ON THE RELATION BETWEEN BUNDLE METHODS FOR MAXIMAL MONOTONE INCLUSIONS AND HYBRID PROXIMAL POINT ALGORITHMS

Claudia A. Sagastizábal [a] * and Mikhail V. Solodov [b] †

[a]IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. On leave from INRIA, BP 105, 78153 Le Chesnay Cedex, France.
Email : sagastiz@impa.br

[b]IMPA– Instituto de Matemática Pura e Aplicada,
Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email : solodov@impa.br

We demonstrate that bundle methods for computing zeroes of general maximal monotone operators can be cast as a special case in a certain class of hybrid proximal point algorithms. This fact is significant for several reasons. First, it provides an alternative convergence proof, which is technically simple, for serious steps of bundle methods by invoking the corresponding results for hybrid proximal point methods. This includes the linear rate of convergence results, which were not available previously. Second, relating the two methodologies supplies a computationally realistic implementation of hybrid proximal point methods for the most general case, i.e., for operators without any particular structure.

1. Introduction

Given a maximal monotone operator T on $\Re^n$, we consider the problem:

$$\text{Find } x \in \Re^n \text{ such that } 0 \in T(x)\,. \tag{1}$$

As is well known, this problem is fundamental in mathematics and related sciences. It also subsumes many important practical applications. When T has some specific structure, specialized algorithms can usually be used to solve (1). For example, when $T = \partial f$, the subdifferential of a convex function f, methods of convex optimization come into play, including bundle techniques, see [15, Chapters XIV, XV] and [5, Part II]. When T represents a variational inequality over a closed convex set with a continuous mapping, then certain projection methods can be used, see [14] and [33, Introduction] for references and relevant discussions. However, when T is a general maximal monotone operator,

*Research partially supported by CNPq under Grant No. 301800/96-0 and by FAPERJ under Grant No. 150-581/00

†Research is supported in part by CNPq Grant 300734/95-6, by PRONEX–Optimization, and by FAPERJ.

relatively few methods are applicable. Until very recently, only *implicit* procedures based on resolvents (i.e., proximal-type methods) were available for solving (1) in the general case. In fact, the bundle method of [8] and the projection method with averaging of [1] appear to be currently the only *explicit* methods for solving general maximal monotone inclusions. To make this remark precise, we note that we consider an algorithm for solving (1) as explicit if:

- It only requires the knowledge of *one* $v \in T(x)$ for each given $x \in \Re^n$ (recall that the ability of an algorithm to work with this minimal information is a well-known practical requirement in nonsmooth optimization, which is an important example of problem (1)).
- All parts of the method are computationally implementable, independently of the structure of T.
- The method converges under no assumptions other than the maximal monotonicity of T and $T^{-1}(0) \neq \emptyset$.

The iterative procedure of [1] is essentially a subgradient-type method with vanishing stepsizes, where convergence is "forced" by Cesàro averaging. Methods of this type have inherently slow (sublinear) convergence even for well-structured problems. In optimization, bundle methods are known to be in practice typically superior to standard subgradient algorithms, e.g., such as in [30]. On the theoretical side, bundle methods also have the potential for the faster linear rate of convergence. It should be mentioned, however, that the method of [1] is developed in a considerably much more general setting than [8] or the present paper (in [1], the Banach space setting is considered, and the problem data can be subject to perturbations).

In this paper we consider bundle methods under the light of inexact proximal point algorithms, namely the hybrid variant of [36], see also [35,37,38]. The insight given by this new interpretation is twofold. First, it provides an alternative convergence proof, which is technically simple, for serious steps of bundle methods by invoking the corresponding results for hybrid proximal point methods. Second, relating the two methodologies supplies a computationally realistic implementation of hybrid proximal point methods for the most general case, i.e., when the operator may not have any special structure.

Our paper is organized as follows. In § 2 we outline the hybrid proximal point algorithm, together with its relevant convergence properties. Some useful theory from [9] and [8] on certain enlargements of maximal monotone operators is reviewed in § 3. Finally, in Section 4 we establish the connection between bundle and hybrid proximal methods and give some new convergence results, including the linear rate of convergence for bundle methods.

2. Hybrid proximal point algorithms

A classical algorithm applicable to solve (1) is the proximal point method, e.g., [24,29, 6,26,23,18]. This algorithm consists of solving a sequence of subproblems of the following structure:

Given $x^k \in \Re^n$, find $x^{k+1} \in \Re^n$ and $v^{k+1} \in T(x^{k+1})$ such that $0 = v^{k+1} + \mu_k(x^{k+1} - x^k)$,(2) where $\mu_k > 0$ is a regularization parameter. It is quite clear that in this form the proximal point algorithm is purely conceptual (in other words, not implementable and/or not practical). This has to do with the obvious fact that subproblems (2) are structurally as difficult to solve as the original problem (1) itself. Nevertheless, the ideas of the proximal point method are useful for developing more sophisticated (and implementable!) computational algorithms. A well-known example is precisely the family of bundle methods for nonsmooth convex optimization, see [15,20,4,21].

To pass from the domain of conceptual to implementable, one has to start with inexact versions of (2). Perhaps the most natural relaxation of (2) can be stated as follows:

$$\text{Given } x^k \in \Re^n\,, \text{ find } y^k \in \Re^n \text{ and } v^k \in T(y^k) \text{ such that } 0 = v^k + \mu_k(y^k - x^k) + r^k\,, \tag{3}$$

where $r^k \in \Re^n$ is some error term induced by the approximation method. Traditionally, see [29,39,7,10,13,11], inexact proximal algorithms defined the next iterate as the approximate solution given by (3), i.e., $x^{k+1} = y^k$. To guarantee convergence, error terms were required to satisfy some *"a priori summability"* condition, such as

$$\|r^k\| \leq \sigma_k\,, \quad \sum_{k=0}^{\infty} \sigma_k < \infty\,. \tag{4}$$

Recently, the so-called hybrid proximal methods appeared as an alternative proposal aimed to develop more constructive (and hopefully practical) approximation criteria. We refer the reader to [36,35,37,38] for more detailed discussions. Here, we shall only consider the approximation criterion of [36]. Specifically, we regard a pair $(y^k, v^k) \in \Re^n \times \Re^n$ as an acceptable approximation to the solution of the k-th proximal subproblem (2), if it satisfies (3) with

$$\|r^k\| \leq \sigma_k \max\{\|v^k\|, \mu_k\|y^k - x^k\|\}\,, \quad \sigma_k \in [0,1)\,. \tag{5}$$

This rule is constructive, in the sense that it is clearly related to the behaviour of the method around the current point for the problem being solved. Furthermore, as opposed to (4), the relaxation parameters σ_k can be kept bounded away from zero (in fact, they can be even fixed), which essentially means that the *relative* error in (3) need not tend to zero. Computationally, the latter is a realistic condition, suitable for applications. Moreover, criterion (5) is weaker than standard conditions, like (4). This can be seen from the fact (see an example in [36]) that a sequence generated according to (3) and (5) with $x^{k+1} = y^k$, need not converge under the assumption $\liminf_{k\to\infty} \sigma_k > 0$. By contrast, a sequence generated using (4) does converge. Nevertheless, approximate solutions satisfying (3) and (5) are "good enough". Specifically, convergence can be forced by introducing a simple (explicit) projection step at no additional computational cost.
The algorithm proposed in [36] is the following.

Algorithm 2.1 Hybrid Projection–Proximal Point Method (HPPM)
Choose any $x^0 \in \Re^n$ and set $k := 0$.

Inexact proximal step. *Having x^k, choose $\mu_k \in (0, +\infty)$, $\sigma_k \in [0, 1)$, and solve (approximately) the k-th proximal subproblem so that* (3) *and* (5) *are satisfied.*
Stop if $v^k = 0$ or $y^k = x^k$. Otherwise, compute

Projection step.

$$x^{k+1} = x^k - \langle v^k, x^k - y^k \rangle \|v^k\|^{-2} v^k \,. \tag{6}$$

Set $k := k+1$*; and repeat.* ∎

Remark 2.1 Since it is not specified how to solve subproblems in order to satisfy (3) and (5), at this stage HPPM is also a conceptual scheme. However, in a number of applications where T has some structure, the constructive nature of (5) allows to solve subproblems by simple and efficient means. Such is the case of *truly* globally convergent Newton-type methods for systems of monotone equations [34] and complementarity problems [32]. The key fact is that under certain natural assumptions, just one step of Newton method is sufficient to solve the k-th proximal subproblem within the error tolerance given by (5). See also [31] for some other possibilities in the context of variational inequalities. Another interesting application is the modified forward-backward splitting method of [40], which can be shown to be a specific realization of the method of [35] (closely related to HPPM) in the case when T is a sum of two maximal monotone operators, one of which is continuous. In § 4, we show how to develop a (constructive) bundle method satisfying (3) and (5) for general maximal monotone T.

Remark 2.2 HPPM is well-defined, because the exact solution of the k-th proximal point subproblem always satisfies (3) and (5) for any $\sigma_k \geq 0$. If $\sigma_k = 0$, only the exact solution is acceptable; otherwise there are many acceptable (approximate) solutions.

Remark 2.3 If $x^k \notin T^{-1}(0)$, it can be seen that (3) and (5) imply that $\langle v^k, x^k - y^k \rangle > 0$. On the other hand, by the monotonicity of T, $\langle v^k, \bar{x} - y^k \rangle \leq 0$ for any $\bar{x} \in T^{-1}(0)$. It follows that the hyperplane $\{x \in \Re^n \mid \langle v^k, x - y^k \rangle = 0\}$ strictly separates x^k from $T^{-1}(0)$. The last step of HPPM, namely (6), projects x^k onto this hyperplane, essentially ensuring global convergence.
The projection step of HPPM plays here the role of "forcing" convergence. It is interesting to note that in the proximal method of [2] a somewhat similar role seems to be played by averaging. We point out, however, that the method of [2] itself is very different from HPPM, which is based on constructive approximation criteria for solving the subproblems.

HPPM preserves all the desirable convergence properties of the classical exact proximal point algorithm. Specifically, assuming that the method does not terminate finitely (which can only occur at the exact solution of (1)), we have the following results.

Theorem 2.1 *[36, Section 2] Suppose that T is a maximal monotone operator, with $T^{-1}(0) \neq \emptyset$, and consider HPPM Algorithm 2.1. Then the following hold:*

(i) The sequence $\{x^k\}$ generated by HPPM is bounded.

(ii) If parameters are chosen so that $\bar{\sigma} := \limsup_{k\to\infty} \sigma_k < 1$ and $\sum_{k=0}^{\infty} \mu_k^{-2} = +\infty$, then $\{x^k\}$ converges to some $\bar{x} \in T^{-1}(0)$.

(iii) If, in addition, there exists $\bar{\mu} > 0$ such that $\mu_k \le \bar{\mu}$ for all k, and two constants $L > 0$ and $\delta > 0$ such that

$$\|v\| \le \delta\,, v \in T(x) \quad \Longrightarrow \quad \mathrm{dist}(x, T^{-1}(0)) \le L\|v\|\,, \tag{7}$$

then $\{x^k\}$ converges to $\bar{x} \in T^{-1}(0)$ R-linearly.

(iv) If, in addition to (7), $\bar{x} \in T^{-1}(0)$ is unique, then the convergence rate is Q-linear.

■

When T is a general maximal monotone operator, up to now HPPM remained as conceptual as the original proximal point method, as so far no techniques have been proposed to solve the proximal point subproblems to satisfy (3) and (5) in this general case. We show in § 4 below how this can be done by the bundle strategy of [8]. Specifically, we demonstrate that a modification of the method of [8] can be viewed as a particular practical realization of HPPM. As a by-product, we obtain an immediate convergence proof for serious steps of the bundle method, including the linear rate of convergence result. We note that the rate of convergence analysis for bundle methods was not previously available.

3. Bundle methods

We shall assume from now on that the domain of T is the whole space $\Re^n$. From a purely theoretical point of view, this assumption may be deemed restrictive. However, it is important to keep in mind that we are considering implementable algorithms which are developed using just one $v \in T(x)$ for each $x \in \Re^n$. Formally speaking, this precludes T from having a restricted domain. However, in most applications infinite values are not inherent in the problem, but are rather induced by the indicator function of the feasible region. If one is to deal with constraints explicitly (via projections, for example, so that T needs to be evaluated only at feasible points), then bundle methods can be extended to constrained problems, after taking care of the usual technicalities (see also Remark 3.4). For the sake of clarity of presentation, we shall not deal with this issue in the present paper.

3.1. Enlargements of maximal monotone operators

It is known that the ε-subdifferential plays a paramount role in the convergence analysis of bundle methods for nonsmooth optimization. For a general maximal monotone operator, a similar tool is a certain outer approximation of T. More precisely, given some $\varepsilon \ge 0$, the ε-enlargement of T at $x \in \Re^n$ [7] is defined by

$$T^\varepsilon(x) := \{u \in \Re^n \mid \langle v - u, y - x\rangle \ge -\varepsilon \text{ for all } y \in \Re^n\,, v \in T(y)\}\,.$$

It is clear that $T^0 = T$, and $T(x) \subset T^\varepsilon(x)$ for all $x \in \Re^n\,, \varepsilon \ge 0\,.$

One of the most important properties of T^ε is that, unlike T itself, it is continuous as an operator of both x and ε. Continuity is crucial when constructing practical algorithms, because it ensures the following key property:

$$\text{Given } \{(\varepsilon_k, x^k)\} \to (0, \bar{x} \in T^{-1}(0))\,, \text{ there exists } \{s^k \in T^{\varepsilon_k}(x^k)\} \text{ such that } s^k \to 0\,.$$

Information accumulated over the course of iterations is related to the ε-enlargement of T through the following consequence of the weak *transportation formula* [9, Theorem 2.2]:

Lemma 3.1 *[8, Corollary 2.3] Take any $z^i \in \Re^n\,, w^i \in T(z^i)\,, i = 1\,, \ldots\,, j$. Let $x \in \Re^n$ and $\rho > 0$ be such that $\|z^i - x\| \le \rho$ for all $i = 1\,, \ldots\,, j$. Let*

$$(\hat{x}, \hat{s}) := (\sum_{i=1}^{j} \lambda_i z^i, \sum_{i=1}^{j} \lambda_i w^i)\,, \qquad \sum_{i=1}^{j} \lambda_i = 1\,, \lambda_i \ge 0\,, i = 1\,, \ldots\,, j\,.$$

Then it holds that

$$\hat{s} \in T^\varepsilon(\hat{x}) \quad \text{with } \|\hat{x} - x\| \le \rho \text{ and } \varepsilon \le 2\rho M\,, \text{ where } M := \max\{\|w^i\| \mid i = 1, \ldots, j\}\,.$$

■

For properties and applications of ε-enlargements of maximal monotone operators, as well as other related issues, we refer the reader to [7,9,8,35,37,41,27].

3.2. Towards an implementable algorithm

Before stating our implementable bundle method, which is technically quite involved, we describe a related conceptual scheme, introduced in [9], which is helpful for transmitting some key ideas.

Algorithm 3.1 Conceptual Algorithmic Scheme (CAS)
Choose positive parameters $\tau\,, R\,, \varepsilon, \alpha$ and θ, with $\alpha, \theta \in (0,1)$. Choose any $x^0 \in \Re^n$. Set $k := 0$.

Compute search direction. *Stop if $0 \in T(x^k)$.*
Otherwise, find m_k, the smallest *nonnegative integer m, such that*

$$\|s^k\| > \alpha^m \tau\,, \text{ where } s^k := \arg\min\{\|v\| \mid v \in T^{\varepsilon\alpha^m}(x^k)\}\,.$$

Line-search step. *Find l_k, the* smallest *nonnegative integer l, such that*

$$\langle v^k, s^k \rangle > \theta \|s^k\|^2\,, \text{ where } v^k \in T(y^k)\,, y^k := x^k - \alpha^l R s^k\,.$$

Projection step. *Compute*

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

Set $k := k + 1$; and repeat. ■

Remark 3.1 In the given form, CAS is a conceptual method because computing the minimal-norm element of $T^{\varepsilon}(x^k)$ (actually, even of $T(x^k)$) is in general prohibitively computationally expensive, and in fact, typically impossible. At the very least, for this one should have available a full functional representation for the set $T^{\varepsilon}(x^k)$ for all ε and x^k, which is certainly too much to wish for. Recalling nonsmooth optimization, one practical requirement in this setting is that algorithms should be developed with the use of only one subgradient at each point. In order to develop stable methods with reliable stopping tests, it is also important to keep track of previous information along iterations, in the form of a *bundle* of information. This is the underlying philosophy in bundle methods for nonsmooth optimization, which constituted a huge progress when compared to subgradient methods. It is precisely the use of bundle methods that allowed efficient numerical solution of complex real-world applications [19,3].

Remark 3.2 CAS can be viewed as an analog of the steepest-descent method for nonsmooth convex optimization. Indeed, if we consider the minimal-norm element v of $\partial f(x)$, then by the optimality condition it holds that $\langle v, u - v\rangle \geq 0$ for all $u \in \partial f(x)$. Hence,

$$\|v\|^2 = \min_{u\in\partial f(x)} \langle v, u\rangle \leq \min_{u\in\partial f(x)} \max_{w\in\partial f(x)} \langle w, u\rangle = \min_{u\in\partial f(x)} f'(x; u)\,,$$

where $f'(x;u)$ stands for the usual directional derivative of f at $x \in \Re^n$ in the direction $u \in \Re^n$. Assuming that x is not a minimizer of f, v is then the direction of maximal descent for f at $x \in \Re^n$, and one can perform a line-search in this direction to decrease the value of the objective function f.

Remark 3.3 It was shown in [9, Proposition 3.2] that CAS is well-defined (m_k and l_k are finite integers for each k), and generates a (finitely or infinitely) convergent sequence, provided $T^{-1}(0) \neq \emptyset$.

Remark 3.4 CAS is related to methods proposed in [16] for solving multi-valued variational inequalities. However, the methods of [16] are not aimed at devising implementable algorithms. On the other hand, [16] can be combined with CAS to produce a constrained version of the conceptual scheme. This constrained version can then serve as the basis for a constrained bundle method to handle the case when T has restricted domain.

4. Bundle method as a hybrid proximal point algorithm

To make the basic scheme of CAS computationally viable, in [8] a bundle strategy is employed to approximate the sets $T^{\varepsilon}(x^k)$, as follows:

- Polyhedral approximations of $T^{\varepsilon}(x^k)$ are defined using the information that has been accumulated along iterations. The line-search step in CAS is then modified to include an important safeguard. This safeguard essentially verifies whether the current polyhedral approximation is "good enough" or "poor". In the bundle methods terminology, these two situations correspond to serious or null steps, respectively. See Algorithm 4.1 below.

– Having in hand these suitable polyhedral approximations of $T^{\varepsilon}(x^k)$, they can be used to approximate minimal-norm elements of $T^{\varepsilon}(x^k)$. The latter problem can be solved by fast and reliable *quadratic programming* techniques, [22,12]. In nonsmooth optimization there exist very efficient quadratic programming solvers customized specifically for bundle methods, [17]. To keep the size of quadratic programs manageable, the method of [8] uses a selection rule to work with a reduced bundle at every step.

We proceed to formally state a bundle method, which is a modification of that in [8], adapted for our purposes. First, some notation is in order.

– Given $z^i \in \Re^n$ and $w^i \in T(z^i)$, $i = 1, \ldots, k$, by Δ_k we denote the bundle of accumulated information, i.e., $\Delta_k := \cup_{i=1}^{k} \{(z^i, w^i)\}$.

– Given a set of points $w^i, i \in I$, by $\operatorname{conv}\{w^i, i \in I\}$ we denote its convex hull, omitting the index set I when it is clear from the context.

– Serious-steps indices (i.e., indices of iterates yielding a significant progress toward a solution) will be collected in the set K_S.

Algorithm 4.1 Bundle Method (BM)

Choose positive parameters $\tau, R, \mathrm{TOL}, \alpha$ *and* θ, *with* $\alpha \in (0,1)$ *and* $\theta \in (1/2, 1)$. *Choose an integer* $\bar{m}$ *such that* $2\alpha^{\bar{m}} R \leq \mathrm{TOL}$. *Choose any* $x^0 \in \Re^n$ *and define* $\Delta_{-1} := \emptyset, K_S := \emptyset$. *Set* $k := 0$.

Major iteration. *Compute some* $u^k \in T(x^k)$. *Stop if* $u^k = 0$.
Otherwise, set $\hat{m}_k := 0$; $(z^k, w^k) := (x^k, u^k)$.

Update the bundle. *Set* $\Delta_k := \Delta_{k-1} \cup \{(z^k, w^k)\}$.

Compute direction. *For* $m = \hat{m}_k, \hat{m}_k + 1, \ldots$,

compute

$$s := \arg\min\{\|v\| \mid v \in \operatorname{conv}\{w^i\}, (z^i, w^i) \in \Delta_k, i \textit{ such that } \|z^i - x^k\| \leq \alpha^m R\},$$

until

$$\|s\| > \alpha^m \tau \quad \textit{or} \quad m = \bar{m}.$$

Set $s^k := s, \hat{m}_k := m$.

Line-search step. *For* $l = 0, 1, \ldots$,

compute

$$y := x^k - \alpha^l R \|s^k\|^{-1} s^k, \quad v \in T(y),$$

until

$$\langle v, s^k \rangle > \theta \min\{\|s^k\|, \|v\|\}^2 \quad \textit{or} \quad l = \hat{m}_k.$$

Set $y^k := y, v^k := v, l_k := l$.

Null step.

If

$$\langle v^k, s^k\rangle \le \theta \min\{\|s^k\|, \|v^k\|\}^2,$$

then

set $(z^{k+1}, w^{k+1}) := (y^k, v^k)$, $x^{k+1} := x^k$, $k := k+1$,

go to **"Update the bundle"** *step.*

Otherwise,

Serious step. *Define*

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

Set $k := k+1$*; and go to* **"Major iteration"** *step.* ■

Remark 4.1 We note that the criterion employed by our algorithm to accept a serious step is weaker (hence, easier to satisfy) than the criterion in [8], which is $\langle v, s^k\rangle > \theta\|s^k\|^2$. This improvement was made possible precisely due to relating our method to HPPM Algorithm 2.1 (see the proof of Theorem 4.1).

Remark 4.2 For each m, the task of computing direction s is equivalent to solving the quadratic program

$$\min_{\lambda}\{\|\sum_i \lambda_i w^i\|^2 \mid \lambda_i \ge 0, \sum_i \lambda_i = 1, i \text{ such that } \|z^i - x^k\| \le \alpha^m R, (z^i, w^i) \in \Delta_k\}.$$

Not that the problem constraints have a very special structure (the feasible region is the unit-simplex), which can be exploited by sophisticated quadratic programming solvers.

Remark 4.3 The "Compute direction" step in BM Algorithm 4.1 differs from CAS Algorithm 3.1 in that the sets $T^{\varepsilon}(x^k)$ in the latter are replaced by their polyhedral approximations in the former. Furthermore, these polyhedral approximations are constructed as a convex hull of certain *selected* elements in the full bundle Δ_k. Thus the method is always working with a *reduced* bundle, which is important for keeping the size of quadratic programs manageable. In addition, to control memory requirements for storing the full bundle Δ_k, one could use bundle compression techniques, similar to those in nonsmooth optimization [15, Chapter XIV].

The following is our main result, proving convergence when there is an infinite number of serious steps.

Theorem 4.1 *Suppose that T is a maximal monotone operator, with $T^{-1}(0) \neq \emptyset$, and consider the sequence of serious steps $\{x^k\}_{k\in K_S}$ generated by BM Algorithm 4.1. Assume further that the set K_S is infinite. Then the following hold:*

(i) The sequence $\{x^k\}_{k\in K_S}$ can be considered as a sequence generated by HPPM Algorithm 2.1. Furthermore, with parameters σ_k and μ_k chosen appropriately, the assumptions of Theorem 2.1(ii) hold, and so $\{x^k\}_{k\in K_S}$ converges to some $\bar{x} \in T^{-1}(0)$.

(ii) If, in addition, assumptions in Theorem 2.1 (iii) [respectively, (iv)] hold, then $\{x^k\}_{k\in K_S}$ converges to $\bar{x} \in T^{-1}(0)$ R-linearly [respectively, Q-linearly].

Proof. Since we assume that BM Algorithm 4.1 performs an infinite number of serious steps, we shall consider only indices k in K_S (note that x^k is not being modified for $k \notin K_S$). BM Algorithm 4.1 accepts a serious step when

$$\langle v^k, s^k\rangle > \theta \min\{\|s^k\|, \|v^k\|\}^2 . \tag{8}$$

Having in mind the framework of HPPM Algorithm 2.1, define

$$\mu_k := (\alpha^{l_k} R)^{-1}\|s^k\| \quad \text{and} \quad r^k := s^k - v^k .$$

With this notation, the line-search step of BM Algorithm 4.1 becomes $y^k = x^k - \mu_k^{-1} s^k$, or, equivalently,

$$\begin{aligned} 0 &= s^k + \mu_k(y^k - x^k) \\ &= v^k + \mu_k(y^k - x^k) + r^k , \quad v^k \in T(y^k) . \end{aligned}$$

Thus, condition (3) in HPPM Algorithm 2.1 is satisfied. To check (5), note that

$$\begin{aligned} \|r^k\|^2 &= \|s^k - v^k\|^2 \\ &= \|s^k\|^2 - 2\langle s^k, v^k\rangle + \|v^k\|^2 \\ &< \|s^k\|^2 - 2\theta \min\{\|s^k\|, \|v^k\|\}^2 + \|v^k\|^2 && \text{[by (8)]} \\ &= \max\{(1-2\theta)\|s^k\|^2 + \|v^k\|^2, (1-2\theta)\|v^k\|^2 + \|s^k\|^2\} \\ &\le 2(1-\theta)\max\{\|s^k\|, \|v^k\|\}^2 \\ &= 2(1-\theta)\max\{\mu_k\|y^k - x^k\|, \|v^k\|\}^2 \\ &\le \sigma_k^2 \max\{\mu_k\|y^k - x^k\|, \|v^k\|\}^2 . && \text{[for } \sigma_k \ge \sqrt{2(1-\theta)}\text{]} \end{aligned}$$

We conclude that the pair $(y^k, v^k \in T(y^k))$ yielding a serious step in BM Algorithm 4.1 is acceptable in the "Inexact proximal step" of HPPM Algorithm 2.1, where μ_k and σ_k are as chosen above. In particular, Theorem 2.1(i) implies that $\{x^k\}$ is bounded. Obviously, $\{y^k\}$ is then also bounded. Being a maximal monotone operator with domain $\Re^n$, T is bounded on bounded sets [28, Theorem 1]. Hence, there exists some $M > 0$ such that $\|w^i\| \le M$ for all $(z^i, w^i) \in \Delta_k$ and all $k = 1, 2, \ldots$. Since $s^k \in \operatorname{conv}\{w^i\}$ for all k, it follows that $\|s^k\| \le M$. Finally,

$$\mu_k = (\alpha^{l_k} R)^{-1}\|s^k\| \le (\alpha^{\bar{m}} R)^{-1} M =: \bar{\mu} .$$

Therefore, our choice of μ_k satisfies the assumption of Theorem 2.1(iii) (hence, also of Theorem 2.1(ii)). If we take $\sigma_k \in (\sqrt{2(1-\theta)}, 1)$, this choice also satisfies the condition of Theorem 2.1(ii), because $\theta \in (1/2, 1)$. Hence, all the assertions now follow from Theorem 2.1 (ii)-(iv). ∎

Note that our implementation of the bundle algorithm has no infinite loops, and so it is always well-defined. We have established that if an infinite number of serious steps is generated, the algorithm falls within the general framework of the hybrid projection-proximal point method, and so all the convergence results, including rates of convergence, readily follow. For completeness, it remains to consider the case when K_S is finite, i.e., there is a last serious step $x^{k_{last}}$, followed by an infinite sequence of null steps.

In our development, we shall need the following technical result.

Lemma 4.1 *[15, Lemma IX.2.1.1] Let $\gamma > 0$ be fixed, and consider two infinite sequences $\{v^j\}$ and $\{s^j\}$ such that for all $j = 1, 2, \ldots$, it holds that*

$$\langle v^k - v^j, s^j \rangle \geq \gamma \|s^j\|^2 \quad \textit{for all } j > k\,. \tag{9}$$

If $\{v^j\}$ is bounded, then $\{s^j\} \to 0$ as $j \to \infty$. ■

Theorem 4.2 *Suppose that T is a maximal monotone operator, with $T^{-1}(0) \neq \emptyset$, and consider BM Algorithm 4.1. Assume further that the set K_S is finite and let k_{last} be the index k yielding the last serious step. Then the following hold:*

(i) The sequence $\{s^k\}$ tends to zero.

(ii) The last serious step $x^{k_{last}}$ is an approximate solution of (1) in the following sense. For each k sufficiently large and each s^k (recall that $s^k = \mathrm{conv}\{w^i\}, (z^i, w^i) \in \Delta_k$), the point $\hat{x}^k = \mathrm{conv}\{z^i\}$ generated by the same convex combination coefficients as s^k, satisfies

$$s^k \in T^{\varepsilon_k}(\hat{x}^k)\,, \quad \|\hat{x}^k - x^{k_{last}}\| \leq \mathrm{TOL}/2 \quad \textit{and} \quad \varepsilon_k \leq M\mathrm{TOL}\,,$$

where $M := \max\{\|w^i\| \mid (z^i, w^i) \in \Delta_k, i \text{ such that } \|z^i - x^{k_{last}}\| \leq \mathrm{TOL}/2\}$.

Proof. [(i)] For all $k \geq k_{last}$, the line-search step in BM always results in a null step. The latter means that

$$\langle v^k, s^k \rangle \leq \theta \min\{\|s^k\|, \|v^k\|\}^2 \quad \text{and} \quad l_k = \hat{m}_k \text{ for all } k \geq k_{last}\,. \tag{10}$$

Observe that $\hat{m}_k$ is set to zero after serious steps, while within a sequence of null steps it can only be increased or stay the same. Hence, $\{\hat{m}_k\}$ is nondecreasing for all $k \geq k_{last}$. Since $\hat{m}_k$ is also bounded above by $\bar{m}$ for all k, it follows that for k sufficiently large, say $k \geq k_1$, $\hat{m}_k$ remains fixed. Taking also into account the second relation in (10), we conclude that

$$l_k = \hat{m}_k = \tilde{m} \leq \bar{m} \text{ for all } k \geq k_1\,. \tag{11}$$

In particular, the line-search step generates y^k satisfying

$$\|y^k - x^{k_{last}}\| = \alpha^{\tilde{m}} R \text{ for all } k \geq k_1\,. \tag{12}$$

Since for all $k \geq k_1$ we have that $\hat{m}_k = \tilde{m}$ in the "Compute direction" step, (12) ensures that the pair (y^k, v^k) appears in the selected bundle used to compute s^j for all $j > k > k_1$. Using the optimality conditions for the minimum-norm problem defining s^j, this means that

$$\langle s^j, v^k - s^j \rangle \geq 0 \text{ for all } j > k > k_1 \,.$$

Or, equivalently,

$$\langle v^k, s^j \rangle \geq \|s^j\|^2 \text{ for all } j > k > k_1 \,.$$

Subtracting from the above inequality the first relation in (10) written with $k := j > k_1$, we obtain that

$$\langle v^k - v^j, s^j \rangle \geq (1-\theta)\|s^j\|^2 \text{ for all } j > k > k_1 \,.$$

Writing (12) with $k := j > k_1$, since $x^{k_{last}}$ is fixed we have that $\{y^j\}_{j>k_1}$ is bounded, and hence, so is $\{v^j\}_{j>k_1}$. Using Lemma 4.1, we conclude that $\{s^j\} \to 0$.

[(ii)] Note first that if $\tilde{m} < \bar{m}$ in (11), the "Compute direction" step implies that $\|s^k\| > \alpha^{\tilde{m}}\tau$, which leads to a contradiction with item (i). Hence, $\tilde{m} = \bar{m}$.
Since $s^k \in \operatorname{conv}\{w^i\}$, we have that

$$s^k = \sum_i \lambda_i^k w^i \,, \quad \text{with } \lambda_i^k \geq 0 \text{ and } \sum_i \lambda_i^k = 1 \,.$$

Consider now the corresponding convex combination $\hat{x}^k := \sum_i \lambda_i^k z^i$ and recall that the bundle Δ_k gathers pairs (z^i, w^i) such that $\|z^i - x^{k_{last}}\| \leq \alpha^{\bar{m}} R$. Then Lemma 3.1 gives the final result:

$$s^k \in T^{\varepsilon_k}(\hat{x}^k) \quad \text{with } \|\hat{x}^k - x^{k_{last}}\| \leq \alpha^{\bar{m}} R \leq \text{TOL}/2 \quad \text{and} \quad \varepsilon_k \leq 2\alpha^{\bar{m}} RM \leq \text{TOL} M \,,$$

where $M := \max\{\|w^i\| \mid (z^i, w^i) \in \Delta_k \,, i \text{ such that } \|z^i - x^{k_{last}}\| \leq \alpha^{\bar{m}} R\}$. ∎

As a practical matter, if BM Algorithm 4.1 happens to generate a finite number of serious steps and the approximate solution $x^{k_{last}}$ does not appear satisfactory, then one can decrease the tolerance parameter TOL, increase $\bar{m}$, and proceed. In [8, Lemma 4.2], it is established that if TOL $= 0$ and the number of serious steps is finite, then $x^{k_{last}} \in T^{-1}(0)$, i.e., the exact solution is generated.

5. Concluding Remarks

We have demonstrated that serious steps of a bundle method for solving maximal monotone inclusions can be viewed as iterations of a hybrid proximal point algorithm. This implies that bundling strategies can be used as a constructive and computationally realistic tool to find acceptable approximate solutions of proximal subproblems. Conversely, our development provides an alternative convergence analysis for bundle methods, including the rate of convergence results, by relating them to the proximal point algorithms. We hope that the exhibited relation adds a further insight into the nature of both methodologies.

Finally, we conjecture that the presented approach can be combined with decomposition principles, such as in [25], in order to develop implementable decomposition algorithms, with a potential for parallel implementation.

REFERENCES

1. Ya.I. Alber. On average convergence of the iterative projection methods, 1997. To be published in *Taiwanese Journal of Mathematics*.
2. Ya.I. Alber. Proximal projection method for variational inequalities and Cesàro averaged approximations, 1998. Preprint of the Technion No. 1051. Haifa, Israel.
3. L. Bacaud, C. Lemaréchal, A. Renaud, and C.A. Sagastizábal. Bundle methods in stochastic optimal power management: a disaggregated approach using preconditioners. *Computational Optimization and Applications*, 2001. Accepted for publication.
4. J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, and C. Sagastizábal. A family of variable metric proximal point methods. *Mathematical Programming*, 68:15–47, 1995.
5. J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Optimisation Numérique: aspects théoriques et pratiques*. Springer–Verlag, Berlin, 1997.
6. R.E. Bruck and S. Reich. Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston Journal of Mathematics*, 3:459–470, 1977.
7. R.S. Burachik, A.N. Iusem, and B.F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, 5:159–180, 1997.
8. R.S. Burachik, C.A. Sagastizábal, and B. F. Svaiter. Bundle methods for maximal monotone operators. In R. Tichatschke and M. Théra, editors, *Ill-posed variational problems and regularization techniques*, Lecture Notes in Economics and Mathematical Systems, No. 477, pages 49–64. Springer–Verlag, 1999.
9. R.S. Burachik, C.A. Sagastizábal, and B.F. Svaiter. ε-Enlargements of maximal monotone operators: Theory and applications. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 25–44. Kluwer Academic Publishers, 1999.
10. J.V. Burke and M. Qian. A variable metric proximal point algorithm for monotone operators. *SIAM Journal on Control and Optimization*, 37:353–375, 1998.
11. R. Cominetti. Coupling the proximal point algorithm with approximation methods. *Journal of Optimization Theory and Applications*, 95:581–600, 1997.
12. R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem.* Academic Press, New York, 1992.
13. J. Eckstein. Approximate iterations in Bregman-function-based proximal algorithms. *Mathematical Programming*, 83:113–123, 1998.
14. P.T. Harker and J.-S. Pang. Finite-dimensional variational inequality problems : A survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.
15. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms.* Springer–Verlag, Berlin, 1993. Number 305-306 in Grund. der math. Wiss.
16. A.N. Iusem and L.R. Lucambio Pérez. An extragradient-type method for non-smooth variational inequalities. *Optimization*, 2000. To appear.
17. K.C. Kiwiel. A method for solving certain quadratic programming problems arising in nonsmooth optimization. *IMA Journal of Numerical Analysis*, 6:137–152, 1986.
18. B. Lemaire. The proximal algorithm. In J.-P. Penot, editor, *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics 87*, pages 73–87. Birkhauser, Basel, 1989.

19. C. Lemaréchal, F. Pellegrino, A. Renaud, and C. Sagastizábal. Bundle methods applied to the unit-commitment problem. In J. Doležal and J. Fidler, editors, *System Modelling and Optimization*, pages 395–402. Chapman and Hall, 1995.
20. C. Lemaréchal and C. Sagastizábal. An approach to variable metric bundle methods. In J. Henry and J.-P. Yvon, editors, *Lecture notes in Control and Information Sciences No. 197, System Modelling and Optimization*, pages 144–162. Springer, Berlin, 1994.
21. C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods : From conceptual to implementable forms. *Mathematical Programming*, 76:393–410, 1997.
22. Y. Y. Lin and J.-S. Pang. Iterative methods for large convex quadratic programs: A survey. *SIAM Journal on Control and Optimization*, 25:383–411, 1987.
23. F.J. Luque. Asymptotic convergence analysis of the proximal point algorithm. *SIAM Journal on Control and Optimization*, 22:277–293, 1984.
24. B. Martinet. Regularisation d'inequations variationelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationelle*, 4:154–159, 1970.
25. T. Pennanen. Dualization of generalized equations of maximal monotone type. *SIAM Journal on Optimization*, 10:809–835, 2000.
26. S. Reich. Weak convergence theorems for nonexpansive mappings in Banach spaces. *Journal of Mathematical Analysis and Applications*, 67:274–276, 1979.
27. J.P. Revalski and M. Théra. Variational and extended sums of monotone operators. In R. Tichatschke and M. Théra, editors, *Ill-posed variational problems and regularization techniques*, Lecture Notes in Economics and Mathematical Systems, No. 477, pages 229–246. Springer–Verlag, 1999.
28. R. T. Rockafellar. Local boundedness of nonlinear monotone operators. *Michigan Mathematical Journal*, 16:397–407, 1969.
29. R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.
30. N.Z. Shor. *Minimization methods for nondifferentiable functions*. Springer–Verlag, Berlin, 1985.
31. M. V. Solodov. A class of globally convergent algorithms for pseudomonotone variational inequalities. In M C. Ferris, O.L. Mangasarian, and J.-S. Pang, editors, *Complementarity: Applications, Algorithms and Extensions*. Kluwer Academic Publishers, 2001.
32. M. V. Solodov and B. F. Svaiter. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM Journal on Optimization*, 10:605–625, 2000.
33. M. V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, 34:1814–1830, 1996.
34. M.V. Solodov and B.F. Svaiter. A globally convergent inexact Newton method for systems of monotone equations. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 355–369. Kluwer Academic Publishers, 1999.
35. M.V. Solodov and B.F. Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7:323–345, 1999.
36. M.V. Solodov and B.F. Svaiter. A hybrid projection–proximal point algorithm. *Jour-*

nal of Convex Analysis, 6:59–70, 1999.

37. M.V. Solodov and B.F. Svaiter. Error bounds for proximal point subproblems and associated inexact proximal point algorithms. *Mathematical Programming*, 88:371–389, 2000.
38. M.V. Solodov and B.F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Mathematics of Operations Research*, 25:214–230, 2000.
39. P. Tossings. The perturbed proximal point algorithm and some of its applications. *Applied Mathematics and Optimization*, 29:125–159, 1994.
40. P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431–446, 2000.
41. L. Veselý. Local uniform boundedness principle for families of ε-monotone operators. *Nonlinear Analysis, Theory, Methods and Applications*, 24:1299–1304, 1995.