third case is when $r \in [r_\gamma, \infty)$. Here the step is rejected, and the trust-region radius is halved, until a value $r < r_\gamma$ is reached.

Since $\phi^\sharp$ is used and f is strictly differentiable at serious iterates, no cutting planes are taken, and we follow the classical trust-region method. In consequence, the serious iterates $x, x^+, x^{++}, \dots$ stay in the diamonds $a, a^+, a^{++}, \dots$ and converge to the origin, which is not a critical point of f. Note that we have to assure that none of the trial points y lies precisely on the x_2-axis. Now it is clear that for a given starting point x the method has a countable number of possible trial steps y^k, and we can choose the initial $x_1 \in (0, \frac{a}{11}]$ such that the x_2-axis is avoided, for instance, by taking an irrational initial value. Alternatively, in the case where y^k hits the x_2-axis, we might use rule (3) to change it slightly to a z^k, which is not on the axis. In both cases the method will never leave the diamond area, hence convergence based on the Cauchy point fails.

6 Parametric Robustness

We consider a plant P of the form

$$P(s): \begin{cases} \dot{x} = Ax + B_p p + B_w w \\ q = C_q x + D_{qp} p + D_{qw} w \ , \\ z = C_z x + D_{zp} p + D_{zw} w \end{cases} \tag{22}$$

where $x \in \mathbb{R}^{n_x}$ is the state, $w \in \mathbb{R}^{m_1}$ the vector of exogenous inputs, and $z \in \mathbb{R}^{p_1}$ the regulated output. As shown schematically in Fig. 2 we put P in an upper feedback loop $\mathcal{F}_u(P, \Delta)$ with the uncertain block Δ via

$$p = \Delta q, \tag{23}$$

where the uncertain matrix Δ has the block-diagonal form

$$\Delta = \text{diag}\left[\delta_1 I_{r_1}, \dots, \delta_m I_{r_m}\right], \tag{24}$$

with $\delta_1, \dots, \delta_m$ representing real uncertain parameters, and r_i giving the number of repetitions of these δ_i. We write $\delta = (\delta_1, \dots, \delta_m)$ and assume without loss that $\delta = 0$ represents the nominal parameter value. Moreover, we consider $\delta \in \mathbb{R}^m$ in one-to-one correspondence with the matrix Δ in (24). Note that every system featuring real-rational uncertain parameters can be represented via such a Linear Fractional Transform $\mathcal{F}_u(P, \Delta)$, see [50].

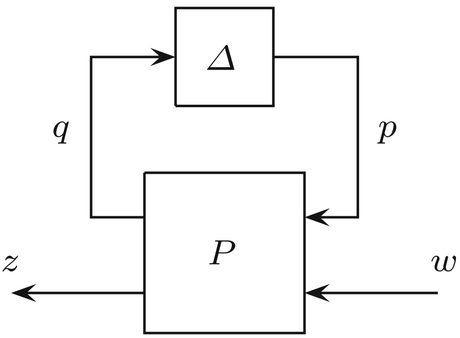

Fig. 2 Robust system interconnection $\mathcal{F}_u(P, \Delta)$, obtained by closing the loop between (22) and (23), where Δ has the structure (24)

6.1 Worst case H_∞-performance over a Parameter set

Our first problem concerns analysis of the performance of the system (22)-(24) in the presence of parametric uncertainty. In order to analyze the robustness of (22)-(24) we compute the worst-case H_∞ performance of the channel $w \to z$ over a given uncertain parameter range normalized to $\Delta = [-1, 1]^m$. In other words, we compute

$$h^* = \max\{\|T_{wz}(\delta)\|_\infty : \delta \in \Delta\}, \tag{25}$$

where $T_{wz}(\delta)$ is the transfer function $z(s) = \mathcal{F}_u(P(s), \Delta)w(s)$, or more explicitly,

$$z(s) = \left[P_{22}(s) + P_{21}(s)\Delta(I - P_{11}(s)\Delta)^{-1}P_{12}(s)\right]w(s).$$

The significance of (25) is that computing a critical parameter value $\delta^* \in \Delta$ which degrades the H_∞-performance of (22)-(24) may be an important jigsaw piece in assessing the properties of a controlled system. We refer to [2], where this is exploited in parametric robust controller synthesis.

Solving (25) leads to a program of the form (1) if we write (25) as minimization of $h_-(\delta) = -\|T_{wz}(\delta)\|_\infty$ over the convex Δ. The specific form of Δ strongly suggests the use of the maximum norm $|\delta|_\infty = \max\{|\delta_1|, \ldots, |\delta_m|\}$ to define trust-regions. Moreover, we will use the standard model $\phi^\sharp$ of $h_-(\delta) = -\|T_{wz}(\delta)\|_\infty$, as is justified by the following

Lemma 7 *Let* $D = \{\delta : T_{zw}(\delta)$ is internally stable$\}$. *Then* $h_- : \delta \mapsto -\|T_{zw}(\delta)\|_\infty$ *is upper-*C^1 *on* D.

Proof It suffices to prove that $h_+ : \delta \mapsto \|T_{wz}(\delta)\|_\infty$ is lower-C^1. To prove this, recall that the maximum singular value has the variational representation

$$\overline{\sigma}(G) = \sup_{\|u\|=1} \sup_{\|v\|=1} \left|u^T G v\right|.$$

Now observe that $z \mapsto |z|$, being convex, is lower-C^1 as a mapping $\mathbb{R}^2 \to \mathbb{R}$, so we may write it as

$$|z| = \sup_{l \in \mathbb{L}} \Psi(z, l)$$

for Ψ jointly of class C^1 and a suitable compact set $\mathbb{L}$. An explicit construction of $\Psi, \mathbb{L}$ could be obtained from Spingarn [47, Thm. 3.9]. Then

$$h_+(\delta) = \sup_{j\omega \in \mathbb{S}^1} \sup_{\|u\|=1} \sup_{\|v\|=1} \sup_{l \in \mathbb{L}} \Psi\left(u^T T_{zw}(\delta, j\omega)v, l\right), \tag{26}$$

where $\mathbb{S}^1 = \{j\omega : \omega \in \mathbb{R} \cup \{\infty\}\}$ is homeomorphic with the 1-sphere. This is a representation of the form (20) for h_+, where the compact space is $\mathbb{K} := \mathbb{S}^1 \times \{u : \|u\| = 1\} \times \{v : \|v\| = 1\} \times \mathbb{L}$, F is $F(\delta, j\omega, u, v, l) := \Psi\left(u^T T_{zw}(\delta, j\omega)v, l\right)$ and $y = (j\omega, u, v, l)$. $\square$

The proof also shows that the non-smoothness in h_+, h_- is due to the maximum singular value and to the semi-infiniteness in the supremum over $\mathbb{S}^1$ in (26).

Theorem 3 (Worst-case H_∞ norm on Δ) *Let* $\delta^j \in \Delta$ *be the sequence generated by the standard trust-region algorithm applied to program (25) based on the standard model of* h_-. *Then the* δ^j *converge to a critical point* δ^* *of (25).*

Proof By Lemma 6 Algorithm 1 coincides with a classical first-order trust-region algorithm, with convergence in the sense of subsequences. Convergence to a single critical point then follows by observing that h_- satisfies a Łojasiewicz inequality. □

6.2 Robust Stability over a Parameter set

In our second problem we wish to check whether the uncertain system (22)-(24) is robustly stable over the uncertain parameter set $\Delta = [-1, 1]^m$. This can be tested by maximizing the spectral abscissa over Δ:

$$\alpha^* = \max\{\alpha\left(A(\delta)\right) : \delta \in \Delta\}, \tag{27}$$

where $A(\delta)$ is the closed-loop system matrix

$$A(\delta) = A + B_p\Delta\left(I - D_{qp}\Delta\right)^{-1} C_q, \tag{28}$$

and where the spectral abscissa of $A \in \mathbb{R}^{n\times n}$ is $\alpha(A) = \max\{\text{Re}(\lambda) : \lambda \text{ eigenvalue of A}\}$. As soon as $\alpha^* \geq 0$, the solution δ^* of (27) represents a destabilizing choice of the parameters, and this may be valuable information in practice, see e.g. [2]. On the other hand, if the global maximum has value $\alpha^* < 0$, then a certificate for robust stability over $\delta \in \Delta$ is obtained.

Global maximization of (27) is NP-hard [11, 39], so it is interesting to use a local optimization method to compute good lower bounds. This can be achieved by Algorithm 1, because (27) is clearly of the form (1) if maximization of α is replaced by minimization of $-\alpha$ over Δ. In our experiment additional speed is gained by adapting the trust-region norm $|\delta|_\infty = \max\{|\delta_1|, \ldots, |\delta_m|\}$ to the special form $\Delta = [-1, 1]^m$ of the set C, and by using the standard model $\phi^\sharp$ of $a_-(\delta) = -\alpha(A(\delta))$ is used. With these arrangements the method converges fast and reliably to a local optimum, which in the majority of cases can be certified *a posteriori* as a global one.

In order to justify the use of the standard model in Algorithm 1 we have to show that a_- is upper-C^1, or at least that its standard model is strict. Here the situation is more delicate than in section 6.1. We start by observing the following.

Lemma 8 *Suppose all active eigenvalues of $A(\delta)$ at δ are semi-simple. Then $a_-(\delta) = -\alpha\left(A(\delta)\right)$ is Clarke subdifferentiable in a neighborhood of δ.*

Proof This follows from [12]. A very concise proof that semi-simple eigenvalue functions are locally Lipschitz could also be found in [33]. Recall that an eigenvalue is semi-simple if its geometric and algebraic multiplicities are equal. □

That $a_\pm(\delta) = \pm\alpha(A(\delta))$ may fail to be locally Lipschitz was first observed in [12]. This may lead to difficulties when a_+ is minimized. In our numerical testing $a_-(\delta) = -\alpha\left(A(\delta)\right)$ is minimized, and we have observed that a_- behaves consistently like an upper-C^1 function. We expect a_- to have a strict standard model if all active eigenvalues of $A(\delta^*)$ are semi-simple, and in [2, Chap. V. C] it is shown that $\phi^\sharp$ is at least directionally strict. See [35] for more information.

Theorem 4 (Worst-case spectral abscissa on Δ) *Let $\delta^j \in \Delta$ be the sequence generated by Algorithm 1 for program (27), where the standard model $\phi^\sharp$ of a_- is used. Suppose that at least one accumulation point δ^* of the sequence δ^j is such that every active eigenvalue at $A(\delta^*)$ is simple. Then the entire sequence δ^j converges to this point δ^*, which is then a critical point of (27).*

Proof We apply Theorem 1 to get convergence in the sense of subsequences, and we use the Łojasiewicz inequality for a_- to prove that the entire sequence converges to δ^*, see [7, 37] for the argument. $\square$

6.3 Distance to Instability

Our third problem is related to the above and concerns computation of the structured distance to instability of (22)-(24). Suppose A in (22) is nominally stable, i.e., $A(\delta)$ is stable at the nominal $\delta = 0$. Then the structured distance to instability is defined as

$$d^* = \max\{d > 0 : A(\delta) \text{ stable for all } |\delta|_\infty < d\}, \tag{29}$$

where $A(\delta)$ is given by (28), and $|\delta|_\infty = \max\{|\delta_1|, \ldots, |\delta_m|\}$. Equivalently, we may consider the following constrained optimization program

$$\begin{aligned} &\text{minimize } t \\ &\text{subject to } -t \le \delta_i \le t \\ &\qquad\qquad\quad \alpha\,(A(\delta)) \ge 0 \end{aligned} \tag{30}$$

with decision variable $x = (t, \delta) \in \mathbb{R}^{m+1}$. Introducing the convex set $C = \{(t, \delta) : -t \le \delta_i \le t, i = 1, \ldots, m\}$, this can be transformed to program (1) if we minimize an exact penalty objective $f(x) = t + c \max\{0, -\alpha\,(A(\delta))\}$ with a penalty constant $c > 0$ over C.

It is clear that the objective of f has essentially the same properties as a_-. It suffices to argue that $\partial \max\{0, -\alpha(A(\delta))\} = \text{co}\,(\{0\} \cup \partial a_-(\delta))$ at points δ where a_- is locally Lipschitz and $a_-(\delta) = 0$, with 'co' denoting convex hull. Indeed, the inclusion $\subset$ holds in general. For the reverse inclusion it suffices to observe that $0 \in \partial \max\{0, -\alpha(A(\delta))\}$ for those δ where $a_-(\delta) = 0$. This is clear, because 0 is a minorant of this max function. We may then use the following

Lemma 9 *Suppose $f = \max\{f_1, f_2\}$ and f_i has a strict model ϕ_i. Then $\phi = \max\{\phi_1, \phi_2\}$ is a strict model of f at those x where $\partial f(x) = co\,(\partial f_1(x) \cup \partial f_2(x))$.*

Proof In fact, the only axiom which does not follow immediately is (M_1). We only know $\partial_1 \phi_i(x, x) \subset \partial f_i(x)$, so $\partial_1 \phi(x, x) = \text{co}\,(\partial_1 \phi_1(x, x) \cup \partial_1 \phi_2(x, x)) \subset \text{co}\,(\partial f_1(x) \cup \partial f_2(x))$. For those x where the maximum rule is exact, this implies indeed $\partial_1 \phi(x, x) \subset \partial f(x)$. $\square$ $\square$

This means that we can use the model $\phi(\delta', t', \delta, t) = t' + c \max\{0, \phi^\sharp(\delta', \delta)\}$ in Algorithm 1 to solve (30), naturally with the same proviso as in Section 6.2, where we need the standard model $\phi^\sharp$ of a_- to be strict.

7 Experiments

In this part experiments with Algorithm 1 applied to programs (25), (27) and (29) are reported.

Table 1 Benchmarks for worst-case H_∞-norm on Δ

♯	Benchmark	n	Structure	$\underline{h}$	h^*	$\overline{h}$	t^*	$\overline{h}/h^*$	t_{wc}/t^*
1	Beam1	11	$1^3 3^1 1^1$	1.70	1.71	1.70	1.02	0.99	13.29
2	Beam2	11	$1^3 3^1 1^1$	1.29	1.29	1.29	0.36	1	32.68
3	DC motor 1	7	$1^1 2^2$	0.72	0.72	0.72	0.51	1.01	14.49
4	DC motor 2	7	$1^1 2^2$	0.50	0.50	0.50	0.13	1	45.02
5	DVD driver 1	10	$1^1 3^3 1^1 3^1$	45.45	45.45	45.46	0.23	1	189.31
6	Four-disk system 1	16	$1^1 3^5 1^4$	3.50	4.56	3.50	0.44	0.77	343.35
7	Four-disk system 2	16	$1^1 3^5 1^4$	0.69	0.68	0.69	0.34	1.01	558.03
8	Four-tank system 1	12	1^4	5.60	5.60	5.60	0.32	1	5.72
9	Four-tank system 2	12	1^4	5.60	5.57	5.60	0.29	1	7.32
10	Hard disk driver 1	22	$1^3 2^4 1^4$	243.9	7526.6	Inf	0.96	Inf	73.10
11	Hard disk driver 2	22	$1^3 2^4 1^4$	0.03	0.03	0.03	0.20	1.12	314.92
12	Hydraulic servo 1	9	1^9	1.17	1.17	1.17	0.34	1	10.94
13	Hydraulic servo 2	9	1^9	0.7	0.70	0.7	0.33	1.01	11.69
14	Mass-spring 1	8	1^2	3.71	6.19	3.71	0.31	0.60	3.54
15	Mass-spring 2	8	1^2	6.84	6.84	7.16	0.13	1.05	7.05
16	Missile 1	35	$1^3 6^3$	5.12	5.15	5.12	0.46	0.99	272.54
17	Missile 2	35	$1^3 6^3$	1.83	1.82	1.83	0.22	1	1183.5
18	Filter 1	8	1^1	4.86	4.86	4.86	0.32	1	3.41
19	Filter 2	3	1^1	2.63	2.64	2.63	0.27	1	4.06
20	Filter-Kim 1	3	1^2	2.95	2.96	2.95	0.24	1	3.4
21	Filter-Kim 2	3	1^2	2.79	2.79	2.79	0.07	1	12.95
22	Satellite 1	11	$1^1 6^1 1^1$	0.16	0.17	0.16	0.33	1	86.17
23	Satellite 2	11	$1^1 6^1 1^1$	0.15	0.15	0.15	0.70	1	41.09
24	Mass-spring-damper 1	13	1^1	7.63	8.85	7.63	0.21	0.86	4.88
25	Mass-spring-damper 2	13	1^1	1.65	1.65	1.65	0.08	1	13.70
26	Robust Toy 1	3	$1^1 2^1$	0.12	0.12	0.12	0.56	1	4.24
27	Robust Toy 2	3	$1^2 2^2 3^1$	20.85	21.70	20.91	0.24	0.96	29.19

7.1 Worst-case H_∞-norm

We apply Algorithm 1 to program (25). Table 1 shows the result for 27 benchmark systems, where n is the number of states, and column 4 gives the uncertain structure $[r_1 \dots r_m]$ according to (24). An expression like $1^3 3^1 1^1$ corresponds to $[r_1\, r_2\, r_3\, r_4\, r_5] = [1\ 1\ 1\ 3\ 1]$. The values achieved by Algorithm 1 are h^* in column 6, obtained in t^* seconds CPU. To certify h^* we use the function WCGAIN of [52], which is a branch-and-bound method tailored to program (25). WCGAIN computes a lower and an upper bound $\underline{h}, \overline{h}$ shown in columns 5,7 within t_{wc} seconds. It also provides $\underline{\delta} \in \Delta$ realizing the lower bound.

The results in Table 1 show that h^* is certified by WCGAIN in the majority of cases 1-5,7-9,11-13,16,17. Case 15 leaves a doubt, while cases 6,14,24 are failures of WCGAIN, because our local solver already gets a value larger than the upper bound of WCGAIN. Based on the medians, Algorithm 1 is approximately 18 times faster than WCGAIN. The fact that the results of both methods are in good agreement can be understood as an endorsement of our approach.

7.2 Robust Stability over $\mathbf{\Delta}$

In our second test Algorithm 1 is applied to program (27). We have used a bench of 32 cases gathered in Table 2, and Algorithm 1 converges to the value α^* in t^* seconds.

Algorithm 2 Zheng-method for global optimization $\alpha^* = \max_{x\in\mathbf{\Delta}} f(x)$

▷ **Step 1** (`Initialize`). Choose initial $\alpha < \alpha^*$.

▷ **Step 2** (`Iterate`). Compute $\alpha^+ = \dfrac{\int_{[f\geq\alpha]} f(x)\, d\mu(x)}{\mu[f \geq \alpha]}$.

▷ **Step 3** (`Stopping`). If progress of α^+ over α is marginal, stop, otherwise update α by α^+ and loop on with step 2.

To certify α^* we have implemented Algorithm 2, known as integral global optimization, or as the Zheng-method (ZM), based on [49].

Here μ is any continuous finite Borel measure on Δ. Numerical implementations use Monte-Carlo to compute the integral, and we refer to [49] for details. Our numerical tests are performed with $2000 \cdot m$ initial samples, and stopping criterion variance $= 10^{-12}$; cf. [49]. The result obtained by ZM are α_{ZM} obtained in t_{ZM} seconds CPU.

A favorable feature of ZM is that it can be initialized with the lower bound α^*, and this leads to a significant speedup. Altogether ZM and Algorithm 1 are in very good agreement on the test set, which we consider an argument in favor of our approach.

7.3 Distance to Instability

In this last part we apply Algorithm 1 to (29) using the test bench of Table 3, which can be found in [19]. The distance computed by Algorithm 1 is d^* in column 2 of Table 3. We certify d^* using ZM [49] and by comparing to the local method of [19].

To begin with, ZM is used in the following way. For a given d^* and a confidence level $\gamma = 0.05$ we compute

$$\underline{\alpha} = \max\{\alpha(A(\delta)) : \delta \in (1-\gamma)d^*\Delta\} \tag{31}$$

and

$$\overline{\alpha} = \max\{\alpha(A(\delta)) : \delta \in (1+\gamma)d^*\Delta\}. \tag{32}$$

If $\underline{\alpha} < 0$ and $\overline{\alpha} > 0$ then d^* is certified by ZM with that confidence level γ. This happens in all cases except 87, where ZM failed due to the large size.

We also compared d^* to the result d_{F} of the technique [19], which is a sophisticated tool tailored to problem (29). Column 6 of Table 3 shows perfect agreement on the test set from [19]. Given the highly dedicated character of [19], this can be understood as an endorsement of our optimization-based approach.

Following [26] one can certify robust stability over Δ by showing that the value of the following polynomial optimization problem is strictly positive:

$$\begin{array}{ll} \text{minimize} & \det(H(\delta)) \\ \text{subject to} & \delta \in \Delta \end{array} \tag{33}$$

Table 2 Benchmarks for worst-case spectral abscissa (27)

♯	Benchmark	n	Structure	α^*	α_{ZM}	t^*	t_{ZM}
28	Beam3	11	$1^3 3^1 1^1$	−1.2e-7	−1.2e-7	0.19	32.70
29	Beam4	11	$1^3 3^1 1^1$	−1.7e-7	−1.7e-7	0.04	33.00
30	Dashpot system 1	17	1^6	0.0186	0.0185	0.23	90.25
31	Dashpot system 2	17	1^6	−1.0e-6	−1.0e-6	0.39	39.63
32	Dashpot system 3	17	1^6	−1.6e-6	−1.6e-6	0.08	39.70
33	DC motor 3	7	$1^1 2^2$	−0.0010	−0.0010	0.02	20.63
34	DC motor 4	7	$1^1 2^2$	−0.0010	−0.0010	0.02	20.74
35	DVD driver 2	10	$1^1 3^3 1^1 3^1$	−0.0165	−0.0165	0.04	49.29
36	Four disk system 3	16	$1^1 3^5 1^4$	0.0089	0.0088	0.10	159.61
37	Four disk system 4	16	$1^1 3^5 1^4$	−7.5e-7	−7.5e-7	0.29	73.86
38	Four disk system 5	16	$1^1 3^5 1^4$	−7.5e-7	−7.5e-7	0.29	74.36
39	Four tank system 3	12	1^4	−6.0e-6	−6.0e-6	0.17	25.81
40	Four tank system 4	12	1^4	−6.0e-6	−6.0e-6	0.02	26.20
41	Hard disk driver 3	22	$1^3 2^4 1^4$	266.70	266.70	0.09	1252.20
42	Hard disk driver 4	22	$1^3 2^4 1^4$	−1.6026	−1.6026	0.06	80.40
43	Hydraulic servo 3	9	1^9	−0.3000	−0.3000	0.04	51.41
44	Hydraulic servo 4	9	1^9	−0.3000	−0.3000	0.02	50.95
45	Mass-spring 3	8	1^2	−0.0054	−0.0054	0.01	31.59
46	Mass-spring 4	8	1^2	−0.0368	−0.0370	0.01	16.94
47	Missile 3	35	$1^3 6^3$	22.6302	22.1682	0.07	104.18
48	Missile 4	35	$1^3 6^3$	−0.5000	−0.5000	0.07	51.78
49	Missile 5	35	$1^3 6^3$	−0.5000	−0.5000	0.07	52.24
50	Filter 3	8	1^1	−0.0148	−0.0148	0.06	7.05
51	Filter 4	8	1^1	−0.0148	−0.0148	0.02	6.89
52	Filter-Kim 3	3	1^2	−0.2500	−0.2500	0.01	12.83
53	Filter-Kim 4	3	1^2	−0.2500	−0.2500	0.01	12.90
54	Satellite 3	11	$1^1 6^1 1^1$	3.9e-5	3.9e-5	0.02	44.02
55	Satellite 4	11	$1^1 6^1 1^1$	−0.0269	−0.0269	0.02	26.02
56	Satellite 5	11	$1^1 6^1 1^1$	−0.0268	−0.0268	0.02	26.08
57	Mass-spring-damper 3	13	1^1	0.2022	0.2022	0.01	8.30
58	Mass-spring-damper 4	13	1^1	−0.1000	−0.1000	0.01	6.91
59	Mass-spring-damper 5	13	1^1	−0.1000	−0.1000	0.01	6.94

where $H(\delta)$ is the so-called Hermite-matrix [26]. For $\Delta = [-1, 1]^m$ in (33), the method [31] gives finite convergence. We follow [26] and apply GloptiPoly [27] to (33), where Maple 14 is used beforehand to compute the determinant of $H(\delta)$ formally. Based on (31) and (32) this leads to a procedure to certify or reject our heuristic d^*.

The method was indeed able to certify d^* in cases 20, 21, 26 and 27. In the tests of Table 3 the method was not able to furnish a decision even when the feasibility radius of the SDP-solver SeDuMi was enlarged to 10^3, and a large number of LMIs was considered. The bottleneck of the proposed method appears to be slow convergence $v_k \to v^*$, the fact that lower bounds cannot be taken into account in (33), and the necessity to compute the

Table 3 Benchmarks for distance to instability (29), available in [53]

♯	Benchmark	n	Structure	d^*	d_F/d^*	D_{ZM}	t^*	t_{ZM}
60	Academic example	5	1^1	0.79	1	$\surd$	0.15	7.3
61	Academicexample	4	1^3	3.41	1	$\surd$	0.13	23.9
62	Academic example	4	2^2	0.58	1	$\surd$	0.15	97.4
63	Inverted pendulum	4	1^3	0.84	1	$\surd$	0.22	24.7
64	DC motor	4	$1^3 2^1\ 1^1$	1.25	1	$\surd$	0.19	37.7
65	Bus steering system	9	$2^1 3^1$	1.32	0.99	$\surd$	0.37	13.8
66	Satellite	9	$2^1 1^2$	1.01	0.99	$\surd$	0.3	20.2
67	Bank-to-turn missile	6	1^4	0.60	0.99	$\surd$	0.17	167.7
68	Aeronautical vehicle	8	1^4	0.61	0.99	$\surd$	0.19	38.9
69	Four-tank system	10	1^4	6.67	0.99	$\surd$	0.27	24.9
70	Re-entry vehicle	6	$3^1 2^1 3^1$	6.20	1	$\surd$	0.44	21.8
71	Missile	14	1^4	7.99	1	$\surd$	0.25	24.9
72	Cassini spacecraft	17	1^4	0.06	1	$\surd$	0.13	25.1
73	Mass-spring-damper	7	1^6	1.17	1	$\surd$	0.17	2536.3
74	Spark ignition engine	4	1^7	1.22	0.99	$\surd$	0.41	42.8
75	Hydraulic servo system	8	1^8	1.50	0.99	$\surd$	0.41	62.8
76	Academic example	41	$2^1 1^3$	1.18	0.99	$\surd$	0.57	36.5
77	Drive-by-wire vehicle	4	$1^2 2^7$	1	0.99	$\surd$	0.96	97.0
78	Re-entry vehicle	7	$1^3 6^1 4^1$	1.02	0.98	$\surd$	0.42	132.4
79	Space shuttle	34	1^9	0.79	0.99	$\surd$	0.8	60.9
80	Rigid aircraft	9	1^{14}	5.42	1	$\surd$	0.54	252.5
81	Fighter aircraft	10	$3^1 15^1 1^6 2^1 1^1$	0.59	0.99	$\surd$	1.31	171.3
82	Flexible aircraft	46	1^{20}	0.22	0.99	$\surd$	1.26	180.3
83	Telescope mockup	70	1^{20}	0.02	0.99	$\surd$	1.37	274.8
84	Hard disk drive	29	$1^8 2^4 1^1 1$	0.82	1	$\surd$	2.87	202.1
85	Launcher	30	$1^2 2^2 1^2 3^1 6^1 1^{12} 2^8$	1.16	0.99	$\surd$	4.08	271.2
86	Helicopter	12	30^4	0.08	0.99	$\surd$	0.85	70.7
87	Biochemical network	7	39^{13}	1.4e-3	1	failed	36.76	-

determinant of $H(\delta)$ formally, which is impossible for matrices larger that 7×7. In all other aspects the method remains very promising.

8 Conclusion

We have presented a bundle trust-region method for nonsmooth, nonconvex minimization, where cutting planes are tangents to a convex local model $\phi(\cdot, x)$ of f, and where a trust-region strategy replaces the proximity control mechanism. Global convergence of our method was proved under natural hypotheses.

By way of an example we have shown that the standard approach in trust-region methods based on the Cauchy point fails for nonsmooth functions. We have identified a particular class $\mathscr{S}$ of nonsmooth functions, where the Cauchy point argument can be salvaged. Functions in $\mathscr{S}$, even when nonsmooth, can be minimized as if they were smooth. The class $\mathscr{S}$ must therefore be regarded as atypical in a nonsmooth optimization program, convex functions with a genuine nonsmoothness are not in $\mathscr{S}$.

Algorithm 1 was validated numerically on a test set of 87 problems in automatic control, where the versatility of Algorithm 1 with regard to the choice of the norm was exploited. We were able to compute good quality lower bounds for three NP-hard optimization problems related to the analysis of parametric robustness in system theory. In the majority of cases, posterior application of a global optimization technique allowed us to certify these results as globally optimal.

References

1. Absil, P.A., Mahony, R., Andrews, B.: Convergence of the iterates of descent methods for analytic cost functions. SIAM J. Optim. **16**(2), 531–547 (2005)
2. Apkarian, P., Dao, M.N., Noll, D.: Parametric robust structured control design. IEEE Trans. Autom. Control. **60**(7), 1857–1869 (2015)
3. Apkarian, P., Noll, D.: Nonsmooth H_∞ synthesis. IEEE Trans. Automat. Control. **51**(1), 71–86 (2006)
4. Apkarian, P., Noll, D.: Nonsmooth optimization for multidisk H_∞ synthesis. Eur. J. Control. **12**(3), 229–244 (2006)
5. Apkarian, P., Noll, D.: IQC analysis and synthesis via nonsmooth optimization. Syst. Control Lett. **55**(12), 971–981
6. Apkarian, P., Noll, D., Prot, O.: A proximity control algorithm to minimize non-smooth and non-convex semi-infinite maximum eigenvalue functions. J. Convex Analysis **16**, 641–666 (2009)
7. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. J. Math. Oper. Res. **35**(2), 2010
8. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
9. Bonnans, F., Gilbert, J.-Ch., Lemaréchal, C., Sagastizábal, C.: Numerical optimization - theoretical and practical aspects, 2nd Edition. Springer-Verlag (2006)
10. Borwein, J.M., Moors, W.B.: A chain rule for essentially strictly differentiable Lipschitz functions. SIAM J. Optim. **8**, 300–308 (1998)
11. Braatz, R.D., Young, P.M., Doyle, J.C., Morari, M.: Computational complexity of μ calculation. IEEE Trans. Autom. Control **39**, 1000–1002 (1994)
12. Burke, J.V., Overton, M.L.: Differential properties of the spectral abscissa and the spectral radius for analytic matrix-valued mappings. Nonlinear Anal. **23**(4), 467–488 (1994)
13. Clarke, F.H.: Optimization and nonsmooth analysis. John Wiley, New York (1983)
14. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-region methods. MPS/SIAM series on optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2000)
15. Daniilidis, A., Georgiev, P.: Approximate convexity and submonotonicity. J. Math. Anal. Appl. **291**, 117–144 (2004)
16. Dao, M.N.: Bundle method for nonconvex nonsmooth constrained optimization. J. Convex Analysis, to appear (2015)
17. Dao, M.N., Gwinner, J., Noll, D., Ovcharova, N.: Nonconvex bundle method with application to a delamination problem. arXiv:1401.6807v1 [math.OC] (2014)
18. Dennis, J.E., Li, S.B., Tapia, R.A.: A unified approach to global convergence of trust-region methods for nonsmooth optimization. Math. Programming **68**, 319–346 (1995)
19. Fabrizi, A., Roos, C., Biannic, J.M.: A detailed comparative analysis of lower bound algorithms. European Control Conference 2014, Jun 2014, Strasbourg, France
20. Frangioni, A.: Generalized bundle methods. SIAM J. Optim. **13**(1), 117–156 (2002)
21. Fuduli, A., Gaudioso, M., Giallombardo, G.: A DC piecewise affine model and a bundling technique in nonconvex nonsmooth optimization. Optimization Method and Software **19**, 89–102 (2004)
22. Hare, W., Sagastizábal, C.: A redistributed proximal bundle method for nonconvex optimization. SIAM J. Optim. **20**(5), 2442–2473 (2010)

23. Helmberg, C., Kiwiel, K.C.: A spectral bundle method with bounds. Math. Programming **93**, 173–194 (2002)
24. Helmberg, C., Oustry, F. In: Vandenberghe, L., Saigal, R., Wolkowitz, H. (eds.): Bundle methods to minimize the maximum eigenvalue function. Handbook of Semidefinite Programming. Theory, Algorithms and Applications, vol. 27 (2000)
25. Helmberg, C., Rendl, F.: Spectral bundle method for semidefinite programming. SIAM J. Optimization **10**, 673–696 (2000)
26. Henrion, D., Lasserre, J.B.: GloptiPoly: global optimization over polynomials with Matlab and SeDuMi. ACM Trans. Math. Software **29**(2), 165–194 (2003)
27. Henrion, D., Arzelier, D., Peaucelle, D., Lasserre, J.-B.: On parameter-dependent Lyapunov functions for robust stability of linear systems. 43rd IEEE Conf. Dec. Con. Atlantis, Paradise Island, Bahamas (2004)
28. Hiriart-Urruty, L.emaréchal.: Convex Analysis and Minimization Algorithms, vol. I and II: Advanced Theory and Bundle Methods, vol. 306 of Grundlehren der mathematischen Wissenschaften. Springer Verlag, Heidelberg, Berlin (1993)
29. Kiwiel, C.: An aggregate subgradient method for nonsmooth convex minimization. Math. Programming **27**, 320–341 (1983)
30. Klatte, D., Kummer, B.: Nonsmooth equations in optimization. Regularity, Calculus, Methods and applications. Nonconvex optimization and applications, vol. 60. Kluwer Academic Publishers (2002)
31. Lasserre, J.-B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, 796–817 (2001)
32. Le, T.H.A., Huynh, V.N., Pham, D.T., Ismael, A., Vaz, F., Vicente, L.N.: Globally convergent DC trust-region methods. J. Glob. Optim. **59**, 209–225 (2014)
33. Lui, S.H.: Pseudospectral mapping theorem II. Electron. Trans. Numer. Anal. **38**, 168–183 (2011)
34. Mifflin, R.: Semismooth and semiconvex functions in optimization. SIAM J. Control. Optim. **15**(6), 959–972 (1977)
35. Moro, J., Burke, J.V., Overton, M.L.: On the Lidskii-Vishik-Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. SIAM J. Matrix Anal. Appl. **18**(4), 793–817 (1997)
36. Noll, D.: Cutting plane oracles to minimize non-smooth non-convex functions. Set-Valued Var. Anal. **18**(3-4), 531–568 (2010)
37. Noll, D.: Convergence of non-smooth descent methods using the Kurdyka-Łojasiewicz inequality. J. Optim. Theory Appl. **160**(2), 553–572 (2014)
38. Noll, D., Prot, O., Rondepierre, A.: A proximity control algorithm to minimize nonsmooth and nonconvex functions. Pac. J. Optim. **4**(3), 571–604 (2008)
39. Poljak, S., Rohn, J.: Checking robust nonsingularity is NP-hard. Math. Cont. Sig. Sys. **6**, 1–9 (1993)
40. Powell, M.J.D.: General algorithms for discrete nonlinear approximation calculations, Report DAMTP 1983/NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge. Cambridge, England (1983)
41. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer Verlag (2004)
42. Ruszczyński, A.: Nonlinear Optimization. Princeton University Press (2007)
43. Sagastizábal, C.: Composite proximal bundle method. Math. Programming **140**(1), 189–233 (2013)
44. Sartenaer, A.: Armijo-type condition for the determination of a Generalized Cauchy Point in trust region algorithms using exact or inexact projections on convex constraints. Belgian Journal of Operations Research, Statistics and Computer Science **33**(4), 61–75
45. Schramm, H.: Eine Kombination von Bundle- und Trust-Region-Verfahren zur Lösung nicht-differenzierbarer Optimierungsprobleme. Bayreuther Mathematische Schriften, 30, Bayreuth (1989)
46. 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**, 121–152 (1992)
47. Spingarn, J.E.: Submonotone subdifferentials of Lipschitz functions. Trans. Amer. Math. Soc. **264**(1), 77–89 (1981)
48. Yuan, Y.: Conditions for convergence of trust region algorithms for nonsmooth optimization. Math. Programming **31**, 220–228 (1985)
49. Zheng, Q., Zhuang, D.: Integral global minimization: algorithms, implementations, and numerical tests. J. Glob. Optim. **7**, 421–454 (1995)
50. Zhou, K., Doyle, J.C., Glover, K.: Robust and Optimal Control. Prentice Hall, New Jersey (1996)
51. Zowe, J. In: Clarke, F.H., Demyanov, V.F., Gianessi, F. (eds.): The BT-Algorithm for minimizing a nonsmooth functional subject to linear constraints. In Nonsmooth Optimization and Related Topics. Plenum Press (1989)
52. Robust Control Toolbox 5.0. MathWorks, Natick, MA, USA, Sept 2013
53. SMAC Toolbox, ONERA 2012-15, http://w3.onera.fr/smac