Bundle methods for maximal monotone operators

Regina S. BURACHIK * Claudia SAGASTIZÁBAL † B. F. SVAITER ‡

Abstract

To find a zero of a maximal monotone operator T we use an enlargement T^{ε} playing the role of the ε-subdifferential in nonsmooth optimization. We define a convergent and implementable algorithm which combines projection ideas with bundle-like techniques and a transportation formula. More precisely, first we separate the current iterate x^k from the zeros of T by computing the direction of minimum norm in a polyhedral approximation of $T^{\varepsilon_k}(x^k)$. Then suitable elements defining such polyhedral approximations are selected following a bundle strategy. Finally, the next iterate is computed by projecting x^k onto the corresponding separating hyperplane.

Keywords: bundle methods, maximal monotone operators, enlargements of maximal monotone operators, transportation formulæ.

AMS subject classification: Primary: 65K05; Secondary: 90C30, 52A41, 90C25.

1 Introduction and motivation

Consider the problem

$$0 \in T(x)\,, \tag{1}$$

where T is a maximal monotone operator on $\mathbb{R}^N$ with nonempty solution set $\mathcal{S}$. When T is single-valued, well-known algorithms for solving (1) are Korpelevich's method [10], Khobotov's [7] and others.

In this paper we present an implementable algorithm for solving (1) when T is a multi-valued mapping. Our departing point is the following, simple, remark: given an arbitrary y and $v \in T(y)$, the monotonicity of T implies that $\mathcal{S}$ is contained in the halfspace

$$H_{y,v} := \{z \in \mathbb{R}^N : \langle z - y, v\rangle \leq 0\}\,. \tag{2}$$

*Engenharia de Sistemas de Computação, COPPE–UFRJ, CP 68511, Rio de Janeiro, RJ, CEP 21945–970, Brazil.

†INRIA, BP 105, 78153 Le Chesnay, France and PUC, Depto. Engenharia Elétrica. Rio de Janeiro, RJ, CEP 22453-900, Brazil. Research supported by FAPERJ (Brazil) under Grant No.E26/150.205/98.

‡IMPA, Estrada Dona Castorina, 110. Rio de Janeiro, RJ, CEP 22460-320, Brazil. Research partially supported by CNPq Grant 301200/93-9(RN) and by PRONEX–Optimization.

We have already presented in [3] a conceptual method for finding zeros of T developing these ideas. More precisely, given a current iterate $x^k \notin \mathcal{S}$,
– first, find y^k and $v^k \in T(y^k)$ such that $x^k \notin H_{y^k,v^k}$,
– then, project x^k onto $H_{y^k,v^k} \supseteq \mathcal{S}$ to obtain a new iterate

$$x^{k+1} := Proj_{H_{y^k,v^k}}(x^k) = x^k - \frac{\langle x^k - y^k, v^k \rangle}{\|v^k\|^2} v^k \,. \tag{3}$$

Being a projection, x^{k+1} is closer to $\mathcal{S}$ than x^k. However, to have a significant progress from x^k to x^{k+1}, adequate choices of (y^k, v^k) are required. Since $v^k \in T(y^k)$ is given by an oracle, the control can only be done when generating y^k. In our proposal we analyzed points of the form

$$y^k = x^k - t^k s^k \,, \quad \text{with} \quad s^k = Proj_{T^{\varepsilon_k}(x^k)}(0) \,, \tag{4}$$

for positive t^k and ε_k. The sets $T^\varepsilon(x)$ are the ε-enlargements of T from [2], further studied in [3]. These enlargements behave in many ways as the ε-subdifferential of a closed convex function, see § 2 below. In particular, they are also Lipschitz-continuous multi-functions of x and ε, provided ε is positive. The scheme (3)-(4) is useful for theoretical purposes. Actually, it allows to clearly identify the crucial elements to obtain convergence: (3) and (4) have to be combined in order to drive ε_k to 0, to generate a convergent sequence. At the same time, ε^k should not be driven to 0 too fast, otherwise the resulting multifunction $x \hookrightarrow T^{\varepsilon^k}$ would not be smooth enough and the improvement when passing from x^k to x^{k+1} would become negligible. Such a "loss of smoothness" is reflected by the sensitivity of s^k to variations in ε_k and x^k. More precisely, when $\varepsilon_k = 0$, s^k ceases to depend continuously on x^k. In this respect, we say that the problem of finding s^k in (4) becomes *ill-posed*.
When coming to implementation concerns, it appears that s^k in (4) cannot be computed without having a *full* knowledge of $T^{\varepsilon_k}(x^k)$, a fairly bold (if not impossible) assumption. Instead, we assume that an oracle, giving one element in $T(z)$ for any z, is available. Then s^k can be approached by projecting 0 onto a polyhedral approximation of $T^{\varepsilon_k}(x^k)$. A suitable polyhedral approximation is obtained by using the transportation formula for T^ε, proved in [3], together with bundle techniques like in [8], [18], [13]: having at the current iteration a raw bundle with all the oracle information collected so far, $\{(z^i, w^i \in T(z^i))\}_{i \le p}$, the convex hull of some *selected* w^i's is a good approximation of $T^{\varepsilon_k}(x^k)$. Again, special attention has to be put when selecting the sub-bundle, in order to control ε_k and preserve convergence.
The paper is organized as follows. In Section 2 we start with some notations and assumptions. Then we recall the definition of T^ε, together with some continuity properties and the transportation formula. Section 3 describes an implementable algorithm combining the projections ideas and bundle techniques outlined above. This algorithm is proved to be convergent in § 4. Finally, in § 5

we give some concluding remarks and we compare our algorithm with a closely related work by Konnov, [9].

2 Preliminary results

We gather in this section notation and general assumptions, as well as some results, straightforward or already proved in [2] and [3].

2.1 Notation and assumptions

Let $\mathcal{P}(\mathbb{R}^N)$ denote the set of all subsets of $\mathbb{R}^N$. Then, given a multifunction $F\colon \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ and a set $A \subseteq \mathbb{R}^N$:

– the closure of A is denoted by $\overline{A}$ and
– the domain of F is $D(F) := \{x \in \mathbb{R}^N \,:\, F(x) \neq \emptyset\}$.
– We define the set $F(A) := \bigcup_{a\in A} F(a)$.
– $B(x,\rho)$ denotes the ball centered in x with radius ρ.
– F is *locally bounded* at x if there exists a neighbourhood U of x such that the set $F(U)$ is bounded.
– F is *monotone* if $\langle u - v, x - y\rangle \geq 0$ for all $x, y \in \mathbb{R}^N$, and all $u \in F(x)$, $v \in F(y)$.
– T is *maximal monotone* if it is monotone and, additionally, whenever there is some monotone F such that $T(x) \subset F(x)$ for all $x \in \mathbb{R}^N$, this implies $T = F$. □

Recall that any maximal monotone operator is locally bounded in the interior of its domain ([17, Theorem 1]).
All along this paper, we suppose that T in (1) is defined in the whole of $\mathbb{R}^N$, so that it maps bounded sets in bounded sets. Finally, recall that the solution set $\mathcal{S}$ is assumed to be not empty.

2.2 The enlargement T^ε. Some useful properties

Given $T\colon \mathbb{R}^N \to \mathcal{P}(\mathbb{R}^N)$ maximal monotone and $\varepsilon \geq 0$, the ε-enlargement of T at x, introduced in [2] and thoroughly studied in [3], is defined by

$$\begin{array}{rlll} T^\varepsilon\colon & \mathbb{R}^N & \to & \mathcal{P}(\mathbb{R}^N) \\ & x & \hookrightarrow & \{u \in \mathbb{R}^N : \langle v - u, y - x\rangle \geq -\varepsilon \text{ for all } y \in \mathbb{R}^N,\ v \in T(y)\}. \end{array} \tag{5}$$

The idea of using a smearing parameter ε to obtain better regularity properties can be traced back to [1] and [4] for the subdifferential of convex and non convex functions, respectively. For monotone operators, not necessarily maximal, relaxed monotonicity was explored in [20]. Other related works are [11], [19], [15] and [16].

2.2.1 Continuity Properties

The ε-enlargement is proved to be locally bounded in [3]. Also, in [2] it was proved that the application $(x,\varepsilon) \hookrightarrow T^\varepsilon(x)$ is continuous (upper and lower semicontinuous). This result was improved in [3, Lemma 2.5], where T^ε is shown to be Lipschitz-continuous whenever $\varepsilon > 0$ (in [3] we also show that the Lipschitz constant diverges when $\varepsilon \to 0$).
In the sequel, we will just make use of the following result, establishing the closedness of the graph of T^ε:

Proposition 2.1 *[2, Propositions 1(iv) and 2(i)]* $T^0 = T$*, and for any sequence* $\{(\varepsilon_i, x^i, w^i \in T^{\varepsilon_i}(x^i))\}_i$ *such that* $\varepsilon_i \geq 0$ *for all* i*,*

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

□

2.2.2 Transportation Formula

In our algorithm, we build polyhedral approximations of $T^\varepsilon(x)$ using elements in a certain bundle of information $\{(z^i, w^i \in T(z^i))\}_{i=1}^m$. To prove convergence, it is crucial to relate elements in the bundle with some $v \in T^\varepsilon(x)$. This is the transportation formula for T^ε proved in [3, Theorem 2.3].

Theorem 2.2 *Consider a set of* m *triplets* $\{(\varepsilon_i, z^i, w^i \in T^{\varepsilon_i}(z^i))\}_{i=1,\dots,m}$*, and take the convex sum:*

$$(\hat{x}, \hat{s}) := (\sum_{i=1}^m \alpha_i z^i, \sum_{i=1}^m \alpha_i w^i)\,,$$

where $\alpha \in \mathbb{R}^m$ *is such that* $\alpha_i \geq 0$ *and* $\sum_{i=1}^m \alpha_i = 1$*. Then* $\hat{s} \in T^{\hat{\varepsilon}}(\hat{x})$*, with*

$$\hat{\varepsilon} := \sum_{i=1}^m \left[\alpha_i \varepsilon_i + \alpha_i \langle z^i - \hat{x}, w^i - \hat{s}\rangle\right].$$

□

Note that when compared to the result obtained for $\partial_\varepsilon f$ (see, for example, [5, Proposition XI.4.2.2.]), our formula is weaker, because it only transports convex sums.
In this paper, we make use of the transportation formula above only when $\varepsilon_i = 0$, that is, when $w^i \in T(z^i)$ is given by an oracle for all z^i.
As a consequence of Theorem 2.2, $\hat{\varepsilon}$ can be uniformly bounded on α. This result will be very useful in the sequel.

Corollary 2.3 *Consider the notations and assumptions of Theorem 2.2. Suppose that* $\varepsilon_i = 0$ *for all* $i \leq m$*. In addition, let* $\tilde{x} \in \mathbb{R}^N$ *and* $\rho > 0$ *be such that*

$\|z^i - \tilde{x}\| \leq \rho$ *for all* $i \leq m$*. Then, the convex sum* $(\hat{x}, \hat{s}) := (\sum_1^m \alpha_i z^i, \sum_1^m \alpha_i w^i)$ *satisfies the following:*

$$\|\hat{x} - \tilde{x}\| \leq \rho\,,$$
$$\hat{s} \in T^{\hat{\varepsilon}}(\hat{x}) \quad \textit{with} \quad \hat{\varepsilon} := \textstyle\sum_1^m \alpha_i \langle z^i - \hat{x}, w^i - \hat{s}\rangle \leq 2\rho M\,,$$

where $M := \max\{\|w^i\| \mid i = 1, \ldots, m\}$*.*

Proof. The convexity of the norm implies that $\|\hat{x} - \tilde{x}\| \leq \rho$, and also $\|\hat{s}\| \leq M$. Theorem 2.2 establishes that $\hat{s} \in T^{\hat{\varepsilon}}(\hat{x})$, for $\hat{\varepsilon} = \sum_1^m \alpha_i \langle z^i - \hat{x}, w^i - \hat{s}\rangle$. The last expression can be rewritten as follows:

$$\begin{aligned} \textstyle\sum_1^m \alpha_i \langle z^i - \hat{x}, w^i - \hat{s}\rangle &= \textstyle\sum_1^m \alpha_i \langle z^i - \tilde{x}, w^i - \hat{s}\rangle + \sum_1^m \alpha_i \langle \tilde{x} - \hat{x}, w^i - \hat{s}\rangle \\ &= \textstyle\sum_1^m \alpha_i \langle z^i - \tilde{x}, w^i - \hat{s}\rangle + 0 \\ &= \textstyle\sum_1^m \alpha_i \langle z^i - \tilde{x}, w^i - \hat{s}\rangle\,. \end{aligned}$$

Because $\|\hat{s}\| \leq M$, together with the Cauchy-Schwarz inequality and our assumptions, the proof is done. □

3 Building bundle-like methods

We present first the conceptual algorithm scheme from [3], working directly with the convex sets $T^\varepsilon(x)$. Further on, we use a bundle technique to build adequate polyhedral approximations of the sets $T^\varepsilon(x)$.

3.1 The conceptual algorithm scheme (CAS)

The scheme from [3] has the following form.

INITIALIZATION:Choose parameters $\tau\,, R\,, \varepsilon > 0$ and $\sigma \in (0, 1)$. Set $k := 0$ and take $x^0 \in \mathbb{R}^N$.

k-STEP:
Step 0: If $0 \in T(x^k)$, then STOP. **(stopping test)**
Step 1: **(computing search direction)**
Compute $s^k := \text{argmin}\{\|v\|^2 \mid v \in T^{\varepsilon 2^{-j}}(x^k)\}$, where j is such that $\|s^k\| > \tau\, 2^{-j}$.
Step 2: **(line search)**
Define $y^k := x^k - R\, 2^{-l} s^k$ and take $v^k \in T(y^k)$, for l such that $\langle v^k, s^k\rangle > \sigma\|s^k\|^2)$.
Step 3: **(projection step)**
Define $x^{k+1} := x^k - \langle v^k, x^k - y^k\rangle v^k / \|v^k\|^2$.
Set $k := k + 1$ and LOOP to **k-**STEP**.** □

Indices j and l are nonnegative integers; in [3] we prove that (CAS) is well defined, with no infinite loops in Steps 1 or 2. (CAS) is also convergent in the following sense: it either stops with a last $x^k \in \mathcal{S}$, or it generates an infinite sequence $\{x^k\}$ converging to a solution of (1). We will see that the same convergence result holds for an implementable version of (CAS) (cf. Theorem 4.8).

3.2 The implementable bundle strategy (BS)

We already mentioned that a major drawback of (CAS) is the expensive, when not impossible, computation of an element s^k of minimum norm in $T^\varepsilon(x^k)$. A proposal to get round this difficulty, inspired from our guiding model $T = \partial f$, is to apply a bundle-like strategy. However, since no functional values are available when $T \neq \partial f$, we need to extend and adapt dual-bundle methods (see [14], [6], also Chapters XI, XIII in [5]) to this more general context.

Let us describe how a bundle strategy can be used in our setting. As usual in bundle methods, we suppose that an *oracle* computes $v = v(y) \in T(y)$ for any given y. In addition, we denote by

$$\Delta_I := \{\lambda \in \mathbb{R}^I_+ : \sum_{i \in I} \lambda_i = 1\}$$

the unit-simplex associated to a set of indices I.

The following mechanism makes the "k-step" in (CAS) implementable. The convergence is then preserved by an appropriate choice of the current bundle in step (1.b) below.

Bundling Strategy (BS)

INITIALIZATION:Choose parameters $\tau\, , R > 0$ and $\sigma \in (0, 1)$. Set $k := 0$, $p := 0$ and take $x^0 \in \mathbb{R}^N$.

IMPLEMENTABLE **k-**STEP**:**

Step 0:

(0.a) Compute $u^k \in T(x^k)$, if $u^k = 0$, then STOP. **(stopping test)**

(0.b) Else, set $p := p + 1$, $(z^p, w^p) := (x^k, u^k)$. Set $n := 0$.

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

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

(1.b) Define $I_{k,n,j} := \{1 \le i \le p \,|\, \|z^i - x^k\| \le R\, 2^{-j}\}$.

(1.c) Compute $\alpha^{k,n,j} := \operatorname{argmin}\{\| \sum_{i \in I_{k,n,j}} \alpha_i w^i\|^2 \,|\, \alpha \in \Delta_{I_{k,n,j}}\}$.

(1.d) Take $s^{k,n,j} := \sum_{i \in I_{k,n,j}} \alpha_i^{k,n,j} w^i$.

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

(1.f) Else, define $j_{k,n} := j$ and $s^{k,n} := s^{k,n,j_{k,n}}$.

Step 2: **(line search)**

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

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

(2.c) If $\left(\langle v^{k,n,l}, s^{k,n} \rangle \le \sigma \|s^{k,n}\|^2 \text{ and } l < j_{k,n} + 1 \right)$, then

Set $l := l + 1$ and LOOP to (2.b).

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

Step 3: **(evaluating the pair (y, v))**

(3.a) If $\langle v^{k,n}, s^{k,n}\rangle \le \sigma\|s^{k,n}\|^2$, then **(null step)**

Set $p := p + 1$, $(z^p, w^p) := (y^{k,n}, v^{k,n})$.

Set $n := n + 1$ and LOOP to (1.b).

(3.b) Else **(serious step)**

Define $n_k := n$, $j_k := j_{k,n_k}$ and $l_k := l_{k,n_k}$,
$s^k := s^{k,n_k}$, $y^k := y^{k,n_k}$, $v^k := v^{k,n_k}$.

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

Set $k := k + 1$ and LOOP to IMPLEMENTABLE **k-**STEP.

□

We comment here some features of this implementable version of (CAS).

Remark 3.1

- Observe that index p is a counter of elements in the raw bundle $\{(z^i, w^i)\}_{i\le p}$. Namely, serious points are added at (0.b), while the information obtained in null steps is incorporated to the bundle in (3.a).
- In *Step 0* a stopping test of the kind $\|u^k\| \le$ TOL could have been used in (0.a). In such a case, the convergence analysis would essentially remain the same by setting the tolerance TOL to 0 in the results (to make the algorithm loop forever and produce an infinite sequence $\{x^k\}$).
- In *Step 1* the search direction is computed. The set $T^{\varepsilon 2^{-j}}(x^k)$ used in (CAS) is now replaced by the convex hull generated by some *selected* elements in the range of T, belonging to a reduced bundle. Observe that, by definition of the index set $I_{k,n,j}$, the vector $u^k \in T(x^k)$ is always in the associated (reduced) bundle. Hence this sub-bundle is not empty and $\alpha^{k,n,j}$ is well defined. We show in Lemma 4.2(i) that if x^k is not a solution, then the loop (1.e)↔(1.b) is always finite.
- In *Step 2* we evaluate the pair $(y^{k,n,l}, v^{k,n,l})$ provided by the direction $s^{k,n}$. Step (2.c) makes a loop (to (2.b)) which eventually ends. The two possible endings are:

  1. *null step:* For $l = j + 1$ it holds that $\langle v^{k,n}, s^{k,n}\rangle \le \sigma\|s^{k,n}\|^2$.
  2. *serious step:* When for some $l \le j+1$ the direction $v^{k,n}$ satisfies $\langle v^{k,n}, s^{k,n}\rangle > \sigma\|s^{k,n}\|^2$.

  In the first case, which leads to step (3.a), the line search performed in step (2.c) gives a pair $(y^{k,n}, v^{k,n})$ such that $\|y^{k,n} - x^k\| = R\,2^{-j-1}$. Then $v^{k,n}$ is added to the current sub-bundle (recall that v^i is incorporated only when $\|y^i - x^k\| \le R\,2^{-j}$), and no progress is made in k. Lemma 4.2(ii) shows that, when x^k is not a solution, there is no infinite loop in (3.a)↔(1.b).
- If a serious step is done, (3.b) is visited. In this case, the direction defined by $v^{k,n}$ defines a separating hyperplane far enough from the current iterate x^k. By "far enough" we mean that the stepsize $\|x^k - x^{k+1}\|$ does not go to zero too fast. A new iterate x^{k+1} is generated by projecting x^k onto the halfspace

H_{y^k,v^k} from (2). In addition, note that at step (3.b) the following holds:

$$\begin{array}{lcl} l_k \leq j_k + 1\,, & & \langle v^k, s^k \rangle > \sigma \|s^k\|^2\,, \\ \|s^k\| > \tau 2^{-j_k}\,, & \text{and} & y^k = x^k - R2^{-l_k} s^k / \|s^k\|\,. \end{array} \tag{6}$$

□

4 Convergence Analysis

We show in this section that either (BS) generates a finite sequence, whose last iterate solves (1), or it generates an infinite sequence, converging to a solution of (1).
The bundle strategy (BS) provides us with a constructive device for (CAS). For this reason the convergence analysis is close to the one in [3]. The main difference appears when analyzing null steps, which lead to an enrichment of the bundle. For proving that the convergence is preserved when replacing $T^{\varepsilon}(x^k)$ by the current polyhedral approximation, we use the following technical result:

Lemma 4.1 *[5, Lemma IX.2.1.1] Let $\gamma > 0$ fixed. Consider two infinite sequences sequences $\{v^m\}$ and $\{\hat{v}^m\}$ satisfying for $m = 1, 2, \ldots$:*

$$\langle v^i - v^{m+1}, \hat{v}^m \rangle \geq \gamma \|\hat{v}^m\|^2 \quad \text{for all } i = 1, \ldots, m\,. \tag{7}$$

If $\{v^i\}$ is bounded then $\hat{v}^m \to 0$ when $m \to \infty$.

□

Each k-step of (BS) has two ending points: (0.a) and (3.b). If some implementable k-step ends at (0.a), then the algorithm stops at a solution of (1). If the implementable k-step ends at (3.b), then x^{k+1} is generated and a new k-step is started with k replaced by $k+1$.
There are three inner loops: (1.e)↔(1.b) on j indices; (3.a)↔(1.b), iterating along n indices and finally (2.c)↔(2.b), incrementing l.
We prove first that infinite loops do not occur on these indices when $x^k \notin \mathcal{S}$.

Lemma 4.2 *Let x^k be the current iterate in (BS) and suppose $x^k \notin \mathcal{S}$. Then the following holds:*

(i) Relative to the loop (1.e)↔(1.b), there exists a finite $j = j_{k,n}$ such that (1.f) is reached:

$$\|s^{k,n}\| > \tau 2^{-j_{k,n}}\,.$$

Furthermore, the loop (2.c)↔(2.b) is finite: (2.d) is reached with $l_{k,n} \leq j_{k,n} + 1$.

(ii) Relative to the loop (3.a)↔(1.b), there exists a finite $n = n_k$ such that (3.b) is reached.

Proof. By assumption, $0 \notin T(x^k)$. To prove (i), suppose, for contradiction, that (BS) loops forever in (1.e)↔(1.b). Then $j \to \infty$ and an infinite sequence $\{s^{k,n,j}\}_{j\in\mathbf{N}}$ is generated, satisfying $\|s^{k,n,j}\| \le \tau 2^{-j}$. Therefore, there exist two subsequences $\{n_q\}$, $\{j_q\}$ such that

$$\|s^{k,n_q,j_q}\| \le \tau 2^{-j_q}\,, \tag{8}$$

with $\lim_{q\to\infty} j_q = \infty$. For such indices, define $I_q := I_{k,n_q,j_q}$. Because of step (1.b), for all $i \in I_q$, $\|z^i - x^k\| \le R2^{-j_q}$. Consider the convex sum given by $\alpha^q := \alpha^{k,n_q,j_q}$ from step (1.c):

$$(\hat{x}^q, \hat{s}^q) := (\sum_{i\in I_q} \alpha_i^q z^i, s^{k,n_q,j_q})\,.$$

Corollary 2.3 applies, with $\rho = R2^{-j_q}$ and $\tilde{x} = x^k$, and we have

$$\hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q)\,, \quad \text{with} \quad \hat{\varepsilon}_q \le 2R2^{-j_q}M\,, \tag{9}$$

where $M := \sup\{\|u\| \mid u \in T(\overline{B(x^k, R)})\}$. In addition,

$$\|\hat{x}^q - x^k\| \le R2^{-j_q}\,. \tag{10}$$

Altogether, letting $q \to \infty$ in (8), (9) and (10), Proposition 2.1 yields:

$$(\hat{\varepsilon}_q, \hat{x}^q, \hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q)) \longrightarrow (0, x^k, 0) \quad \Longrightarrow \quad 0 \in T^0(x^k) = T(x^k)\,,$$

a contradiction. Hence, there exists a finite j such that the loop (1.e)↔(1.b) ends. For this index j, step (1.f) is reached, $j_{k,n} = j$ is defined with $\|s^{k,j_{k,n}}\| > \tau 2^{-j_{k,n}}$, and the first part in (i) holds.
Furthermore, the test in (2.c) will eventually be false and the loop (2.c)↔(2.b) ends, with a value of $l = l_{k,n} \le j_{k,n} + 1$.
Now we prove (ii). If an infinite loop occurs at (3.a)↔(1.b), then $n \to \infty$. Thus, at step (2.d) an infinite sequence $\{(y^{k,n}, v^{k,n} \in T(y^{k,n})\}_{n\in\mathbf{N}}$ is generated. We have that for each n, the loop (1.e)→(1.b) ends with an index j such that

$$s^{k,n,j} = s^{k,n,j_{k,n}} = s^{k,n} \quad \text{and} \quad \|s^{k,n}\| > \tau 2^{-j_{k,n}}\,. \tag{11}$$

We proved in (i) that j eventually reaches its final value, say J. Therefore, there exists $\bar{n}$ such that $j_{k,n} = j_{k,\bar{n}} = J$ for any $n \ge \bar{n}$. Consider now the sequence $\{(y^{k,n}, v^{k,n})\}_{n\ge\bar{n}}$. At Step (3.a) infinite null steps are made:

$$\langle v^{k,n}, s^{k,n}\rangle \le \sigma \|s^{k,n}\|^2 \tag{12}$$

with

$$\|y^{k,n} - x^k\| = R\,2^{-l_{k,n}} = R\,2^{-j_{k,n}-1} = R\,2^{-J-1}\,, \tag{13}$$

for all $n \ge \bar{n}$ (if (2.a) leads to (3.a) then $l_{k,n} = j_{k,n} + 1$). This means that $(y^{k,n}, v^{k,n})$ is incorporated to the sub-bundle associated to $I_{k,n,j_{k,n}}$ for any $n \ge$

$\bar{n}$. In particular, choose an index $\tilde{n}$ such that $\bar{n} < \tilde{n} < n$: since $j_{k,n} = J$ in (13), $(y^{k,\tilde{n}}, v^{k,\tilde{n}})$ is incorporated to the sub-bundles defining $s^{k,n}$, the projection of 0 onto the convex hull of $\{w^i\}_{i \in I_{k,n,j}}$ (cf. steps (1.c)-(1.f)). Therefore, by the classical projection property (see for instance [21])

$$\langle v^{k,\tilde{n}}, s^{k,n} \rangle \geq \|s^{k,n}\|^2 , \qquad \text{for any } n > \tilde{n} > \bar{n} \ .$$

Together with (12) written for $n > \tilde{n}$, we obtain

$$\langle v^{k,\tilde{n}} - v^{k,n}, s^{k,n} \rangle \geq (1 - \sigma)\|s^{k,n}\|^2 \text{ for any } n > \tilde{n} > \bar{n} \ . \tag{14}$$

We claim that the assumptions of Lemma 4.1 hold, after a suitable renaming is done. Actually, define, for all $i \geq 1$,

$$t^i := v^{k,\bar{n}+i}, \quad \tilde{t}^i := s^{k,\bar{n}+i+1} \, .$$

Using now (14), we obtain

$$\langle t^i - t^{m+1}, \tilde{t}^m \rangle \geq (1 - \sigma)\|\tilde{t}^m\|^2 \text{ for any } i = 1, \ldots, m \, .$$

Therefore, the sequences $\{t^m\}$ and $\{\tilde{t}^m\}$ satisfy condition (7) in Lemma 4.1 with $\gamma = 1 - \sigma$. Moreover, because $\{t^i\} \subseteq T(B(x^k, R))$ and T is locally bounded, the last part in Lemma 4.1 also applies:

$$\tilde{t}^m \to 0 \quad \text{when} \quad m \to \infty \, .$$

However, (11) and the choice of $\bar{n}$ yields

$$\|\tilde{t}^m\| = \|s^{k,\bar{n}+m+1}\| > \tau \, 2^{-J} > 0 \, ,$$

a contradiction. Altogether, the loop (3.a)$\leftrightarrow$(1.b) must eventually finish with a finite value of n and (ii) is proved. ☐
In the next result we analyze all the possibilities for an iteration of (BS).

Proposition 4.3 *Let x^k be the current iterate in (BS). Then*

- (i) *if x^k is a solution, either the oracle answers $u^k = 0$ and (BS) stops in (0.a), or (BS) loops forever after this last serious step, without updating k.*

- (ii) *Else, x^k is not a solution, and (BS) reaches step (3.b) after finitely many inner iterations. Furthermore,*

$$\|s^{k,n_k^*,j_k-1}\| \leq \tau 2^{-j_k+1}, \tag{15}$$

where n_k^* is the smallest value of n equating $j_{k,n} = j_k$, whenever $j_k > 0$.

Proof. Suppose first that x^k is a solution. If the oracle gives $u^k = 0$, the stopping test holds in (0.a) and (BS) stops. Otherwise, $u^k \neq 0$. Then, suppose for contradiction that (3.b) is reached. Recall that by (6), we have

$$\begin{array}{rcl}\langle x^k - y^k, v^k\rangle & = & R2^{-l_k}\langle s^k, v^k\rangle/\|s^k\| \\ & > & 2^{-l_k-j_k}\sigma\tau R > 0\,, \quad \text{with} \quad v^k \in T(y^k)\,.\end{array}$$

Because $0 \in T(x^k)$, the inequality above contradicts the monotonicity of T and (i) is proved.
Let us prove (ii). Suppose $j_k > 0$. If x^k is not a solution, Lemma 4.2 shows that no infinite loop occurs inside iteration k and (3.b) is eventually reached. To prove (15), define

$$n_k^* := \min\{n \le n_k |\, j_{k,n} = j_k\}\,.$$

Then in (1.e)↔(1.b), the indices

$$j := j_{k,n_k^*} - 1 < j_k \quad \text{and} \quad j+1 = j_{k,n_k^*} = j_k\,,$$

are such that (1.e) holds for index j and (1.f) holds for index $j+1$:

$$\|s^{k,n_k^*,j_{k,n_k^*}-1}\| \le \tau 2^{-(j_{k,n_k^*}-1)}\,, \quad \text{and} \quad \|s^{k,n_k^*,j_{k,n_k^*}}\| > \tau 2^{-(j_{k,n_k^*})}\,,$$

and the conclusion follows. □
As a consequence of the last result, the sequence of serious points $\{x^k\}$ generated by (BS) is either finite, ending at a solution; or infinite, with no iterate being a solution. Before proving the convergence for the infinite sequence, we need some preliminary technical results.

Proposition 4.4 *Let x^k be the current iterate in (BS) and assume $x^k \notin \mathcal{S}$. Then, after x^{k+1} is generated in (3.b), the following holds:*

(i) Let H_{y^k,v^k} be the halfspace defined in (2)*, written with* $(y,v) := (y^k, v^k)$*. Then $x^k \notin H_{y^k,v^k}$ and $x^{k+1} = P_{H_{y^k,v^k}}(x^k)$.*

(ii) For all $x^* \in \mathcal{S}$, $\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2$.

(iii) Finally, $\|x^{k+1} - x^k\| > R\sigma\tau 2^{-2j_k-1}/\|v^k\|$.

Proof. To prove (i), recall from (6) that the pair $(y^k = x^k - R2^{-l_k}s^k, v^k = v^{k,n_k})$ is such that

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

so $x^k \notin H_{y^k,v^k}$. To see that x^{k+1} is its projection, just recall the definition of x^{k+1} in step (3.a) of (BS). Because x^{k+1} is an orthogonal projection onto H_{y^k,v^k} and $\mathcal{S} \subset H_{y^k,v^k}$, (ii) follows from the properties of orthogonal projections (see e.g. [22].) As for (iii), it is also straightforward from (6) and the definition of x^{k+1}. □
The boundedness of the variables generated by (BS) now follows from the Fejér convergence of $\{x^k\}$ to $\mathcal{S}$, via the following obvious statement.

Proposition 4.5 *Consider a sequence $\{x^k\}$ be such that*

$$\|x^{k+1} - x\| \leq \|x^k - x\|\,, \quad \textit{for any } x \in \mathcal{S}.$$

with $\mathcal{S} \neq \emptyset$. Then

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

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

□

All the variables generated by (BS), namely x^k, s^k, y^k, v^k, $\{(y^{k,n}, v^{k,n})\}$ and $\{(z^p, w^p)\}$ are bounded. Our final theorem only needs the boundedness of the last two sequences.

Lemma 4.6 *The sequences $\{(y^{k,n}, v^{k,n})\}$ and $\{(z^p, w^p)\}$ generated by (BS) in Steps (2.d) and (0.b), respectively, are bounded.*

Proof. First we show that the sequence $\{x^k\}$ is bounded. If the sequence is finite, the boundedness is trivial. If $k \to \infty$, by Proposition 4.4(ii) the assumptions of Proposition 4.5 hold. Hence, using Proposition 4.5(i), we obtain that $\{x^k\}$ is bounded.
Therefore, there exists some compact set K_0 such that the (bounded) sequence $\{x^k\} \subset K_0$. Define $K_1 := K_0 + \overline{B(0,R)}$. Then, from steps (2.b), (2.d) and (3.b) in (BS), it follows that the variables $y^{k,n}$ and y^k are contained in K_1. Since z^p is extracted from either $\{x^k\}$ or $\{y^k\}$, the sequence is also contained in K_1. Finally, because T is locally bounded, $\{v^k\}$ and $\{w^p\}$ are bounded too and the proof is finished. □
In our last Lemma we show that index j_k in (BS) goes to infinity together with k. Along the lines of [3, Section 3.3], convergence of (BS) will be proved using Proposition 4.5 (ii), by exhibiting a subsequence of triplets $(\hat{\varepsilon}_q, \hat{x}^q, \hat{s}^q \in T^{\varepsilon_q}(\hat{x}^q))$ tending to $(0, \bar{x}, 0)$ as q is driven by j_k to infinity.

Lemma 4.7 *Suppose (BS) loops forever on k (i.e., $k \to \infty$). Then* $\lim_{k\to\infty} j_k = +\infty$.

Proof. Combine Proposition 4.4(iii) and (ii) to obtain

$$R\sigma\tau 2^{-2j_k-1}/\|v^k\| < \|x^{k+1} - x^k\| \to 0\,, \quad \text{when } k \to \infty\,.$$

Since $\|v^k\|$ is bounded (Lemma 4.6), the result follows. □
Finally, we state our main convergence result.

Theorem 4.8 *Consider the sequence $\{x^k\}$ generated by (BS). Then the sequence is either finite with last element in $\mathcal{S}$, or it converges to a solution of* (1).

Proof. We already dealt with the finite case in Proposition 4.3.
If there are infinitely many x^k, keeping Proposition 4.5(ii) in mind, we only need to show that some accumulation point of the bounded sequence $\{x^k\}$ is a solution of (1). Let $\{x^{k_q}\}$ be a convergent subsequence, with limit point $\bar{x}$. Because of Lemma 4.7, we can suppose $j_{k_q} > 0$, for q large enough. Then Proposition 4.3(ii) applies: for n^*_k defined therein, we have

$$\|s^{k_q,n^*_{k_q},j_{k_q}-1}\| \leq \tau\, 2^{-j_{k_q}+1}\,. \tag{16}$$

Consider the associated index set $I_q := I_{k_q,n^*_{k_q},j_{k_q}-1}$. By developing an argument similar to the one used in the proof of Lemma 4.2, mutatis mutandis, define

$$\begin{aligned} \alpha^q &:= \alpha^{k_q,n^*_{k_q},j_{k_q}-1}\,, \\ \hat{x}^q &:= \textstyle\sum_{i\in I_q} \alpha^q{}_i z^i\,, \\ \hat{s}^q &:= s^{k_q,n^*_{k_q},j_{k_q}-1} \quad = \quad \textstyle\sum_{i\in I_q} \alpha^q_i w^i\,. \end{aligned}$$

We have that

$$\|\hat{x}^q - x^{k_q}\| \leq R\, 2^{-j_{k_q}+1}. \tag{17}$$

Let M be an upper bound for $\|w^p\|$ (these variables are bounded by Lemma 4.6). Then Corollary 2.3 yields

$$\hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q) \text{ with } \hat{\varepsilon}_q \leq 2R\, 2^{-j_{k_q}+1} M. \tag{18}$$

Using Lemma 4.7 we have $\lim_{q\to\infty} j_{k_q} = \infty$. Hence, by (16), (17), (18) we conclude that

$$(\hat{\varepsilon}_q, \hat{x}^q, \hat{s}^q \in T^{\hat{\varepsilon}_q}(\hat{x}^q)) \longrightarrow (0, \bar{x}, 0)\,,$$

when $q \to \infty$. Now Proposition 2.1 applies, implying that $0 \in T(\bar{x})$. □

5 Conclusions and Perspectives

We presented a convergent bundle-like method that can be used for finding a zero of maximal monotone operators. Although it already has an implementable form, further modifications are possible:

1. Parameters R and τ in (BS) can be replaced by $\|u^k\|$, in order to deal with “relative” values.

2. To avoid memory overload, a variant of (BS) working with sub-bundles of limited size could be defined. Using techniques of compression and selection which are now standard for bundle methods in nonsmooth optimization (NSO) (see, for instance, Chapter XIV in [5]), Theorem 2.2 ensures that after solving (1.c) in (BS), the *aggregated* triplet

$$(\hat{\varepsilon}, \hat{x} := \sum_{i\in I_{k,n,j}} \alpha_i^{k,n,j} z^i, \hat{s} := s^{k,n,j})$$

synthesizes the most essential information contained in the sub-bundle (we have shortened $\hat{\varepsilon} = \sum_i \alpha_i \langle z^i - \hat{x}, w^i - \hat{s} \rangle$).

3. We insisted on the importance of driving ε to 0 at an appropriate speed to avoid in Step (1.c) what we have called an *ill-posed* problem. Because of Corollary 2.3, we achieve this goal indirectly through j-indices. In NSO, it is well known that the practical efficiency of bundle methods depends strongly on an appropriate management of the parameters. For example, (BS) could be modified in order to define ε as a *parameter* whose value is reduced along iterations. To achieve this direct control on ε, a possibility (not explored yet) could be to solve the following subproblem in *Step 1*:

$$\left\{ \begin{array}{l} \min_{\alpha \in \Delta_I} \frac{1}{2} \| \sum_I \alpha_i w^i \|^2 \\ \sum_I \langle z^i - \hat{x}, w^i - \hat{s} \rangle \leq \varepsilon \,, \end{array} \right.$$

where $\hat{x}, \hat{s}$ have been computed in the previous iteration, and I defines a sub-bundle.

Finally, let us mention the recent paper [9], describing an algorithm for solving variational inequality problems (VIP) with convex feasible sets. The methodology presented there has been developed independently and can be related to our (BS). Roughly speaking, **Procedure P** therein corresponds to our computation of $s^{k,n,j}$ using an "economic" bundle formed by only two elements, namely $v^{k,n}$ and $s^{k,n,j-1}$ (or $s^{k,n-1,j}$). In his paper, Konnov himself refers to the conjugate subgradients method for NSO in [21] (see also [12]). Since these very first "economic" (conjugate subgradient) formulations a huge progress has been done on bundle methods for NSO. In particular, it is known that the use of "rich" bundles (properly managed to avoid overload) improves dramatically the speed of convergence. In this sense, our contribution, based on "rich" sub-bundles, may open the way to defining new variants, hopefully more efficient and robust. Moreover, since our development is supported by the theoretical background of ε-"subgradients" in $T^{\varepsilon}(x)$, it gives a different insight of the mechanism that makes algorithms of this type work.

References

[1] A. Brøndsted and R.T. Rockafellar. On the subdifferentiability of convex functions. *Proc. of the Amer. Math. Soc.*, 16:605–611, 1965.

[2] 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. Also (extended version) Tech. Rep. B-110/97, IMPA, Rio de Janeiro, Brazil.

[3] R.S. Burachik, C.A. Sagastizábal, and B. F. Svaiter. ε-Enlargements of maximal monotone operators: Theory and Application. In *Reformulation –*

Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, pages 25–43. Kluwer, 1998.

[4] I. Ekeland and G. Lebourg. Sous-gradients approchés et applications. *C.R. Acad. Sc. Paris, série A*, 281:219–222, 1975.

[5] 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).

[6] J.-J.Strodiot and V.H. Nguyen. On the numerical treatment of the inclusion $0 \in \partial f(x)$. In J. J. Moreau, P. D. Panagiotopolus, and G. Strang, editors, *Topics in Nonsmooth Mechanics*, pages 267–294. Birkhäuser Verlag, 1988.

[7] E. N. Khobotov. Modifications of the extragradient method for solving variational inequalities and certain optimization problems. *USSR Computational Mathematics and Mathematical Physics*, 27(5):120–127, 1987.

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

[9] I.V. Konnov. A combined relaxation method for variational inequalities with nonlinear constraints. *Mathematical Programming*, 80:239–252, 1997.

[10] G.M. Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[11] J. E. Martínez Legaz and M. Théra. ϵ-subdifferentials in terms of subdifferentials. *Set-Valued Analysis*, 4:327–332, 1996.

[12] C. Lemaréchal. An extension of Davidon methods to nondifferentiable problems. *Mathematical Programming Study*, 3:95–109, 1975.

[13] C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods: from conceptual to implementable forms. *Mathematical Programming*, 76:393–410, 1997.

[14] C. Lemaréchal, J.-J. Strodiot, and A. Bihain. On a bundle method for nonsmooth optimization. In O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, editors, *Nonlinear Programming 4*, pages 245–282. Academic Press, 1981.

[15] M. Nisipeanu. *Somme variationnelle d'opérateurs et applications.* PhD thesis, Université de Limoges - France, 1997.

[16] J. P. Revalski and M. Théra. Enlargements and sums of monotone operators. Working paper, 1998.

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

[18] 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.

[19] D. Torralba. *Convergence épigraphique et changements d'échelle en Analyse Variationnelle et Optimisation*. PhD thesis, Université de Montpellier II - France, 1996.

[20] L. Veselý. Local uniform boundedness principle for families of ε-monotone operators. *Nonlinear Analysis, Theory, Methods & Applications*, 24(9):1299–1304, 1995.

[21] P. Wolfe. A method of conjugate subgradients for minimizing nondifferentiable functions. *Mathematical Programming Study*, 3:145–173, 1975.

[22] E. H. Zarantonello. Projections on convex sets in Hilbert spaces and spectral theory. In E. H. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, pages 237–424. Academic Press, 1971.