$\mathcal{E}$-Enlargements of maximal monotone operators: theory and applications

Regina S. Burachik[*], Claudia A. Sagastizábal[†] and B. F. Svaiter[‡]

October 14, 2004

Abstract

Given a maximal monotone operator T, we consider a certain ε-enlargement T^ε, playing the role of the ε-subdifferential in nonsmooth optimization. We establish some theoretical properties of T^ε, including a transportation formula, its Lipschitz continuity, and a result generalizing Brønsted & Rockafellar's theorem. Then we make use of the ε-enlargement to define an algorithm for finding a zero of T.

Keywords: maximal monotone operators, enlargement of an operator, Brønsted & Rockafellar's theorem, transportation formula, algorithmic scheme.

AMS Subject Classification: Primary: 65K05; Secondary: 90C30, 52A41, 90C25.

[*]Engenharia de Sistemas de Computação, COPPE–UFRJ, CP 68511, Rio de Janeiro–RJ, 21945–970, Brazil.regi@cos.ufrj.br

[†]INRIA, BP 105, 78153 Le Chesnay, France. Also Depto. Engenharia Elétrica, PUC-Rio. Caixa Postal 38063, 22452-970 Rio de Janeiro, Brazil. Claudia.Sagastizabal@inria.fr or sagastiz@ele.puc-rio.br

[‡]IMPA, Estrada Dona Castorina, 110. Rio de Janeiro, RJ, CEP 22460-320, Brazil. benar@impa.br

1 Introduction and motivation

Given a convex function $f\colon \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$, the *subdifferential* of f at x, i.e. the set of *subgradients* of f at x, denoted by $\partial f(x)$, is defined as

$$\partial f(x) = \{u \in \mathbb{R}^N : f(y) - f(x) - \langle u, y - x\rangle \geq 0 \text{ for all } y \in \mathbb{R}^N\}\,.$$

This concept has been extended in [BR65], where the *ε-subdifferential* of f at x was defined for any $\varepsilon \geq 0$ as follows:

$$\partial_\varepsilon f(x) := \{u \in \mathbb{R}^N : f(y) - f(x) - \langle u, y - x\rangle \geq -\varepsilon \text{ for all } y \in \mathbb{R}^N\}\,.$$

The introduction of a "smearing" parameter ε gives an enlargement of $\partial f(x)$ with good continuity properties (see § 2.3 below for a definition of continuous multifunctions). Actually, the ε-subdifferential calculus appears as a powerful tool in connexion with nonsmooth optimization: the initial ideas presented by [BM73] and [Nur86] resulted in the so-called methods of ε-descent, [HUL93, Chapter XIII], which are the antecessors of bundle methods, [Kiw90], [SZ92], [LNN95], [BGLS97]. The convergence of such methods is based on some crucial continuity properties of the multifunction $x \hookrightarrow \partial_\varepsilon f(x)$, such as its upper and Lipschitz continuity and also a transportation formula. We refer to [HUL93, Chapter XI] for a deep study of all these classical properties of the ε-subdifferential.

In this paper we address some continuity issues of the ε-enlargement for maximal monotone operators T^ε, introduced in [BIS97] and defined as follows:

Definition 1.1 *Given $T\colon H \to \mathcal{P}(H)$, a maximal monotone operator on a Hilbert space H, and $\varepsilon \geq 0$, the* ε-enlargement *of T at x is defined by*

$$\begin{array}{llll} T^\varepsilon\colon & H & \to & \mathcal{P}(H) \\ & x & \hookrightarrow & \{u \in H : \langle v - u, y - x\rangle \geq -\varepsilon\,, \forall y \in H\,, v \in T(y)\}\,. \end{array} \qquad (1)$$

□

Observe that the definition above is similar to the definition of $\partial_\varepsilon f(\cdot)$. When $T = \partial f$, $\partial_\varepsilon f(x) \subseteq T^\varepsilon(x)$; this result, as well as examples showing that the inclusion can be strict, can be found in [BIS97].

First, we prove that the multifunction $x \hookrightarrow T^\varepsilon(x)$ shares with $x \hookrightarrow \partial_\varepsilon f(x)$ some properties that make ε-descent methods work, specifically, a theorem generalizing Brønsted & Rockafellar's, a transportation formula, and the Lipschitz-continuity.

Second, we develop an ε-descent-like method to find a zero of T using the ε-enlargements $T^\varepsilon(x)$. More precisely, consider the unconstrained convex program

$$\text{Find } x \text{ such that } 0 \in \partial f(x)\,.$$

It is well-known that $\partial f(\cdot)$ is a maximal monotone operator (see [Mor65]). In this sense, the problem above can be generalized to

$$\text{Find } x \text{ such that } 0 \in T(x)\,, \tag{2}$$

which is the classical problem of finding a zero of a maximal monotone operator T. We denote the solution set of (2) by $\mathcal{S}$.

The algorithm we introduce to solve (2) can be outlined as follows. Given an arbitrary y and $v \in T(y)$, because T is monotone, the set $\mathcal{S}$ is contained in the halfspace

$$H_{y,v} := \{z \in H : \langle z - y, v\rangle \leq 0\}\,. \tag{3}$$

Let x^k be a current (non optimal) iterate, i.e., $0 \notin T(x^k)$, and let $(y^k, v^k \in T(y^k))$ be such that $x^k \notin H_{y^k,v^k}$. Then project x^k onto H_{y^k,v^k} to obtain a new iterate:

$$x^{k+1} := \mathrm{P}_{H_{y^k,v^k}}(x^k) = x^k - \frac{\langle x^k - y^k, v^k\rangle}{\|v^k\|^2} v^k\,. \tag{4}$$

Because x^{k+1} is a projection, it is closer to the solution set than x^k. This simple idea exploiting projection properties has also been used successfully in other contexts, for example to solve variational inequality problems ([Kor76], [Ius94], [SS96], [IS97], [Kon97], [SS97]).

The use of extrapolation points y^k for computing the direction v^k comes back to [Kor76] and her *extragradient* method for solving continuous variational inequality problems. When solving (2), in order to have convergence and good performances, appropriate y^k and v^k should be found. For instance, since $0 \notin T(x^k)$, it may seem natural to simply find $s^k = \mathrm{P}_{T(x^k)}(0)$ and pick $y^k = x^k - t^k s^k$, for some positive stepsize t^k. Nevertheless, because $x \hookrightarrow T(x)$ is not continuous, (in the sense defined in §2.3) the method can converge to a point which is not a solution, as an example in [Luc97, page 4], shows. In addition, such a method may behave like a steepest-descent algorithm for nonsmooth minimization, whose numerical instabilities are well known, [HUL93, Chapter VII.2.2].

When T is the subdifferential of f, a better direction s^k can be obtained by projecting 0 onto a set bigger than $\partial f(x^k)$, namely the $\partial_\varepsilon f(x^k)$. Accordingly, when T in (2) is an arbitrary maximal monotone operator, it seems a good idea to parallel this behaviour when generating y^k in (4). Here is where the ε-enlargements $T^\varepsilon(x)$ come to help: the directions s^k will be generated by computing $\mathrm{P}_{T^\varepsilon(x^k)}(0)$, for appropriate values of ε.

The paper is organized as follows. In Section 2 we establish a generalization of the Brønsted & Rockafellar's theorem, a "transportation formula" and the Lipschitz-continuity for T^ε. Then, in § 3, we define an algorithm combining the projections and ε-descent techniques outlined above. This conceptual algorithm is proved to be convergent in § 3.2 and is the basis to develop the implementable bundle-like method described in the companion paper [BSS97]. We finish in § 4 with some concluding remarks.

2 Properties of T^ε

We gather in this section some results which are important not only for theoretical purposes, but also in view of designing the implementable algorithm of [BSS97]. These properties are related to the continuity of the ε-enlargement as a multifunction and also to "transportation formulæ" relating elements from T^ε with those from T.

We need first some notation. Given a set $A \subseteq H$ and a multifunction $S\colon H \to \mathcal{P}(H)$

– the closure of A is denoted by $\overline{A}$,

– we define the set $S(A) := \bigcup_{a\in A} S(a)\,.$

– The domain, image and graph of S are respectively denoted by

$$\begin{array}{rcl} D(S) & := & \{x \in H : S(x) \neq \emptyset\}\,, \\ R(S) & := & S(H) \text{ and} \\ G(S) & := & \{(x,v) \,:\, x \in D(S) \text{ and } v \in S(x)\}\,. \end{array}$$

– S is *locally bounded* at x if there exists a neighbourhood U of x such that the set $S(U)$ is bounded.

– S is *monotone* if $\langle u - v, x - y\rangle \geq 0$ for all $u \in S(x)$ and $v \in S(y)$, for all $x, y \in H$.

– S is *maximal monotone* if it is monotone and, additionally, its graph is not properly contained in the graph of any other monotone operator. □

Recall that any maximal monotone operator is locally bounded in the interior of its domain ([Roc69, Theorem 1]).

2.1 Extending Brønsted & Rockafellar's theorem

For a closed proper convex function f, the theorem of Brønsted & Rockafellar, see for instance [BR65], states that any ε-subgradient of f at a point x_ε can be approximated by some *exact* subgradient, computed at some x, possibly different from x_ε.

The ε-enlargement of Definition 1.1 also satisfies this property:

Theorem 2.1 *Let $T : H \to \mathcal{P}(H)$ be maximal monotone, $\varepsilon > 0$ and $(x_\varepsilon, v_\varepsilon) \in G(T^\varepsilon)$. Then for all $\eta > 0$ there exists $(x, v) \in G(T)$ such that*

$$\|v - v_\varepsilon\| \leq \frac{\varepsilon}{\eta} \quad \textit{and} \quad \|x - x_\varepsilon\| \leq \eta\,. \tag{5}$$

Proof. For an arbitrary positive coefficient β define the multifunction

$$\begin{array}{rccl} G_\beta\colon & H & \to & \mathcal{P}(H) \\ & y & \hookrightarrow & \beta T(y) + \{y\}\,. \end{array}$$

Since βT is maximal monotone, by Minty's theorem [Min62], G_β is a surjection:

$$\exists\,(x,v) \in G(T) \quad \text{such that} \quad G_\beta(x) = \beta v + x = \beta v_\varepsilon + x_\varepsilon\,.$$

This, together with Definition 1.1, yields

$$\begin{aligned} \langle v_\varepsilon - v, x_\varepsilon - x\rangle &= -\beta\|v - v_\varepsilon\|^2 \\ &= -\frac{1}{\beta}\|x - x_\varepsilon\|^2 \quad \geq -\varepsilon\,. \end{aligned}$$

Choosing $\beta := \eta^2/\varepsilon$, the result follows. □

Observe that the proof above only uses the ε-inequality characterizing elements in T^ε. Accordingly, the same result holds for any other enlargement of T, as long as it is contained in T^ε.

The result above can also be expressed in a set-formulation:

$$T^\varepsilon(x) \subset \bigcap_{\eta>0} \bigcup_{y\in B(x,\eta)} \{T(y) + B(0, \frac{\varepsilon}{\eta})\}\,, \tag{6}$$

where $B(x,\rho)$ denotes the unit ball centered in x with radius ρ. This formula makes clear that the value $\eta = \sqrt{\varepsilon}$ is a compromise between the distance to x and the degree of approximation. The value of η which makes such quantities equal gives the following expression

$$T^\varepsilon(x) \subset \bigcup_{y\in B(x,\sqrt{\varepsilon})} \{T(y) + B(0, \sqrt{\varepsilon})\}\,.$$

It follows that T^ε is locally bounded together with T. The following result, which extends slightly Proposition 2 in [BIS97], gives further relations between the two multifunctions.

Corollary 2.2 *With the notations above, the following hold:*

(i) $R(T) \subset R(T^\varepsilon) \subset \overline{R(T)}$,

(ii) $D(T) \subset D(T^\varepsilon) \subset \overline{D(T)}$,

(iii) If $d(\cdot,\cdot)$ *denotes the point-to-set distance, then* $d((x_\varepsilon, v_\varepsilon); G(T)) \leq \sqrt{2\varepsilon}$

Proof. The leftmost inclusions in (i) and (ii) are straightforward from Definition 1.1. As for the right ones, they follow from Theorem 2.1, making $\eta \to +\infty$ and $\eta \to 0$ in (i) and (ii) respectively.

To prove (iii), take $\eta = \sqrt{\varepsilon}$ in (5), write

$$d((x_\varepsilon, v_\varepsilon); G(T))^2 \le \|x - x_\varepsilon\|^2 + \|v - v_\varepsilon\|^2 \le 2\varepsilon\,,$$

and take square roots. □

2.2 Transportation Formula

We already mentioned that the set $T^\varepsilon(x)$ approximates $T(x)$, but this fact is of no use as long as there is no way of *computing* elements of $T^\varepsilon(x)$. The question is then how to construct an element in $T^\varepsilon(x)$ with the help of some elements $(x^i, v^i) \in G(T)$. The answer is given by the "transportation formula" stated below. Therein we use the notation

$$\Delta_m := \{\alpha \in \mathbb{R}^m \mid \alpha_i \ge 0\,, \sum_{i=1}^m \alpha_i = 1\}$$

for the unit-simplex in $\mathbb{R}^m$.

Theorem 2.3 *Let T be a maximal monotone operator defined in a Hilbert space H. Consider a set of m triplets*

$$\{(\varepsilon_i, x^i, v^i \in T^{\varepsilon_i}(x^i))\}_{i=1,\dots,m}\,.$$

For any $\alpha \in \Delta_m$ define

$$\begin{array}{rcl} \hat{x} &:=& \sum_1^m \alpha_i x^i \\ \hat{v} &:=& \sum_1^m \alpha_i v^i \\ \hat{\varepsilon} &:=& \sum_1^m \alpha_i \varepsilon_i + \sum_1^m \alpha_i \langle x^i - \hat{x}, v^i - \hat{v}\rangle\,. \end{array} \tag{7}$$

Then $\hat{\varepsilon} \ge 0$ and $\hat{v} \in T^{\hat{\varepsilon}}(\hat{x})$.

Proof. Recalling Definition 1.1, we need to show that the inequality in (1) holds for any $(y, v) \in G(T)$. Combine (7) and (1), with (ε, x, u) replaced by $(\varepsilon_i, x^i, v^i)$, to obtain

$$\begin{array}{rcl} \langle \hat{x} - y, \hat{v} - v\rangle &=& \sum_1^m \alpha_i \langle x^i - y, \hat{v} - v\rangle \\ &=& \sum_1^m \alpha_i [\langle x^i - y, \hat{v} - v^i\rangle + \langle x^i - y, v^i - v\rangle] \\ &\ge& \sum_1^m \alpha_i \langle x^i - y, \hat{v} - v^i\rangle - \sum_1^m \alpha_i \varepsilon_i\,. \end{array} \tag{8}$$

Since

$$\begin{array}{rcl} \sum_1^m \alpha_i \langle x^i - y, \hat{v} - v^i\rangle &=& \sum_1^m \alpha_i [\langle x^i - \hat{x}, \hat{v} - v^i\rangle + \langle \hat{x} - y, \hat{v} - v^i\rangle] \\ &=& -\sum_1^m \alpha_i \langle x^i - \hat{x}, v^i - \hat{v}\rangle + 0 \\ &=& -\sum_1^m \alpha_i \langle x^i - \hat{x}, v^i - \hat{v}\rangle\,, \end{array}$$

with (8) and (7) we get

$$\langle \hat{x} - y, \hat{v} - v \rangle \geq -\hat{\varepsilon}\,. \tag{9}$$

For contradiction purposes, suppose that $\hat{\varepsilon} < 0$. Then $\langle \hat{x} - y, \hat{v} - v \rangle > 0$ for any $(y, v) \in G(T)$ and the maximality of T implies that $(\hat{x}, \hat{v}) \in G(T)$. In particular, the pair $(y, v) = (\hat{x}, \hat{v})$ yields $0 > 0$!. Therefore $\hat{\varepsilon}$ must be nonnegative. Since (9) holds for any $(y, v) \in G(T)$, we conclude from (1) that $\hat{v} \in T^{\hat{\varepsilon}}(\hat{x})$. □

Observe that when $\varepsilon_i = 0$, for all $i = 1, \ldots, m$, this theorem shows how to construct $\hat{v} \in T^{\hat{\varepsilon}}(\hat{x})$, using $(x^i, v^i) \in G(T)$.

The formula above holds also when replacing T^ε by $\partial_\varepsilon f$, with f a proper closed and convex function. This is Proposition 1.2.10 in [Lem80], where an equivalent expression is given for $\hat{\varepsilon}$:

$$\hat{\varepsilon} = \sum_{i=1}^{m} \alpha_i \varepsilon_i + \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j \langle x^i - x^j, v^i - v^j \rangle\,.$$

Observe also that, when compared to the standard transportation formula for ε-subdifferentials, Theorem 2.3 is a *weak* transportation formula, in the sense that it only allows to express some *selected* ε-subgradients in terms of subgradients.

The transportation formula can also be used for the ε-subdifferential to obtain the lower bound:

$$\text{If} \quad \{(v^i \in \partial_{\varepsilon_i} f(x^i))\}_{i=1,2} \quad \text{then} \quad \langle x^1 - x^2, v^1 - v^2 \rangle \geq -(\varepsilon_1 + \varepsilon_2)\,.$$

In the general case, we have a weaker bound:

Corollary 2.4 *Take $v^1 \in T^{\varepsilon_1}(x^1)$ and $v^2 \in T^{\varepsilon_2}(x^2)$. Then*

$$\langle x^1 - x^2, v^1 - v^2 \rangle \geq -(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2 \tag{10}$$

Proof. If ε_1 or ε_2 are zero the result holds trivially. Otherwise choose $\alpha \in \Delta_2$ as follows

$$\alpha_1 := \frac{\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \qquad \alpha_2 := 1 - \alpha_1 = \frac{\sqrt{\varepsilon_1}}{\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2}} \tag{11}$$

and define the convex sums $\hat{x}$, $\hat{v}$ and $\hat{\varepsilon}$ as in (7). Because $\hat{\varepsilon} \geq 0$, we can write

$$\begin{aligned} 0 \leq \hat{\varepsilon} &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \langle x^1 - \hat{x}, v^1 - \hat{v} \rangle + \alpha_2 \langle x^2 - \hat{x}, v^2 - \hat{v} \rangle \\ &= \alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2 + \alpha_1 \alpha_2 \langle x^2 - x^1, v^2 - v^1 \rangle\,, \end{aligned}$$

where we have used the identities $x^1 - \hat{x} = \alpha_2(x^1 - x^2)$, $x^2 - \hat{x} = \alpha_1(x^2 - x^1)$, $v^1 - \hat{v} = \alpha_2(v^1 - v^2)$ and $v^2 - \hat{v} = \alpha_1(v^2 - v^1)$ first, and then $\alpha_1 \alpha_2^2 + \alpha_1^2 \alpha_2 = \alpha_1 \alpha_2$. Now, combine the expression above with (7) and (11) to obtain

$$\sqrt{\varepsilon_1 \varepsilon_2} + \frac{\sqrt{\varepsilon_1 \varepsilon_2}}{(\sqrt{\varepsilon_1} + \sqrt{\varepsilon_2})^2} \langle x^1 - x^2, v^1 - v^2 \rangle \geq 0$$

Rearranging terms and simplifying the resulting expression, (10) is proved. □

2.3 Lipschitz-Continuity of T^ε

All along this subsection, $H = \mathbb{R}^N$ and $D := D(T)^0$ is the interior of $D(T)$. The reason for taking $H = \mathbb{R}^N$ is that for the results below we need bounded sets to be compact.

A closed-valued locally bounded multifunction S is *continuous* at $\bar{x}$ if for any positive ϵ there exist $\delta > 0$ such that

$$\|x - \bar{x}\| \leq \delta \quad \Longrightarrow \quad \begin{cases} S(x) & \subset & S(\bar{x}) + B(0,\epsilon) \\ S(\bar{x}) & \subset & S(x) + B(0,\epsilon) \end{cases} .$$

Furthermore, S is *Lipschitz continuous* if, given a (nonempty) compact set $K \subseteq D$, there exists a nonnegative constant L such that for any $y^1, y^2 \in K$ and $s^1 \in S(y^1)$ there exists $s^2 \in S(y^2)$ satisfying $\|s^1 - s^2\| \leq L\|y^1 - y^2\|$.

We will prove that the application $(\varepsilon, x) \hookrightarrow T^\varepsilon(x)$ is Lipschitz continuous in D. We start with a technical lemma.

Lemma 2.5 *Assume that D is nonempty. Let $K \subset D$ be a compact set and take $\rho > 0$ such that*

$$\tilde{K} := K + \overline{B(0,\rho)} \subset D \,. \tag{12}$$

Define $\tilde{M} := \sup\{\|u\| \mid u \in T(\tilde{K})\}$. Then, for all $\varepsilon \geq 0$, we have that

$$\sup\{\|u\| : u \in T^\varepsilon(K)\} \leq \frac{\varepsilon}{\rho} + \tilde{M} \,. \tag{13}$$

Proof. Because T is locally bounded and $\tilde{K}$ is compact, $T(\tilde{K})$ is bounded. Then $\tilde{M}$ is finite. To prove (13), take $x_\varepsilon \in K$ and $v_\varepsilon \in T^\varepsilon(x_\varepsilon)$. Apply Theorem 2.1 with $\eta := \rho$: there exists a pair $(x, v) \in G(T)$ such that

$$\|x - x_\varepsilon\| \leq \rho \qquad \text{and} \qquad \|v - v_\varepsilon\| \leq \frac{\varepsilon}{\rho} \,.$$

Then $x \in \tilde{K}$, $\|v\| \leq \tilde{M}$, and therefore

$$\|v_\varepsilon\| \leq \|v_\varepsilon - v\| + \|v\| \leq \frac{\varepsilon}{\rho} + \tilde{M} \,,$$

so that (13) follows. □

Now we prove the Lipschitz-continuity of T^ε. Our result strengthens Theorem 1(ii) in [BIS97].

Theorem 2.6 *Assume that D is nonempty. Let $K \subset D$ be a compact set and $0 < \underline{\varepsilon} \le \bar{\varepsilon} < +\infty$. Then there exist nonnegative constants A and B such that for any (ε_1, x^1), $(\varepsilon_2, x^2) \in [\underline{\varepsilon}, \bar{\varepsilon}] \times K$ and $v^1 \in T^{\varepsilon_1}(x^1)$, there exists $v^2 \in T^{\varepsilon_2}(x^2)$ satisfying*

$$\|v^1 - v^2\| \le A\|x^1 - x^2\| + B|\varepsilon_1 - \varepsilon_2| \,. \tag{14}$$

Proof. With ρ, $\tilde{K}$ and $\tilde{M}$ as in Lemma 2.5, we claim that (14) holds for the following choice of A and B:

$$A \;:=\; \left(\frac{1}{\rho} + \frac{2\tilde{M}}{\underline{\varepsilon}}\right)\left(\frac{\bar{\varepsilon}}{\rho} + 2\tilde{M}\right), \qquad B \;:=\; \left(\frac{1}{\rho} + \frac{2\tilde{M}}{\underline{\varepsilon}}\right). \tag{15}$$

To see this, take x^1, x^2, ε_1, ε_2 and v^1 as above. Take $l := \|x^1 - x^2\|$ and let x^3 be in the line containing x^1 and x^2 such that

$$\|x^3 - x^2\| \;=\; \rho \quad , \qquad \|x^3 - x^1\| \;=\; \rho + l \,, \tag{16}$$

as shown in Figure 1.

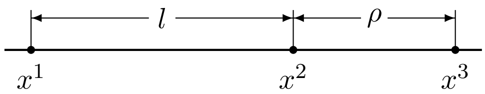

Figure 1:

Then, $x^3 \in \tilde{K}$ and

$$x^2 = (1-\theta)x^1 + \theta x^3 \quad , \text{ with } \quad \theta = \frac{l}{\rho + l} \in [0, 1)\,.$$

Now, take $u^3 \in T(x^3)$ and define

$$\tilde{v}^2 := (1-\theta)v^1 + \theta u^3 \,.$$

Theorem 2.3 yields $\tilde{v}^2 \in T^{\tilde{\varepsilon}_2}(x^2)$, with

$$\begin{aligned}\tilde{\varepsilon}_2 &= (1-\theta)\varepsilon_1 + (1-\theta)\langle x^1 - x^2, v^1 - \tilde{v}^2\rangle + \theta\,\langle x^3 - x^2, u^3 - \tilde{v}^2\rangle \\ &= (1-\theta)\varepsilon_1 + \theta(1-\theta)\,\langle x^1 - x^3, v^1 - u^3\rangle \,.\end{aligned}$$

Use Lemma 2.5 with $v^1 \in T^{\varepsilon_1}(x^1)$, together with the definition of $\tilde{M}$, to obtain

$$\|v^1 - u^3\| \le \|v^1\| + \|u^3\| \le \tilde{M} + (\frac{\varepsilon_1}{\rho} + \tilde{M}) \le \frac{\varepsilon_1}{\rho} + 2\tilde{M}\,. \tag{17}$$

Using now Cauchy-Schwarz, (17), (16), and recalling the definition of θ, we get

$$\begin{aligned} \tilde{\varepsilon}_2 &\leq (1-\theta)\varepsilon_1 + \theta(1-\theta)\|x^1 - x^3\|\,\|v^1 - u^3\| \\ &\leq (1-\theta)\varepsilon_1 + \theta(1-\theta)(\rho + l)\left(\frac{\varepsilon_1}{\rho} + 2\tilde{M}\right) \\ &= \varepsilon_1 + \frac{\rho l}{\rho + l} 2\tilde{M}. \end{aligned} \tag{18}$$

The definition of $\tilde{v}^2$ combined with (17) yields

$$\|v^1 - \tilde{v}^2\| = \theta\|v^1 - u^3\| \leq \theta\left(\frac{\varepsilon_1}{\rho} + 2\tilde{M}\right), \tag{19}$$

as well as

$$\|v^1 - \tilde{v}^2\| \leq \|x^1 - x^2\|\frac{1}{\rho}\left(\frac{\varepsilon_1}{\rho} + 2\tilde{M}\right). \tag{20}$$

Now consider two cases:

(i) $\tilde{\varepsilon}_2 \leq \varepsilon_2$,

(ii) $\tilde{\varepsilon}_2 > \varepsilon_2$.

If (i) holds, $\tilde{v}^2 \in T^{\tilde{\varepsilon}_2}(x^2) \subseteq T^{\varepsilon_2}(x^2)$. Then, choosing $v^2 := \tilde{v}^2$ and using (20) together with (15), (14) follows.

In case (ii), define $\beta := \dfrac{\varepsilon_2}{\tilde{\varepsilon}_2} < 1$ and $v^2 := (1-\beta)u^2 + \beta\tilde{v}^2$, with $u^2 \in T(x^2)$. Because of Theorem 2.3, $v^2 \in T^{\varepsilon_2}(x^2)$. Furthermore, (13) together with (19) lead to

$$\begin{aligned} \|v^2 - v^1\| &\leq (1-\beta)\|u^2 - v^1\| + \beta\|\tilde{v}^2 - v^1\| \\ &\leq (1-\beta)\left(\tfrac{\varepsilon_1}{\rho} + 2\tilde{M}\right) + \beta\theta\left(\tfrac{\varepsilon_1}{\rho} + 2\tilde{M}\right) \\ &= (1-\beta(1-\theta))\left(\tfrac{\varepsilon_1}{\rho} + 2\tilde{M}\right). \end{aligned} \tag{21}$$

Using (18) we have that $\beta \geq \dfrac{\varepsilon_2}{\varepsilon_1 + \frac{\rho l}{\rho + l}2\tilde{M}}$.

Some elementary algebra, the inequality above and the definitions of θ and l, yield

$$\begin{aligned} 1 - \beta(1-\theta) &\leq l\left(\frac{\varepsilon_1 + \rho 2\tilde{M}}{(\rho + l)\varepsilon_1 + \rho l 2\tilde{M}}\right) + \frac{\rho(\varepsilon_1 - \varepsilon_2)}{(\rho + l)\varepsilon_1 + \rho l 2\tilde{M}} \\ &\leq \|x^1 - x^2\|\left(\frac{1}{\rho} + \frac{2\tilde{M}}{\varepsilon_1}\right) + |\varepsilon_1 - \varepsilon_2|\,\frac{1}{\varepsilon_1}. \end{aligned} \tag{22}$$

Altogether, with (21), (22) and our assumptions on $\varepsilon_1, \varepsilon_2, \underline{\varepsilon}, \overline{\varepsilon}$, the conclusion follows. □

The continuity of $T^\varepsilon(x)$ as a multifunction is straightforward. In particular, it also has a closed graph:

For any sequence $\{(\varepsilon_i, x^i, v^i \in T^{\varepsilon_i}(x^i))\}_i$ such that $\varepsilon_i > 0$ for all i,

$$\lim_{i\to\infty} x^i = x, \quad \lim_{i\to\infty} \varepsilon_i = \varepsilon, \quad \lim_{i\to\infty} v^i = v \implies v \in T^{\varepsilon}(x)\,. \tag{23}$$

Let us mention that this result was already proved in [BIS97, Proposition 1(iv)].

3 Conceptual algorithmic patterns

The way is now open to define iterative algorithms for solving (2). We have at hand an enlargement $T^{\varepsilon}(x)$, which is continuous on x and ε. We present in this section a conceptual algorithm, working directly with the convex sets $T^{\varepsilon}(x)$, together with its convergence proof. As already mentioned, this algorithm will be the basis of the implementable algorithm in [BSS97], where the sets $T^{\varepsilon}(x)$ are replaced by suitable polyhedral approximation by means of a bundle strategy which makes an extensive use of the transportation formula in Theorem 2.3.

The ε-enlargement T^{ε} was also used in [BIS97] to formulate and analyze a conceptual algorithm for solving monotone variational inequalities. The method is of the proximal point type, with Bregman distances and inexact resolution of the related subproblems.

We assume from now on that $D(T) = \mathbb{R}^N$, so that T maps bounded sets in bounded sets, and we have a nonempty solution set $\mathcal{S}$.

3.1 The algorithm

Before entering into technicalities, we give an informal description on how the algorithm works. Following our remarks in § 1, iterates will be generated by projecting onto halfspaces $H_{y,v}$ from (3). We also said that to ensure convergence, the pair $(y, v \in T(y))$ has to be "good enough", in a sense to be precised below. Since $y = x^k - ts$, where $s = \mathrm{P}_{T^{\varepsilon_k}(x^k)}(0)$ for some $\varepsilon_k > 0$, this comes to say that ε_k has to be good enough. The control on $\varepsilon_k = \varepsilon 2^{-j_k}$ is achieved in Step 1 below. More precisely, Lemma 3.6 and Theorem 3.7 in§ 3.2 show how the decreasing sequence $\{\varepsilon_k\}$ converges to 0.

Now we describe the algorithm.

Conceptual Algorithmic Scheme (CAS):

Choose positive parameters τ, R, ε and σ, with $\sigma \in (0,1)$.

INITIALIZATION: Set $k := 0$ and take $x^0 \in \mathbb{R}^N$.

K-STEP:

Step 0 **(stopping test)**

(0.a) If $0 \in T(x^k)$, then STOP.

Step 1 **(computing search direction)**

(1.a) Set $j := 0$.

(1.b) Compute $s^{k,j} := \operatorname{argmin}\{\|v\|^2 \mid v \in T^{\varepsilon 2^{-j}}(x^k)\}$.

(1.c) If $\|s^{k,j}\| \le \tau\, 2^{-j}$ then set $j := j+1$ and LOOP to (1.b).

(1.d) Else, define $j_k := j$ and $s^k := s^{k,j_k}$.

Step 2 **(line search)**

(2.a) Set $l := 0$.

(2.b) Define $y^{k,l} := x^k - R\,2^{-l} s^k$ and take $v^{k,l} \in T(y^{k,l})$.

(2.c) If $\langle v^{k,l}, s^k\rangle \le \sigma \|s^k\|^2$, then set $l := l+1$ and LOOP to (2.b).

(2.d) Else, define $l_k := l$, $v^k := v^{k,l_k}$ and $y^k := y^{k,l_k}$.

Step 3 **(projection step)**

(3.a) Define $x^{k+1} := x^k - \langle v^k, x^k - y^k\rangle v^k / \|v^k\|^2$.

(3.b) Set $k := k+1$ and LOOP to k-step. □

A few comments are now in order.

In *Step 1* a search direction s^k is computed. It satisfies $s^k \in T^{\varepsilon 2^{-j_k}}(x^k)$ with $\|s^{k,j_k}\| > \tau\, 2^{-j_k}$. We show in Proposition 3.3(i) that an infinite loop cannot occur in this step.

In *Step 2*, using the direction s^k, a pair $(y^k, v^k \in T(y^k))$ is obtained, by iterating on l. This pair is such that not only $x^k \notin H_{y^k,v^k}$, but it is also "far enough" and gives a nonnegligible progress on $\|x^{k+1} - x^k\|$. This is shown in Proposition 3.4.

Variational inequality problems (VIP) can be considered as constrained versions of (2). Accordingly, a "constrained" variant of (CAS) could be devised in a straightforward manner by adding a subalgorithm ensuring feasibility of iterates. Rather than introducing this extra complication, we chose to focus our efforts in designing an algorithmic pattern oriented to future implementations. As a result, in [BSS97] we analyze a "Bundling Strategy", whose K-STEP gives an *implementable* version of the one in (CAS). As already mentioned, for this mechanism to work transportation formulæ like the one in Theorem 2.3 are crucial, because they allow the iterative construction of polyhedral approximations of $T^{\varepsilon 2^{-j_k}}(x^k)$ while preserving convergence. Further on, in § 3.2, we analyze the convergence of (CAS). One of the key arguments for the proof in Lemma 3.6 below is the Lipschitz-continuity of T^ε, stated in Theorem 2.6. In turn, the proof of the latter result is based on the (useful!) transportation formula.

Along these lines, the related recent work [Luc97] introduces two algorithms to solve VIP with multivalued maximal monotone operators. These methods, named *Algorithms I and II*, ensure feasibility of iterates by making inexact orthogonal projections. For the unconstrained case, both algorithms first choose a direction in some enlargement $T^{\varepsilon_k}(x^k)$, like in (1.b) of (CAS), and perform afterwards a line search similar to *Step 2* in (CAS). The important difference between (CAS) and the methods in [Luc97] is how the parameter ε_k is taken along iterations:

– *Algorithm I* uses a constant $\varepsilon_k = \varepsilon$ for all k. As a result, see [Luc97, Theorem 3.4.1], $\{x^k\}$ is a bounded sequence whose cluster points $\bar{x}$ are ε-solutions of (2), i.e., points satisfying $0 \in T^{\varepsilon}(\bar{x})$.

– *Algorithm II* makes use of a dynamical ε_k, varying along iterations. In this case, if the sequence of iterates $\{x^k\}$ is infinite, it converges to a solution of (2). If finite termination occurs, the last generated point x^{k_f} is shown to be an ε_{k_f}-solution ([Luc97, Theorem 3.6.1]).

In (CAS), instead, a close and careful control on ε_k is made during the iterative process. This control is decisive to prove convergence to a solution, even for the finite termination case (see Theorem 3.7 below).

Summing up, we believe that both *Algorithm II* in [Luc97] and (CAS) are very important steps for developing implementable schemes for solving VIP, because they are on the road to extending bundle methods to a more general framework than nonsmooth convex optimization. However, when compared to the methods in [Luc97], we also think that (CAS) makes a more effective use of the enlargements $T^{\varepsilon_k}(x^k)$ and gives a deeper insight of the mechanism that makes algorithms of this type work, both in theory and in practice.

3.2 Convergence Analysis

For proving convergence of (CAS), we use the concept of Fejér-convergent or Fejér-monotone sequence:

Definition 3.1 A sequence $\{x^k\}$ is said to be *Fejér-monotone* with respect to a set $\mathcal{C}$ if $\|x^{k+1} - x\| \leq \|x^k - x\|\,,\quad$ for any $x \in \mathcal{C}$. □

The following elementary result, that we state here without proof, will be used in the sequel.

Proposition 3.2 *Let the sequence $\{x^k\}$ be Fejér-monotone with respect to a non-empty set $\mathcal{C}$. Then*

(i) $\{x^k\}$ *is bounded.*

(ii) If $\{x^k\}$ *has an accumulation point which is in* $\mathcal{C}$*, then the full sequence converges to a limit in* $\mathcal{C}$*.*

□

Note that a Fejér-monotone sequence *is not necessarily* convergent: a sequence $\{x^k\}$ could approach a set $\mathcal{C}$ *without* converging to an element of $\mathcal{C}$. For our (CAS), convergence will follow from Proposition 3.2 and adequate choices of y^k and v^k in (4).

In (CAS) each k-step has two ending points: (0.a) and (3.b). If some k-step exits at (0.a), then the algorithm stops. If the k-step exits at (3.b), then x^{k+1} is generated and a new k-step is started with k replaced by $k+1$.

There are also two loops: (1.c)↔(1.b) on j indices; and the other (2.c)↔(2.b), iterating along l indices. We start proving that infinite loops do not occur on these indices.

Proposition 3.3 *Let x^k be the current iterate in (CAS). Then, either x^k is a solution in* (2) *and (CAS) stops in (0.a); or x^k is not a solution and the following holds:*

(i) Relative to the loop (1.c)↔(1.b), there exists a finite $j = j_k$ such that (1.d) is reached:

$$\|s^k\| > \tau 2^{-j_k} \quad \textit{with} \quad \|s^{k,j_k-1}\| \leq \tau 2^{-j_k+1} \,,$$

whenever $j_k > 0$.

(ii) Relative to the loop (2.c)↔(2.b), there exists a finite $l = l_k$ such that (2.d) is reached:

$$\langle v^{k,l_k}, s^k \rangle > \sigma \|s^k\|^2 \quad \textit{with} \quad \langle v^{k,l_k-1}, s^k \rangle \leq \sigma \|s^k\|^2 \,,$$

whenever $l_k > 0$.

Proof. If x^k is a solution, then $0 \in T(x^k)$ and (CAS) ends at (0.a). Otherwise, we have that $0 \notin T(x^k)$. To prove (i), suppose, for contradiction, that (CAS) loops forever in (1.c)↔(1.b). Then it generates an infinite sequence $\{s^{k,j}\}_{j\in\mathbf{N}}$ such that

$$s^{k,j} \in T^{\varepsilon 2^{-j}}(x^k) \quad \text{and} \quad \|s^{k,j}\| \leq \tau 2^{-j} \,.$$

Letting j go to infinity and using (23), this implies that $0 \in T(x^k)$, a contradiction. Hence, there exists a finite j such that the loop (1.c)↔(1.b) ends. For this index j, Step (1.d) is reached, $j_k = j$ is defined and $\|s^{k,j_k}\| > \tau 2^{-j_k}$, so that (i) holds.

Now we prove (ii). If an infinite loop occurs at (2.c)↔(2.b), then an infinite sequence $\{(y^{k,l}, v^{k,l})\}_{l\in\mathbf{N}} \subseteq G(T)$ is generated. This sequence is such that

$$\lim_{l\to\infty} y^{k,l} = \lim_{l\to\infty} x^k - R\,2^{-l} s^k = x^k \quad \text{and} \quad \langle v^{k,l}, s^k \rangle \leq \sigma \|s^k\|^2 \,,$$

for all $l \in \mathbf{N}$. Because $\{y^{k,l}\}_{l\in\mathbf{N}}$ is bounded (it is a convergent sequence), so is $\{v^{k,l}\}_{l\in\mathbf{N}}$, by the local boundedness of T. Extracting a subsequence if needed, there exists a $\bar{v}$ such that

$$\lim_{i\to\infty} v^{k,l_i} = \bar{v} \in T(x^k) \subseteq T^{\varepsilon 2^{-j_k}}(x^k)\,,$$

with $\langle \bar{v}, s^k\rangle \le \sigma\|s^k\|^2$. However, since s^k is the element of minimum norm in $T^{\varepsilon 2^{-j_k}}(x^k)$, we also have that $\langle \bar{v}, s^k\rangle \ge \|s^k\|^2$. Therefore, $\sigma \ge 1$, a contradiction.

Hence, there exists a finite l such that the loop (2.c)$\leftrightarrow$(2.b) ends. For this index l, Step (2.d) is reached, $l_k = l$ is defined and $\langle v^{k,l}, s^k\rangle > \sigma\|s^k\|^2$; altogether, (ii) is proved. □

We proved that the sequence generated by (CAS) is either finite, with last point which is a solution, or it is infinite, with no iterate solving (2). We are now in a position to show how (CAS) follows the scheme discussed in the introduction.

Proposition 3.4 *Let x^k be the current iterate in (CAS). Then, if x^k is not a solution in* (2)*, after x^{k+1} is generated in (3.b), the following holds:*

1. *Consider the halfspace defined in* (3)*. Then $x^k \notin H_{y^k,v^k}$ and $x^{k+1} = P_{H_{y^k,v^k}}(x^k)$.*
2. *For all $x^* \in \mathcal{S}$, $\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2$.*
3. *Finally, $\|x^{k+1} - x^k\| > R\sigma\tau^2 2^{-l_k-2j_k}/\|v^k\|$.*

Proof. To prove (i), use Proposition 3.3(ii) and (i): at step (3.b) the pair $(y^k = x^k - R2^{-l_k}s^k, v^k = v^{k,l_k})$ is such that

$$\langle x^k - y^k, v^k\rangle \ge R2^{-l_k}\sigma\|s^k\|^2 > R\sigma\tau^2 2^{-l_k-2j_k} > 0\,,$$

so that $x^k \notin H_{y^k,v^k}$. To see that x^{k+1} is its projection, just recall (4) and the definition of x^{k+1} in (CAS). Because x^{k+1} is an orthogonal projection, (ii) follows. As for (iii), it is also straightforward. □

We can now prove that (CAS) generates a Fejér-monotone sequence, with all the involved variables being bounded.

Lemma 3.5 *If (CAS) generates infinite x^k, then the sequence is Fejér-monotone with respect to the solution set $\mathcal{S}$.*

Moreover, x^k, s^k, y^k, v^k, as well as the sets $\{s^{k,j}\}_{0\le j\le j_k}$ and $\{(y^{k,l}, v^{k,l})\}_{0\le l\le l_k}$, are bounded.

Proof. It follows from Definition 3.1 and Proposition 3.4(ii) that the sequence is Fejér-monotone. Because of Proposition 3.2(i), there exists some compact set K_0 such that the (bounded) sequence $\{x^k\} \subset K_0$. Since T^ε is locally bounded,

$M_0 := \sup\{\|u\| : u \in T^\varepsilon(K_0)\}$ is finite and $s^{k,j} \in T^{\varepsilon 2^{-j}}(x^k) \subseteq T^\varepsilon(x^k)$ is bounded for all $j \le j_k$:

$$\|s^{k,j}\| \le M_0 \text{ for all } 0 \le j \le j_k \text{ and } \|s^k\| \le M_0\,,$$

for all k. Now, define $K_1 := K_0 + \overline{B(0, RM_0)}$ to obtain

$$y^{k,l} \in K_1 \text{ for all } 0 \le l \le l_k \text{ and } y^k \in K_1\,,$$

for all k. Using once more the local-boundedness of T, we conclude that $\{v^{k,l}\}_{0\le l\le l_k}$ is bounded, and the proof is finished. □

Before proving our convergence result, we need a last technical lemma.

Lemma 3.6 *Suppose that (CAS) generates an infinite sequence $\{x^k\}$. If the sequence $\{j_k\}_{k\in N}$ is bounded, then $\{l_k\}_{k\in N}$ is also bounded.*

Proof. Let J be an upper bound for the sequence $\{j_k\}$ and let $\varepsilon_k := \varepsilon 2^{-j_k}$, then

$$\varepsilon_k \in [\underline{\varepsilon}, \bar{\varepsilon}] := [\varepsilon 2^{-J}, \varepsilon]\,. \tag{24}$$

Now, for those k such that $l_k > 0$, Proposition 3.3(ii) applies. In addition, the pairs $(y^{k,l_k-1}, v^{k,l_k-1})$ satisfy

$$\|y^{k,l_k-1} - x^k\| = R2^{-l_k+1}\|s^k\| \text{ and } v^{k,l_k-1} \in T(y^{k,l_k-1}) \subseteq T^{\varepsilon_k}(y^{k,l_k-1})\,, \tag{25}$$

and. Using Lemma 3.5 and (24), we have that $x^1 := y^{k,l_k-1}$, $x^2 := x^k$ and $\varepsilon^1 = \varepsilon^2 := \varepsilon_k$ are bounded and Theorem 2.6 applies. Then, together with (25), we obtain,

$$\|v^{k,l_k-1} - v^2\| \le AR2^{-l_k+1}\|s^k\|\,, \tag{26}$$

for a positive A and some $v^2 \in T^{\varepsilon_k}(x^k)$, for each k such that $l_k > 0$,.

Consider the projection $\bar{v}^k := \mathrm{P}_{T^{\varepsilon_k}(x^k)}(v^{k,l_k-1})$. We have

$$\|s^k\| \le \|\bar{v}^k\|\,, \tag{27}$$

as well as

$$\|v^{k,l_k-1} - \bar{v}^k\| \le \|v^{k,l_k-1} - w\|\,, \tag{28}$$

for any $w \in T^{\varepsilon_k}(x^k)$; in particular, for $w = v^2$ from (26). Altogether, with (28) and (27),

$$\|v^{k,l_k-1} - \bar{v}^k\| \le AR2^{-l_k+1}\|\bar{v}^k\|\,, \tag{29}$$

for all k such that $l_k > 0$.

Consider now the scalar product $\langle v^{k,l_k-1}, s^k\rangle$ and apply succesively Cauchy-Schwartz, (29) and (27) again to write the following chain of (in)equalities:

$$\begin{array}{rcl}\langle v^{k,l_k-1}, s^k\rangle & = & \langle v^{k,l_k-1} - \bar{v}^k, s^k\rangle + \langle \bar{v}^k, s^k\rangle \\ & \geq & -AR2^{-l_k+1}\|s^k\|^2 + \|s^k\|^2 \\ & = & \left(1 - AR2^{-l_k+1}\right)\|s^k\|^2 .\end{array}$$

Because Proposition 3.3(ii) holds, we obtain

$$\sigma \geq \left(1 - AR2^{-l_k+1}\right) ,$$

which in turn yields

$$l_k \leq \max\left\{0, \log_2 \frac{2AR}{1-\sigma}\right\}$$

and the conclusion follows. □

Finally, we state our main convergence result.

Theorem 3.7 *Consider the sequence* $\{x^k\}$ *generated by (CAS). Then the sequence is either finite with last element solving* (2)*, or it converges to a solution in* (2)*.*

Proof. We already dealt with the finite case in Proposition 3.3. If (CAS) does not stop, then there is no x^k solving (2). Because of Proposition 3.4(ii), we have that $\sum \|x^{k+1} - x^k\|^2 < +\infty$ and therefore

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0 . \tag{30}$$

On the other side, Proposition 3.4(iii) leads to

$$\|x^{k+1} - x^k\| > \frac{R\sigma\tau^2}{M} 2^{-l_k-2j_k} ,$$

where M is an upper bound for $\|v^k\|$ given by Lemma 3.5. Together with (30), we obtain

$$\lim_{k\to\infty} l_k + 2j_k = +\infty .$$

Then, by Lemma 3.6, the sequence $\{j_k\}$ must be unbounded.

Because $\{x^k\}$ is bounded (this is again Lemma 3.5), we can extract a convergent subsequence as follows

$$\begin{array}{rcl}\lim_{q\to\infty} x^{k_q} & = & \hat{x} , \\ \lim_{q\to\infty} j_{k_q} & = & +\infty , \\ j_{k_q} > 0 & \text{for all} & q .\end{array} \tag{31}$$

For this subsequence, using Proposition 3.3(i), it holds that

$$\lim_{q\to\infty} s^{k_q, j_{k_q}-1} = 0\,.$$

Since $s^{k_q,j_{k_q}-1} \in T^{\varepsilon 2^{-j_{k_q}+1}}(x^{k_q})$, together with (31) and (23) we get

$$(\varepsilon 2^{-j_{k_q}+1}, x^{k_q}, s^{k_q,j_{k_q}-1}) \longrightarrow (0,\hat{x},0) \quad \Longrightarrow \quad 0 \in T(\hat{x})\,,$$

so that $\hat{x} \in \mathcal{S}$.

Therefore the Fejér-monotone sequence $\{x^k\}$ has an accumulation point which is in $\mathcal{S}$. By Proposition 3.2, the proof is done. □

4 Concluding Remarks

We presented a conceptual algorithm for finding zeros of a maximal monotone operator based on projections onto suitable halfspaces. Our algorithmic scheme extends the pattern of ε-descent methods for nonsmooth optimization by generating directions of minimum norm in $T^{\varepsilon_k}(x^k)$, with $\varepsilon_k = \varepsilon 2^{-j_k} \downarrow 0$. To develop implementable algorithms, a constructive device to approximate the convex set $T^{\varepsilon}(x)$ is needed. This is done in [BSS97] by extending the bundle machinery of nonsmooth optimization to this more general context. An important consideration to reduce iterations in the implementable version concerns backtracking steps in *Step 1.* The following modification, suggested by one of the referees, avoids unnecessary iterations on j-indices:

Take $\varepsilon_0 = \varepsilon$.

Step 1' **(computing search direction)**

(1.a') Set $j := 0$.
(1.b') Compute $s^{k,j} := \operatorname{argmin}\{\|v\|^2 \,|\, v \in T^{\varepsilon_k 2^{-j}}(x^k)\}\,.$
(1.c') If $\|s^{k,j}\| \le \tau\, 2^{-j}$ then set $j := j+1$ and LOOP to (1.b).
(1.d') Else, define $j_k := j$, $\varepsilon_{k+1} := \varepsilon_k 2^{-j_k}$ and $s^k := s^{k,j_k}$.

Finally, observe that (CAS) makes use of the extension $T^{\varepsilon}(x)$ from Definition 1.1. However, the same algorithmic pattern can be applied for any extension $E(\varepsilon,x)$ of the maximal monotone $x \hookrightarrow T(x)$, provided that it is *continuous* and a *transportation formula* like in Theorem 2.3 hold. In particular, any continuous $E(\varepsilon,x) \subset T^{\varepsilon}(x)$ for all $x \in D(T)$ and $\varepsilon \ge 0$ can be used.

Acknowledgments We are indebted to Igor Konnov, whose interesting remarks were very helpful during our revision of this paper. We also want to thank the referees for their careful and exhaustive reading.

References

[BGLS97] F. Bonnans, J.Ch. Gilbert, C.L. Lemaréchal and C.A. Sagastizábal. *Optimisation Numérique, aspects théoriques et pratiques.* Collection "Mathématiques et applications", SMAI-Springer-Verlag, Berlin, 1997.

[BM73] D.P. Bertsekas and S.K. Mitter. A descent numerical method for optimization problems with nondifferentiable cost functionals. *SIAM Journal on Control*, 11(4):637–652, 1973.

[BR65] A. Brøndsted and R.T. Rockafellar. On the subdifferentiability of convex functions. *Proceedings of the American Mathematical Society*, 16:605–611, 1965.

[BIS97] R.S. Burachik, A.N. Iusem, and B.F. Svaiter. Enlargements of maximal monotone operators with application to variational inequalities. *Set Valued Analysis*, 5:159–180, 1997.

[BSS97] R.S. Burachik, C.A. Sagastizábal, and B. F. Svaiter. Bundle methods for maximal monotone operators. Submitted, 1997.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms.* Number 305-306 in Grund. der math. Wiss. Springer-Verlag, 1993. (two volumes).

[Ius94] A. N. Iusem. An iterative algorithm for the variational inequality problem. *Computational and Applied Mathematics*, 13:103–114, 1994.

[IS97] A. N. Iusem. and B.F. Svaiter. A variant of Korpolevich's method for variational inequalities with a new search strategy. *Optimization*, 42:309–321, 1997.

[Kiw90] K.C. Kiwiel. Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming*, 46:105–122, 1990.

[Kon97] I.V. Konnov. A combined relaxation method for variational inequalities with nonlinear constraints. *Mathematical Programming*, 1997. Accepted for publication.

[Kor76] G.M. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematischeskie Metody*, 12:747–756, 1976.

[Lem80] C. Lemaréchal. Extensions diverses des méthodes de gradient et applications, 1980. Thèse d'Etat, Université de Paris IX.

[LNN95] C. Lemaréchal, A. Nemirovskii, and Yu. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–148, 1995.

[Luc97] L.R. Lucambio Pérez. Iterative Algorithms for Nonsmooth Variational Inequalities, 1997. Ph.D. Thesis, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil.

[Min62] G.L. Minty. Monotone nonlinear operators in a Hilbert space. *Duke Mathematical Journal*, 29:341–346, 1962.

[Mor65] J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

[Nur86] E.A. Nurminski. ε-subgradient mapping and the problem of convex optimization. *Cybernetics*, 21(6):796–800, 1986.

[Roc69] R.T. Rockafellar. Local boundedness of nonlinear monotone operators. *Michigan Mathematical Journal*, 16:397–407, 1969.

[SS96] M.V. Solodov and B.F. Svaiter. A new projection method for variational inequality problems. Technical Report B-109, IMPA, Brazil, 1996. *SIAM Journal on Control and Optimization*, submitted.

[SS97] M.V. Solodov and B.F. Svaiter. A hybrid projection-proximal point algorithm. Technical Report B-115, IMPA, Brazil, 1997.

[SZ92] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization*, 2(1):121–152, 1992.