Optim Lett
DOI 10.1007/s11590-016-1068-7

ORIGINAL PAPER

On the minimization of possibly discontinuous functions by means of pointwise approximations

E. G. Birgin[1] · **N. Krejić**[2] · **J. M. Martínez**[3]

Received: 19 November 2015 / Accepted: 19 July 2016

Abstract A general approach for the solution of possibly discontinuous optimization problems by means of pointwise (perhaps smooth) approximations will be proposed. It will be proved that sequences generated by pointwise approximation techniques eventually satisfy well justified stopping criteria. Numerical examples will be given.

Keywords Discontinuous functions · Pointwise approximations · Smoothing · Minimization

---

This work has been partially supported by the Brazilian agencies FAPESP (Grants 2010/10133-0, 2013/03447-6, 2013/05475-7, 2013/07375-0, and 2014/18711-3) and CNPq (Grants 309517/2014-1 and 303750/2014-6) and by the Serbian Ministry of Education, Science, and Technological Development (Grant 174030).

✉ E. G. Birgin
egbirgin@ime.usp.br

N. Krejić
natasak@uns.ac.rs

J. M. Martínez
martinez@ime.unicamp.br

1 Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, São Paulo, SP 05508-090, Brazil

2 Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 4, Novi Sad, Serbia

3 Department of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing, State University of Campinas, Campinas, SP 13083-859, Brazil

Published online: 27 July 2016

1 Introduction

Many Engineering problems involve the minimization of discontinuous functions. For example, process models with discontinuous investment costs and fixed charges [17], continuous review (s,q) inventory systems with constant demand and batch arrivals [10], design of flow sheets for systems that satisfy fixed demands of steam, electricity, and mechanical power [12], models for expansion of capacity of telecommunications networks [16], capacity and flow assignment problems [11], and optimal plastic design [15]. Different techniques have been considered for solving these problems, like integer programming [11], decomposition [16], simulated annealing [12], among others.

The problem considered in this paper is

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\emptyset \neq \Omega \subseteq \mathbb{R}^n$. (Note that $f(x) < \infty$ for all $x \in \mathbb{R}^n$.) We will consider problem (1) with no assumptions on the continuity of the objective function. The idea is to replace the objective function with (possibly smooth) approximations with which it is easier to deal numerically. We will derive conditions under which optima of the approximating functions are suitable approximations for the original problem.

Many times, practical methods related with this approach rely on smoothing ideas. Smoothing methods for minimizing non-differentiable continuous function have been extensively considered in the optimization literature (see [9] and the references therein) whereas the case in which the objective function is discontinuous is rarely encountered [13,20]. Direct search methods and methods for discontinuous bounded factorable functions have been introduced recently in [18] and [19], respectively.

Uniform convergence of a sequence of continuous functions implies continuity of the limit function f. So, in order to exploit the approximation of a discontinuous function by continuous ones we must rely in pointwise approximation. In this work we deal with the problem of minimizing a (possible discontinuous) function f with constraints. The goal is to investigate the minimization properties of pointwise approximations of f.

This paper is organized as follows. In Sect. 2, we assume that finding almost global minimizers of the approximating functions is possible and we investigate the properties of an algorithm that proceeds by almost-minimization of such approximations with respect to the global minimization of the objective function. In Sect. 3 we address the most frequent situation in which smooth global minimization is not affordable. Assuming smoothness of the constraints, an algorithm is defined that, at each iteration, finds an Approximate KKT point of a smooth approximation of f. We prove that the points so far obtained are interesting candidates to approximate a solution of the original problem since they obey a theoretical property that is shared by local minimizers. In Sect. 4 we provide some numerical examples of the approach described in the previous sections. Some conclusions are drawn in the last section.

Notation $\mathbb{N}$ denotes the set of natural numbers $\{1, 2, \dots\}$. $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^n$.

2 Global results

Let $\{f_k\}$ be a sequence of functions such that $f_k : \mathbb{R}^n \to \mathbb{R}$. Assumption A1 below states that this sequence converges pointwise to f.

Assumption A1 *For all $k \in \mathbb{N}$, $f_k : \mathbb{R}^n \to \mathbb{R}$ is such that*

$$\lim_{k\to\infty} f_k(x) = f(x) \text{ for all } x \in \Omega.$$

For an arbitrary nonempty set $A \subseteq \mathbb{R}$ we say that $s \in \mathbb{R}\cup\{-\infty\}$ is the infimum of A if s is the biggest lower bound of A. In Theorem 2.1 we will prove that, under Assumption A1, the infimum of f_k is, asymptotically, not greater than the infimum of f onto Ω.

Theorem 2.1 *If Assumption A1 holds, we have that*

$$\liminf_{k\in\mathbb{N}} \inf_{x\in\Omega} f_k(x) \leq \inf_{x\in\Omega} f(x). \tag{2}$$

Proof Let $a \in \mathbb{R}$ be such that

$$a > \inf_{x\in\Omega} f(x).$$

Therefore, there exists $z \in \Omega$ such that $f(z) < a$. Then, by Assumption A1, for all k large enough we have that $f_k(z) < a$. Therefore, for all k large enough,

$$\inf_{x\in\Omega} f_k(x) < a.$$

This implies that

$$\liminf_{k\in\mathbb{N}} \inf_{x\in\Omega} f_k(x) \leq a.$$

Thus, since $a > \inf_{x\in\Omega} f(x)$ was arbitrary,

$$\liminf_{k\in\mathbb{N}} \inf_{x\in\Omega} f_k(x) \leq \inf_{x\in\Omega} f(x),$$

as we wanted to prove. □

The usefulness of the sequence of functions $\{f_k\}$ for solving (1) relies on the possibility that minimizing f_k could be much easier than minimizing f. For example, if the functions f_k have enough structure, one can use structure-oriented global optimization methods for their minimization [5]. The following algorithm describes a simple way to use the pointwise approximation sequence f_k with the purpose of minimizing f.

Algorithm 2.1. Let $\tau \in (0, 1)$ and a sequence $\{f_k\}$ satisfying Assumption A1 be given. Initialize $k \leftarrow 1$ and $\varepsilon_1 > 0$.

Step 1. Employing some global optimization method, suitable for functions with the structure of f_k, find $x^k \in \Omega$ such that

$$f_k(x^k) \le \inf_{x\in\Omega} f_k(x) + \varepsilon_k. \tag{3}$$

Step 2. Set $\varepsilon_{k+1} = \tau\varepsilon_k$, $k \leftarrow k+1$, and go to Step 1.

Property (3) is usually guaranteed by global optimization solvers. Theorem 2.2 provides the elements to prove the main property of Algorithm 2.1, without using the specific structures of the function f or the pointwise approximations f_k.

Theorem 2.2 *If Assumption* A1 *holds and the sequence* $\{x^k\}$ *is generated by Algorithm* 2.1*, we have that*

$$\liminf_{k\in\mathbb{N}} f_k(x^k) \le \inf_{x\in\Omega} f(x). \tag{4}$$

Proof Let $a \in \mathbb{R}$ be such that $a > \inf_{x\in\Omega} f(x)$. Therefore, there exists $z \in \Omega$ such that $f(z) < a$. Define $\eta = (a - f(z))/2 > 0$. Since $f_k(z) \to f(z)$, there exists $k_0 \in \mathbb{N}$ such that $f_k(z) \le f(z) + \eta$ for all $k \ge k_0$. Therefore, for all $k \ge k_0$,

$$\inf_{x\in\Omega} f_k(x) \le f_k(z) \le f(z) + \eta.$$

Then, by (3),

$$f_k(x^k) \le \inf_{x\in\Omega} f_k(x) + \varepsilon_k \le f(z) + \eta + \varepsilon_k$$

for all $k \ge k_0$. But, since $\eta < a - f(z)$, this implies that

$$f_k(x^k) \le a + \varepsilon_k \text{ for all } k \ge k_0.$$

Since this inequality holds for all $a > \inf_{x\in\Omega} f(x)$, we have that

$$f_k(x^k) \le \inf_{x\in\Omega} f(x) + \varepsilon_k \text{ for all } k \ge k_0.$$

Since ε_k tends to zero, this implies the desired result. □

Observe that inequality (2) cannot be converted into an equality. In fact, assume that $f(x) = 0$ for all $x \in \mathbb{R}^n$, $f_k(x) = 0$ for all $x \ne 1/k$, and $f_k(1/k) = -1$. The sequence $f_k(x)$ converges to 0 for all $x \in \mathbb{R}^n$ and the right-hand side of (2) is 0. However, the left-hand side is -1. In Theorem 2.3 the equality is obtained through the assumption that $f_k(x)$ over-estimates $f(x)$. This assumption is stated below.

Assumption A2 *For all* $k \in \mathbb{N}$*,* $f_k : \mathbb{R}^n \to \mathbb{R}$ *is such that*

$$f_k(x) \ge f(x) \text{ for all } x \in \Omega.$$

Theorem 2.3 *If Assumptions* A1 *and* A2 *hold, we have that*

$$\liminf_{k\in\mathbb{N}} \inf_{x\in\Omega} f_k(x) = \inf_{x\in\Omega} f(x).$$

Proof Since $f_k(x) \geq f(x)$ we have that

$$\liminf_{k\in\mathbb{N}} \inf_{x\in\Omega} f_k(x) \geq \liminf_{k\in\mathbb{N}} \inf_{x\in\Omega} f(x) = \inf_{x\in\Omega} f(x).$$

Therefore, the thesis follows from Theorem 2.1. □

Theorem 2.4 *If Assumptions A1 and A2 hold and the sequence $\{x^k\}$ is generated by Algorithm 2.1, we have that*

$$\liminf_{k\in\mathbb{N}} f_k(x^k) = \inf_{x\in\Omega} f(x).$$

Proof It follows from $f_k(x^k) \leq \inf_{x\in\Omega} f_k(x) + \varepsilon_k$ and the fact that ε_k tends to zero. □

Corollary 2.1 *In addition to the hypotheses of Theorem 2.4, assume that x^* is a global minimizer of $f(x)$ onto Ω. Then, there exists $\mathbb{K}$, an infinite subsequence of $\mathbb{N}$, such that*

$$\lim_{k\in\mathbb{K}} f_k(x^k) = f(x^*).$$

Proof Firstly, note that, by the hypothesis, $f(x^*) = \inf_{x\in\Omega} f(x)$. The rest of the proof comes from the manipulation of the concepts of infimum and limit. □

Remarks The simple results presented here use neither continuity of f or its approximations f_k. Moreover, convergence of f_k to f does not need to be uniform and Ω is an arbitrary set. The main application of these results corresponds to the case in which the functions f_k are more smooth than f. For example, when f is not continuous while the approximations f_k have continuous derivatives, one can use smooth standard algorithms for minimizing f_k in order to obtain approximations of the minimum of f.

3 Affordable algorithm and stopping criterion

In this section we will assume that a sequence $\{f_k\}$ that satisfies Assumption A1 is available and that the set Ω has the form given in Assumption A3 below.

Assumption A3 *The domain Ω is the set of $x \in \mathbb{R}^n$ such that*

$$h(x) = 0 \text{ and } g(x) \leq 0, \tag{5}$$

where $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ admit continuous first derivatives for all $x \in \mathbb{R}^n$.

Moreover, we will use the following assumption.

Assumption A4 *The functions f_k admit continuous first derivatives for all $x \in \mathbb{R}^n$.*

Whereas Algorithm 2.1 assumes that we are able to find (almost) global minimizers of $f_k(x)$ subject to $x \in \Omega$, the algorithm defined in this section is less ambitious. We will only assume that it is possible to find almost stationary points of $f_k(x)$ onto Ω. Well established algorithms that are successful for that purpose will be called Standard Algorithms here. The main convergence property of Standard Algorithms is given below.

Property P1 *Assume that* $\{z^\ell\}$ *is a sequence generated by a Standard Algorithm applied to the minimization of* $f_k(z)$ *subject to* $h(z) = 0$ *and* $g(z) \leq 0$*, where* f_k *satisfies Assumption A4 and* h *and* g *are as in Assumption A3. Assume, moreover, that* $\{z^\ell\}$ *admits a limit point* $z^{\lim}$*. Then, at least one of the following two possibilities hold:*

1. *For all* $\varepsilon > 0$*, there exists an iterate* z^ℓ *such that*
$$\begin{aligned} \|\nabla f_k(z^\ell) + \nabla h(z^\ell)\lambda^\ell + \nabla g(z^\ell)\mu^\ell\| &\leq \varepsilon, \\ \|h(z^\ell)\| &\leq \varepsilon, \\ \|g(z^\ell)_+\| &\leq \varepsilon, \\ \min\{\mu_i^\ell, -g_i(z^\ell)\} &\leq \varepsilon \text{ for all } i = 1, \ldots, p, \end{aligned} \tag{6}$$
for some $\lambda^\ell \in \mathbb{R}^m$ *and* $\mu^\ell \in \mathbb{R}^p_+$*, and the distance between* z^ℓ *and* $z^{\lim}$ *is smaller than* ε*.*
2. *The limit point* $z^{\lim}$ *is infeasible and stationary for the infeasibility measure. This means that*
$$\|h(z^{\lim})\| + \|g(z^{\lim})_+\| > 0 \text{ and } \nabla\left[\|h(z^{\lim})\|_2^2 + \|g(z^{\lim})_+\|_2^2\right] = 0. \tag{7}$$

Property P1 is satisfied by many well established smooth optimization algorithms, for example the Augmented Lagrangian method Algencan, described in [2,7]. Essentially, Property P1 means that Standard Algorithms applied to the minimization of $f_k(z)$ subject to $h(z) = 0$ and $g(z) \leq 0$, under Assumptions A3 and A4, converge to AKKT (Approximate KKT) points [3,7,14]. Note that, for the fulfillment of Property P1, no constraint qualifications are necessary at all. The fulfillment of Property P1 by existent methods when applied to the problems that consists on minimizing $f_k(z)$ onto Ω allows us to define the following algorithm.

Algorithm 3.1. Let $x^0 \in \mathbb{R}^n$, $\tau \in (0, 1)$, a sequence $\{f_k\}$ satisfying Assumptions A1 and A4, $\varepsilon > 0$, and $k_{\text{big}} \in \mathbb{N}$ be given. Initialize $k \leftarrow 1$ and $\varepsilon_1 \geq \varepsilon$.

Step 1. Employing some Standard Algorithm for minimizing $f_k(z)$ subject to $x \in \Omega$ (possibly starting from x^{k-1}), find x^k that satisfies (6) with ε substituted by ε_k.

Step 2. If $k \geq k_{\text{big}}$ and $\varepsilon_k \leq \varepsilon$, stop.

Step 3. Set $\varepsilon_{k+1} = \max\{\varepsilon, \tau\varepsilon_k\}$, $k \leftarrow k + 1$, and go to Step 1.

Due to the definitions given, the proof of the following theorem is straightforward.

Theorem 3.1 *Let Assumptions A1, A3, and A4 hold. Assume that the sequence* $\{x^k\}$ *is generated by Algorithm* 3.1*. Assume, further, that, for all* $k \in \mathbb{N}$*, the sequence potentially generated by the Standard Algorithm for minimizing* $f_k(z)$ *subject to* $z \in \Omega$

is bounded and that any limit point of such sequence belongs to Ω*. Then, Algorithm 3.1 is well-defined and stops satisfying* $k \geq k_{\text{big}}$ *and* (6)*.*

Proof By the hypothesis, for all $k \in \mathbb{N}$, the sequence generated by the Standard Algorithm for minimizing $f_k(z)$ subject to $z \in \Omega$ is bounded and, so, it admits a limit point. Also by the hypothesis, such limit point is feasible. Therefore it does not satisfy (7). This means that, by Property P1, there exists $\ell \in \mathbb{N}$ such that (6) holds. Therefore, Algorithm 3.1 is well defined for all $k \in \mathbb{N}$. Taking $k \geq k_{\text{big}}$ we obtain the desired result. □

Algorithm 3.1 is well defined under the assumptions that the sequences generated by the Standard Algorithm are bounded and that limit points of those sequences are feasible. Boundedness is generally guaranteed by assumptions on the constraints, for example, when some of these constraints define a box $\ell \leq x \leq u$. The non-existence of infeasible limit points is implied by the non-existence of infeasible stationary points of the constraints. This property depends only on the smooth constraints $h(x) = 0$ and $g(x) \leq 0$ and it is not possible to guarantee that such points do not exist, except in particular cases. For this reason, in practice, we must include a stopping criterion that detects that, possibly, the Standard Algorithm iteration is close to a limit point that fulfills (7).

In the case that the assumptions of Theorem 3.1 hold, even if the sequence $\{x^k\}$ converges to a point at which f is not differentiable, the gradient $\nabla f_k(x^k)$ tends to be a linear combination of the gradients of the constraints, as stated in (6). In particular, if there are no constraints at all, the gradient $\nabla f_k(x^k)$ tends to zero.

We have presented an algorithm that, under reasonable conditions, stops at a point that satisfies $k \geq k_{\text{big}}$ and (6). The question that naturally arises is: Is the stopping criterion based on $k \geq k_{\text{big}}$ and (6) reasonable? In other words, do these properties tell us something about the optimality of the solution found by Algorithm 3.1?

Let us give an example: Imagine, for a moment that we deal with smooth unconstrained minimization. Why we believe that the approximate annihilation of the gradient of the objective function is a suitable stopping criterion for related algorithms? The answer, of course, is that at local minimizers that gradient is null. Therefore, although we cannot be sure that we stop at local minimizers, at least we stop at points that verify a property that is satisfied by local minimizers.

In the same sense, we will prove now that, roughly speaking, if $\bar{x}$ is a local minimizer of $f(x)$ subject to $x \in \Omega$, the stopping conditions $k \geq k_{\text{big}}$ and (6) hold at $\bar{x}$. Note that the formulation of the stopping conditions $k \geq k_{\text{big}}$ and (6) is based on a sequence of approximating functions that satisfy Assumptions A1 and A4, but not necessarily A2. In order to guarantee that those conditions hold at a local minimizer we need to assume A2. The condition A2 plays, with respect to the stopping criterion based on $k \geq k_{\text{big}}$ and (6), the same role that constraint qualifications (for example, linear independence of active constraints) play with respect to KKT conditions in nonlinear programming: Constraint qualifications guarantee that local minimizers satisfy KKT conditions, but one considers that the fulfillment of KKT is a good symptom of optimality, independently of the satisfaction of the constraint qualification. (This observation should not be confused with the notion of AKKT conditions for smooth constrained optimization, which do not employ constraint qualifications at all.)

Theorem 3.2 *Assume that $\bar{x}$ is a local minimizer of $f(x)$ subject to $x \in \Omega$ and that Ω and $\{f_k\}$ fulfill Assumptions A1, A2, and A4. Then, there exists a sequence $\{x^k\}$ that converges to $\bar{x}$ and such that, for all $k_{\text{big}} \in \mathbb{N}$ and $\varepsilon > 0$, there exists $k \geq k_{\text{big}}$, $\lambda^k \in \mathbb{R}^m$, and $\mu^k \in \mathbb{R}^p_+$ satisfying (6).*

Proof Define

$$\bar{f}(x) = f(x) + \|x - \bar{x}\|_2^2$$

and

$$\bar{f}_k(x) = f_k(x) + \|x - \bar{x}\|_2^2$$

for all $x \in \Omega$ and $k \in \mathbb{N}$. By Assumption A1,

$$\lim_{k \to \infty} \bar{f}_k(x) = \bar{f}(x) \text{ for all } x \in \Omega. \tag{8}$$

Moreover, as $\bar{x}$ is a local minimizer of f onto Ω, there exists $\delta > 0$ such that, for all $x \in \Omega$ with $0 < \|x - \bar{x}\|_2 \leq \delta$, one has that $\bar{f}(x) > \bar{f}(\bar{x}) = f(\bar{x})$. Thus, $\bar{x}$ is a strict local minimizer of $\bar{f}(x)$ onto Ω.

For all $k \in \mathbb{N}$, let B_k be the closed Euclidean ball with center $\bar{x}$ and radius δ/k, and let S_k be the boundary of this ball. Since Ω is closed, $\Omega \cap B_k$ and $\Omega \cap S_k$ are compact. Therefore, by continuity, $\bar{f}_k(x)$ admits a global minimizer onto $\Omega \cap B_k$ for all $k \in \mathbb{N}$.

Since $\bar{x}$ is a local minimizer on B_k we have that, for all $x \in S_k$,

$$f(\bar{x}) \leq f(x).$$

Therefore,

$$f(\bar{x}) < f(x) + \frac{\|x - \bar{x}\|^2}{2}.$$

This implies that

$$f(\bar{x}) < f(x) + \|x - \bar{x}\|^2 - \frac{\|x - \bar{x}\|^2}{2}.$$

So, by the definition of $\bar{f}$,

$$f(\bar{x}) < \bar{f}(x) - \frac{\|x - \bar{x}\|^2}{2} = \bar{f}(x) - (\delta/k)^2/2. \tag{9}$$

By Assumption A1 there exists $j_k \in \mathbb{N}$ such that $j_k > j_{k-1}$ if $k > 1$ and

$$f_{j_k}(\bar{x}) < f(\bar{x}) + (\delta/k)^2/4.$$

Then, by (9),

$$\bar{f}_{j_k}(\bar{x}) < \bar{f}(x) - (\delta/k)^2/4.$$

Then, by Assumption A2,

$$\bar{f}_{j_k}(\bar{x}) < \bar{f}_{j_k}(x) - (\delta/k)^2/4.$$

Since this inequality holds for all $x \in S_k$ it turns out that the global minimizer of $\bar{f}_{j_k}(x)$ subject to $x \in B_k \cap \Omega$ does not belong to S_k.

Let z^k be the global minimizer of $\bar{f}_{j_k}(x)$ subject to $x \in B_k \cap \Omega$. We proved above that $\|z^k - \bar{x}\| < \delta/k$. Therefore z^k is a local minimizer of $f_{j_k}(x)$ subject to $h(x) = 0$ and $g(x) \leq 0$. Then, by the AKKT property of local minimizers of smooth functions [3,7], we have that there exist $x^k \in \mathbb{R}^n$, $\lambda^k \in \mathbb{R}^m$ and $\mu^k \in \mathbb{R}^p_+$ such that $\|x^k - z^k\| < \delta/(2k)$, and, in addition, (6) holds. Moreover, since $j_k > j_{k-1}$ for $k > 1$ it turns out that, eventually, $j_k \geq k_{\text{big}}$. This completes the proof. □

4 Numerical examples

In this section we present illustrative examples on the application of Algorithm 3.1.

4.1 Avoiding the greediness phenomenon in penalty methods for nonlinear programming

In this section, we present an application of a smoothing process to the solution of optimization problems of the form

$$\text{Min } \varphi(x) \text{ subject to } g(x) \leq 0, \tag{10}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable. Let $\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ be the feasible region and let $\varphi^{\text{ub}} \in \mathbb{R}$ be such that $\varphi(x) \leq \varphi^{\text{ub}}$ for all $x \in \Omega$, assumed to be known. It is easy to see that solving problem (10) is equivalent to solving the discontinuous unconstrained optimization problem given by

$$\text{Min } f(x),$$

where

$$f(x) = \begin{cases} \varphi(x), & \text{if } x \in \Omega, \\ \varphi^{\text{ub}} + \Phi(x), & \text{otherwise,} \end{cases}$$

and

$$\Phi(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\}^2.$$

Let $\{\omega_k\}_{k=1}^{\infty}$ such that $\omega_k \to 0$ when $k \to \infty$ and $\{\kappa_k\}_{k=1}^{\infty}$ such that $\kappa_k \to \infty$ when $k \to \infty$ be given sequences of positive numbers. We define

$$\Omega_k = \{x \in \mathbb{R}^n \mid g(x) \leq -\omega_k\} \subseteq \Omega \text{ and } \bar{\Omega}_k = \Omega \backslash \Omega_k,$$

and we also define $\varphi_k(x) : \Omega \to \mathbb{R}$ as

$$\varphi_k(x) = \begin{cases} \varphi(x), & \text{if } x \in \Omega_k, \\ \varphi(x)\,[1 - H_k(x)] + \left(\varphi^{\text{ub}} + \Phi(x)\right)[H_k(x)]\,, & \text{if } x \in \bar{\Omega}_k, \end{cases} \tag{11}$$

where

$$H_k(x) = \frac{\kappa_k \Phi_k(x)}{1 + \kappa_k \Phi_k(x)}$$

and

$$\Phi_k(x) = \sum_{i=1}^{m} \max\{0, g_i(x) + \omega_k\}^2.$$

Note that (11) is equivalent to:

$$\varphi_k(x) = \varphi(x)\,[1 - H_k(x)] + \left(\varphi^{\text{ub}} + \Phi(x)\right)[H_k(x)] \text{ for all } x \in \Omega.$$

We now define the sequence of continuously differentiable functions $f_k : \mathbb{R}^n \to \mathbb{R}$ given by

$$f_k(x) = \begin{cases} \varphi_k(x), & \text{if } x \in \Omega, \\ \varphi^{\text{ub}} + \Phi(x), & \text{otherwise.} \end{cases}$$

Clearly, $f_k(x) \geq f(x)$ for all k and all $x \in \mathbb{R}^n$, since, by definition, $\varphi_k(x) \geq \varphi(x)$ for all k and all $x \in \Omega$. Moreover, it is also easy to see that $\lim_{k\to\infty} f_k(x) = f(x)$ for all $x \in \mathbb{R}^n$. Thus, Assumptions A1, A2, and A4 are fulfilled by the sequence $\{f_k\}$.

Summing up, in the application being described, the original nonlinear programming problem (10) is modeled as an equivalent discontinuous unconstrained minimization problem, which is tackled by solving a sequence of smooth unconstrained problems of the form

$$\text{Min } f_k(x)$$

for $k = 1, 2, \ldots$. The smooth functions $f_k(x)$ were constructed in such a way that the original objective function $\varphi(x)$ in (10) plays no role whenever $x \notin \Omega$. This was done with the purpose of avoiding an inconvenience of penalty methods known as greediness, which is the tendency of being attracted by spurious infeasible points at which the objective function goes to minus infinity (see [4,8] for details).

Table 1 Results of applying Algorithm 3.1 to Example 4.1.1.

k	ε_k	ω_k	κ_k	# it	$\|x^k\|_2$	$\prod_{i=1}^{10} x_i^k$
1	10^{-4}	10^{-4}	10	0	5.6250000D−01	7.3735355D−08
2	10^{-5}	10^{-5}	10^2	0	5.6250000D−01	7.3735355D−08
3	10^{-6}	10^{-6}	10^3	215	9.9999903D−01	−9.9999512D−06
4	10^{-7}	10^{-7}	10^4	5	9.9999990D−01	−9.9999950D−06
5	10^{-8}	10^{-8}	10^5	5	9.9999999D−01	−9.9999994D−06

Example 4.1.1

$$\text{Min} \prod_{i=1}^{10} x_i \text{ subject to } 0.25 \le \|x\|_2^2 \le 1.$$

We consider a random initial guess $x^0 = 0.75\ \bar{x}^0/\|\bar{x}^0\|_2$, where $\bar{x}^0$ is such that its components $\bar{x}_i^0$ are random values with uniform distribution in the interval $[-1, 1]$. It is worth noting that when a classical optimization method based on penalization is applied to this problem, the sequence of iterates x^k diverges very quickly and $\prod_{i=1}^{10} x_i^k \to -\infty$ when $k \to \infty$. This is what happens when, for example, the Augmented Lagrangian method Algencan [2,7] is applied to this problem.

We now describe the application of Algorithm 3.1. Since, for each k, the subproblem of minimizing f_k at Step 1 of Algorithm 3.1 is an unconstrained smooth problem, the method proposed in [6], named Gencan, was used as the “Standard Algorithm” required to tackle the subproblems. We considered $\varepsilon_1 = 10^{-4}$. In the definition of f_k, we set $\omega_1 = 10^{-4}$ and $\kappa_1 = 10$ and, for $k > 1$, $\omega_k = \omega_{k-1}/10$ and $\kappa_k = 10\kappa_{k-1}$. Table 1 shows the results. For each $k \in \{1, \dots, 5\}$ the table displays the values of ω_k and κ_k that define f_k, the tolerance ε_k used to stop the process of minimizing f_k at a point x^k such that $\|\nabla f_k(x^k)\|_\infty \le \varepsilon_k$, the number of iterations ‘#it’ required to achieve this stopping criterion, the values of $\|x^k\|_2$ (from which feasibility can be observed), and the objective functional value $\prod_{i=1}^{10} x_i^k$. In the five calls, Gencan successfully solved the subproblems. In the first two subproblems, the initial guess satisfies the stopping criterion, meaning that $x^2 = x^1 = x^0$. The main work is done minimizing f_3 to obtain x^3 that already is a reasonable approximation to the solution. The last two optimizations (of f_4 and f_5) are very quick and simple increase the number of correct digits in the solution. At the end, we have, as expected, $x_i^5 \approx 1/\sqrt{10}$ for $i = 1, \dots, 10$.

4.2 Additional illustrative problems

In this section we show the performance of the introduced method when applied to the four problems considered in [18], where direct search methods for minimizing discontinuous functions were introduced. The four problems are two-dimensional bound-constrained problems of the form

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega \equiv [-1, 1]^2,$$

where function f is discontinuous and is given by one of the four functions defined below:
Example 4.2.1

$$f(x) = \begin{cases} f_1(x) \equiv x_1^2 + x_2^2, & \text{if } x \in \Omega_1 \equiv \{x \in \mathbb{R}^2 \mid x_1/2 \le x_2 \le 2x_1\}, \\ f_2(x) \equiv x_1^2 + x_2^2 + 10, & \text{if } x \in \Omega_2 \equiv \Omega \backslash \Omega_1. \end{cases}$$

Example 4.2.2

$$f(x) = \begin{cases} f_1(x) \equiv 10x_1^2 + x_2^2, & \text{if } x \in \Omega_1 \equiv \{x \in \mathbb{R}^2 \mid x_1 \ge 0\}, \\ f_2(x) \equiv 10x_1^2 + 10x_2^2, & \text{if } x \in \Omega_2 \equiv \Omega \backslash \Omega_1. \end{cases}$$

Example 4.2.3

$$f(x) = \begin{cases} f_1(x) \equiv x_1^2 + x_2^2, & \text{if } x \in \Omega_1 \equiv \{x \in \mathbb{R}^2 \mid x_2 = 2x_1\}, \\ f_2(x) \equiv x_1^2 + x_2^2 + 10, & \text{if } x \in \Omega_2 \equiv \Omega \backslash \Omega_1. \end{cases}$$

Example 4.2.4

$$f(x) = \begin{cases} f_1(x) \equiv x_1^2 + x_2^2, & \text{if } x \in \Omega_1 \equiv \{x \in \mathbb{R}^2 \mid x_1/2 \le x_2 \le 2x_1\}, \\ f_2(x) \equiv x_1^2 + x_2^2 + 5, & \text{if } x \in \Omega_2 \equiv \{x \in \mathbb{R}^2 \mid x_1 \le 0 \text{ and } x_2 \le 0\} \backslash \Omega_1, \\ f_3(x) \equiv x_1^2 + x_2^2 + 10, & \text{if } x \in \Omega_3 \equiv \{x \in \mathbb{R}^2 \mid x_2 \le x_1/2 \text{ and } x_1 \ge 0\} \backslash (\Omega_1 \cup \Omega_2), \\ f_4(x) \equiv x_1^2 + x_2^2 + 15, & \text{if } x \in \Omega_4 \equiv \Omega \backslash (\Omega_1 \cup \Omega_2 \cup \Omega_3). \end{cases}$$

These problems were tackled by solving a sequence of bound-constrained smooth problems given by

$$\text{Minimize } f_k(x) \text{ subject to } x \in \Omega \equiv [-1, 1]^2, \tag{12}$$

for $k = 1, 2, \ldots$. Note that, by definition, the feasible set Ω satisfies Assumption A3. In Examples 4.2.1–4.2.3, we have that

$$f_k(x) = \left(1 - H_k^1(x)\right) f_1(x) + H_k^1(x) f_2(x),$$

$$H_k^1(x) = \frac{\kappa_k \omega_1(x)}{1 + \kappa_k \omega_1(x)},$$

and $\omega_1(x)$ is the squared infeasibility measure associated with Ω_1. This means that if we assume that Ω_1 is defined as $\Omega_1 \equiv \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ and } g(x) \le 0\}$, where $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are smooth functions (as it is in fact the case in the considered examples) then $\omega_1(x)$ is given by $\|h(x)\|_2^2 + \|g(x)_+\|_2^2$. The sequence $\{\kappa_k\}$ is a sequence of positive numbers that goes to infinity when k goes to infinity. In the numerical experiments, we considered $\kappa_k = 10^k$. In Example 4.2.4, we define the squared infeasibility measures ω_i associated with the sets Ω_i and the corresponding H_k^i functions for $i = 1, 2, 3$. In this case, f_k is given by

$$f_k(x) = \left(1 - H_k^1(x)\right) f_1(x) + H_k^1(x) \times \left\{ \left(1 - H_k^2(x)\right) f_2(x) + H_k^2(x) \left[\left(1 - H_k^3(x)\right) f_3(x) + H_k^3(x) f_4(x) \right] \right\}.$$

In the four cases, the sequence $\{f_k\}$ satisfies Assumptions A1 and A4 although, in general, does not satisfy Assumption A2.

We now describe the application of Algorithm 3.1. This time, for each k, the subproblem that is solved at Step 1 of Algorithm 3.1 is the bound-constrained smooth problem (12). Once again, the subproblems were solved using the bound-constrained

Table 2 Results of applying Algorithm 3.1 to Examples 4.2.1–4.2.4.

k	ε_k	κ_k	#it	#fcnt	#gcnt	$f_k(x^k)$	x^k
Example 4.1.1							
1	10^{-4}	10	5	28	14	4.3e−24	(1.7e−12, 8.8e−13)
2	10^{-5}	10^2	0	1	2	8.1e−24	(1.7e−12, 8.8e−13)
3	10^{-6}	10^3	1	2	4	3.9e−24	(1.7e−12, 8.8e−13)
4	10^{-7}	10^4	0	1	2	3.9e−24	(1.7e−12, 8.8e−13)
5	10^{-8}	10^5	0	1	2	3.9e−24	(1.7e−12, 8.8e−13)
Example 4.1.2							
1	10^{-4}	10	1	2	6	1.7e−16	(4.2e−9, 7.5e−10)
2	10^{-5}	10^2	0	1	2	1.7e−16	(4.2e−9, 7.5e−10)
3	10^{-6}	10^3	0	1	2	1.7e−16	(4.2e−9, 7.5e−10)
4	10^{-7}	10^4	0	1	2	1.7e−16	(4.2e−9, 7.5e−10)
5	10^{-8}	10^5	1	2	4	4.6e−19	(−1.2e−12, 6.8e−10)
Example 4.1.3							
1	10^{-4}	10	3	25	10	1.3e−17	(1.6e−9, 3.2e−9)
2	10^{-5}	10^2	0	1	2	1.3e−17	(1.6e−9, 3.2e−9)
3	10^{-6}	10^3	1	2	4	1.3e−17	(1.6e−9, 3.2e−9)
4	10^{-7}	10^4	0	1	2	1.3e−17	(1.6e−9, 3.2e−9)
5	10^{-8}	10^5	0	1	2	1.3e−17	(1.6e−9, 3.2e−9)
Example 4.1.4							
1	10^{-4}	10	3	25	11	5.7e−14	(2.1e−7, 1.0e−7)
2	10^{-5}	10^2	0	1	2	5.7e−14	(2.1e−7, 1.0e−7)
3	10^{-6}	10^3	0	1	2	5.7e−14	(2.1e−7, 1.0e−7)
4	10^{-7}	10^4	1	2	4	3.3e−33	(5.9e−17, 2.5e−17)
5	10^{-8}	10^5	0	1	2	3.3e−33	(5.9e−17, 2.5e−17)

solver Gencan. We considered $\varepsilon_1 = 10^{-4}$ and $\kappa_1 = 10$ and, for $k > 1$, $\varepsilon_k = \varepsilon_{k-1}/10$ and $\kappa_k = 10\kappa_{k-1}$. Table 2 shows the results. For each $k \in \{1, \dots, 5\}$, the table displays the value of κ_k that defines f_k, the tolerance ε_k used to stop the process of minimizing f_k at a point x^k such that $\| P_\Omega(x^k - \nabla f_k(x^k)) - x^k \|_\infty \le \varepsilon_k$ (that, for convex-constrained minimization is a stopping criterion equivalent to (6)), the number of iterations, functional evaluations, and gradient evaluations (#it, #fcnt, and #gcnt, respectively) required to achieve this stopping criterion for the kth subproblem, the value of $f_k(x^k)$, and the point x^k itself. The initial point was always a random initial point $x^0 \in [-1, 1]^2$.

As it can be seen in Table 2, when Algorithm 3.1 is applied to Example 4.1.1, almost all the work is done when $k = 1$. The point x^1 is also a solution for the subproblem corresponding to $k = 2$ and this is the reason why nothing is done at the second iteration of the algorithm. Then, a single iteration of the bound-constrained solver Gencan is done when minimizing f_3 starting from $x_2 = x_1$ to obtain x_3 that is almost identical to the previously obtained points. The process continues with $x_3 = x_4 = x_5$. The performance is analogous in the other three examples and the method found a good approximation to the known solution $x^* = (0, 0)^T$ in all cases using a reduced number of functional evaluations. On the one hand, if these results are compared to the ones reported in [18], it can be concluded that Algorithm 3.1 is very effective and efficient (the method proposed in [18] fails, as supported by the underlying theory, when applied to Example 4.1.3 because Ω_1 has an empty interior). The method proposed in the present work also performs a very reduced number of functional evaluations when compared to the one introduced in [18]; although in the present context a single evaluation of f_k means evaluating all parts of the piecewise defined functions f and $\nabla f_k(\cdot)$ is also evaluated. On the other hand, building f_k requires to have access to the piecewise definition of f; while the method proposed in [18] may be applied to black-box discontinuous functions.

5 Conclusions

The theoretical results presented in this paper showed that, under very mild conditions, assuming global minimization of subproblems, global minimizers of the original discontinuous problem are obtained. Moreover, if we only assume that the subproblems solver finds asymptotically stationary points, the sequence generated by the algorithm converges to points that satisfy a sequential optimality condition. Obviously, it is not possible to claim universal robustness or efficiency of smoothing methods that obey the assumptions given here. However, illustrative numerical experiments showed that this approach can be useful to solve interesting problems. For example, applications to smoothed quantile regression [1] seems to be attractive and will deserve to be considered in the near future.

Acknowledgments The authors would like to thank the associate editor and the three referees for their comments and remarks that helped a lot to improve the first version of this work.

References

1. Abraham, Z., Xin, F., Tan, P.-N.: Smoothed quantile regression for statistical downscaling of extreme events in climate modeling. In: Srivastava, A.N., Chawla, N.V., Perera, A.S. (eds.) Proceedings of the 2011 Conference on Intelligent Data Understanding (CIDU 2011), pp. 92–106. Mountain View, CA (2011)
2. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On Augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**, 1286–1309 (2008)
3. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)
4. Birgin, E.G., Castelani, E.V., Martinez, A.L.M., Martínez, J.M.: Outer trust-region method for constrained optimization. J. Optim. Theory Appl. **150**, 142–155 (2011)
5. Birgin, E.G., Floudas, C.A., Martínez, J.M.: Global minimization using an Augmented Lagrangian method with variable lower-level constraints. Math. Program. **125**, 139–162 (2010)
6. Birgin, E.G., Martínez, J.M.: Large-scale active-set box-constrained optimization method with spectral projected gradients. Comput. Optim. Appl. **22**, 101–125 (2002)
7. Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM, Philadelphia (2014)
8. Castelani, E.V., Martinez, A.L., Martínez, J.M., Svaiter, B.F.: Addressing the greediness phenomenon in nonlinear programming by means of proximal Augmented Lagrangians. Comput. Optim. Appl. **46**, 229–245 (2010)
9. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71–99 (2012)
10. Heuts, R., de Klein, J.: An (s-q) inventory model with stochastic and interrelated lead times. Naval Res. Logist. **42**, 839–859 (1995)
11. Kamimura, K., Nishino, H.: An efficient method for determining economical configurations of elementary packet-switched networks. IEEE Trans. Cybern. **39**, 278–288 (1991)
12. Maia, L.O.A., Vidal de Carvalho, L.A., Qassim, R.Y.: Synthesis of utility systems by simulated annealing. Comput. Chem. Eng. **19**, 481–488 (1995)
13. Martínez, J.M.: Minimization of discontinuous cost functions by smoothing. Acta Appl. Math. **71**, 245–260 (2002)
14. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
15. Rozvany, G.: Optimal plastic design with discontinuous cost functions. J. Appl. Mech. **41**, 309–310 (1974)
16. Shulman, A., Vachani, R.: A decomposition algorithm for capacity expansion of local access networks. IEEE Trans. Cybern. **41**, 1063–1073 (1993)
17. Turkay, M., Grossmann, I.E.: Disjunctive programming techniques for the optimization of process systems with discontinuous investment costs-multiple size regions. Ind. Eng. Chem. Res. **35**, 2611–2623 (1996)
18. Vicente, L.N., Custódio, A.L.: Analysis of direct searches for discontinuous functions. Math. Program. **133**, 299–325 (2012)
19. Wechsung, A., Barton, P.I.: Global optimization of bounded factorable functions with discontinuities. J. Global Optim. **58**, 1–30 (2014)
20. Zang, I.: Discontinuous optimization by smoothing. Math. Oper. Res. **6**, 140–152 (1981)