Optimization Methods and Software
Vol. 19, No. 5, October 2004, pp. 557–575

Taylor & Francis
Taylor & Francis Group

A CLASS OF DECOMPOSITION METHODS FOR CONVEX OPTIMIZATION AND MONOTONE VARIATIONAL INCLUSIONS VIA THE HYBRID INEXACT PROXIMAL POINT FRAMEWORK

M. V. SOLODOV*

Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil

(Received 29 October 2003; In final form 13 February 2004)

Dedicated to Olvi Mangasarian on his 70th birthday

We show that the decomposition method for convex programming proposed by Chen and Teboulle can be regarded as a special case in the hybrid inexact proximal point framework. We further demonstrate that the more general decomposition algorithms for variational inequalities introduced by Tseng are also either a special case in this framework or very closely related to it. This analysis provides a new insight into the nature of those decomposition schemes, as well as paves the way for deriving more practical methods by solving subproblems approximately (for example, using appropriate bundle methods). As a by-product, we also improve some convergence results and extend the approach to a more general class of problems.

Keywords: Maximal monotone operator; Variational inclusion; Decomposition; Enlargement of operator; Hybrid inexact proximal point method; Bundle method

1 INTRODUCTION

Let Z be an Euclidean space and $\mathcal{M}(Z)$ be the set of all subsets of Z. We start our discussion with the classical problem

$$\text{Find } z \in Z \text{ such that } 0 \in T(z), \tag{1}$$

where $T\colon Z \to \mathcal{M}(Z)$ is an (multi-valued) operator on Z. Throughout we assume that T is maximal monotone. A wide variety of important problems, such as convex minimization, monotone variational inequalities over convex sets, certain min–max problems, etc., can be stated in the form of (1). If a given problem, and therefore the associated operator T, has some separable structure, decomposition methods come into play. Many of those methods, e.g.

* E-mail: solodov@impa.br

ISSN 1055-6788 print; ISSN 1029-4937 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/1055678042000218957

Refs. [12,13,18,34,35], are explicitly or implicitly derived from the proximal point algorithm [19,25] for solving (1).

More recently, some new decomposition schemes had been proposed in Refs. [9,37], which, however, did not appear to fit any previously known (proximal point) framework for solving (1) with some appropriate T. Consider, for example, the problem

$$\begin{aligned} &\text{minimize} \quad f_1(x_1) + f_2(x_2) \\ &\text{subject to} \quad Ax_1 - x_2 = 0, \end{aligned} \tag{2}$$

where f_1 and f_2 are closed proper convex functions on Euclidean spaces X_1 and X_2, respectively, and $A\colon X_1 \to X_2$ is a linear operator (a matrix of appropriate dimensions). The method of Chen and Teboulle [9] applies proximal point iterations to the subdifferential of the Lagrangian function $L(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + \langle y, Ax_1 - x_2\rangle$, alternately fixing the variables or the multipliers. Specifically, given some $(x_1^k, x_2^k, y^k) \in X_1 \times X_2 \times X_2$, the exact version of the method performs the following updates:

$$\begin{aligned} \hat{y}^k &= y^k + \alpha_k(Ax_1^k - x_2^k), \\ x_1^{k+1} &= \operatorname*{arg\,min}_{x_1 \in X_1} \left\{ f_1(x_1) + \langle A^\top \hat{y}^k, x_1\rangle + \frac{1}{2\alpha_k}\|x_1 - x_1^k\|^2 \right\}, \\ x_2^{k+1} &= \operatorname*{arg\,min}_{x_2 \in X_2} \left\{ f_2(x_2) - \langle \hat{y}^k, x_2\rangle + \frac{1}{2\alpha_k}\|x_2 - x_2^k\|^2 \right\}, \\ y^{k+1} &= y^k + \alpha_k(Ax_1^{k+1} - x_2^{k+1}), \end{aligned} \tag{3}$$

where $0 < \alpha_k < (2\max\{\|A\|, 1\})^{-1}$.

This method has some nice features not shared by previous decomposition algorithms, when the latter are applied to (2). In particular, the minimization is carried out separately in the spaces X_1 and X_2, and the two minimization problems decompose further according to the separable structure of the functions f_1 and f_2. Other methods do not achieve such a fine degree of decomposition for the given problem, see Ref. [37] for a more detailed discussion.

As is well known [24], the problem (2) is equivalent to (1) if we define

$$\begin{aligned} T(z) &= \partial_{x_1,x_2} L(x_1, x_2, y) \times (-\partial_y L(x_1, x_2, y)), \\ z &= (x_1, x_2, y) \in Z = X_1 \times X_2 \times X_2. \end{aligned} \tag{4}$$

Although clearly related to the proximal point method, the decomposition scheme of Chen and Teboulle did not appear to fit into any general proximal-based framework for problem (1) with T given by (4) (or with any other appropriate choice of T). In Ref. [37], Tseng extended this decomposition scheme to more general problems (see Section 3.2), and interpreted it as an 'alternating projection-proximal method'. This, however, still did not fit any proximal-based framework for problem (1) that existed at the time.

In this article, we show that the decomposition method of Chen and Teboulle is a special case in the framework of inexact hybrid proximal point methods developed in Refs. [27,28,31]. The same assertion essentially applies to the methods of Tseng (although depending on the problem structure, one part of his iteration can be slightly different, see Section 3.2). Apart from providing a new unified insight into those methods, this would allow us to extend them to a class of more general problems, improve some convergence results and suggest a practical approximation criteria for solving the subproblems.

We proceed with describing the inexact hybrid proximal point framework. Given some $z^k \in Z$ and having chosen the regularization parameter $\alpha_k > 0$, the *exact* proximal iteration [17,19,25] for (1) amounts to solving the following system: find $z^{k+1} \in Z$ and $v^{k+1} \in Z$ such that

$$\begin{cases} v^{k+1} \in T(z^{k+1}), \\ 0 = \alpha_k v^{k+1} + z^{k+1} - z^k. \end{cases}$$

Then z^{k+1} is declared the next iterate. To handle approximate solutions, it is useful to relax *both* the inclusion and the equation in the above system [for some motivation as to why it is important to relax both parts of the subproblem and not just one, see Ref. 29 and Section 3.1 below]. To relax the inclusion, the following notion is fundamental: for $\varepsilon \geq 0$, the ε-enlargement of T at $z \in Z$ is given by

$$T^\varepsilon(z) = \{v \in Z \mid \langle v - v', z - z' \rangle \geq -\varepsilon \text{ for all } z' \in Z \text{ and } v' \in T(z')\}.$$

It holds that for any $z \in Z$ and $\varepsilon \geq 0$, $T(z) \subset T^\varepsilon(z)$ and $T^0(Z) = T(z)$. If f is a proper closed convex function, then $\partial_\varepsilon f(z) \subset (\partial f)^\varepsilon(z)$, where $\partial_\varepsilon f$ is the usual ε-subdifferential of f. For other properties and applications of ε-enlargements of maximal monotone operators, see Refs. [5,7,8,26,29].

The following iterative procedure will be referred to as the hybrid inexact proximal point method (HIPPM): Given z^k and having chosen the regularization parameter $\alpha_k > 0$ and the error tolerance parameter $\sigma_k \in [0, 1)$, find $\hat{z}^k \in Z$, $\hat{v}^k \in Z$ and $\varepsilon_k \geq 0$ such that

$$\begin{cases} \hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k), \\ \delta^k = \alpha_k \hat{v}^k + \hat{z}^k - z^k, \end{cases} \tag{5}$$

and

$$\|\delta^k\|^2 + 2\alpha_k \varepsilon_k \leq \sigma_k(\|\alpha_k \hat{v}^k\|^2 + \|\hat{z}^k - z^k\|^2). \tag{6}$$

To obtain the next iterate, set

$$z^{k+1} = z^k - \tau_k \alpha_k \hat{v}^k, \tag{7}$$

where

$$\tau_k = \theta_k \frac{\langle \hat{v}^k, z^k - \hat{z}^k \rangle - \varepsilon_k}{\alpha_k \|\hat{v}^k\|^2}, \quad \theta_k \in (0, 2). \tag{8}$$

For future reference, we note that under the conditions (5) and (6), it holds that [see Ref. 31, Proposition 4]

$$\exists \theta_k \in (0, 2) \text{ such that } \tau_k = 1.$$

In particular,

$$z^{k+1} = z^k - \alpha_k \hat{v}^k$$

is a special case of (7). The described version of HIPPM had been introduced in Ref. [31]. The HIPPM has all the desirable convergence properties of the exact proximal point algorithm (if the solution set $T^{-1}(0) \neq \emptyset$, the sequence $\{z^k\}$ converges to some $\bar{z} \in T^{-1}(0)$ for any starting point $z^0 \in Z$; if T^{-1} is Lipschitz-continuous at zero then the rate of convergence is linear). At the same time, in HIPPM the approximation criterion for solving the proximal subproblems is considerably more constructive and more suited for practical use when compared to the

classical (summability-type) approximation conditions [see Refs. 27,29,30,32 for detailed discussions of this issue and for some applications where this feature of HIPPM is useful; see also Section 3.1].

To motivate and set the stage for further developments, we next show that the decomposition method (3) of Chen and Teboulle falls within the framework of HIPPM. The idea is to consider the point $\hat{z}^k = (x_1^{k+1}, x_2^{k+1}, \hat{y}^k) \in Z = X_1 \times X_2 \times X_2$ as an approximate solution, in the sense of (5) and (6), of the proximal point iteration applied to T defined in (4). To establish the claim, we have to exhibit the associated $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, verify that (6) holds and show that (7) can produce z^{k+1} coinciding with $(x_1^{k+1}, x_2^{k+1}, y^{k+1})$ given by (3).

By the optimality conditions for the two minimization problems in (3), we have that

$$\begin{aligned} r_1^k &:= 0 = \alpha_k \hat{u}_1^k + x_1^{k+1} - x_1^k, \quad \hat{u}_1^k \in \partial f_1(x_1^{k+1}) + A^\top \hat{y}^k = \partial_{x_1} L(x_1^{k+1}, x_2^{k+1}, \hat{y}^k), \\ r_2^k &:= 0 = \alpha_k \hat{u}_2^k + x_2^{k+1} - x_2^k, \quad \hat{u}_1^k \in \partial f_1(x_1^{k+1}) - \hat{y}^k = \partial_{x_2} L(x_1^{k+1}, x_2^{k+1}, \hat{y}^k). \end{aligned} \tag{9}$$

Furthermore, by the first relation in (3), we have that

$$\begin{aligned} -\alpha_k(Ax_1^{k+1} - x_2^{k+1}) + \hat{y}^k - y^k &= \alpha_k A(x_1^k - x_1^{k+1}) + \alpha_k(x_2^{k+1} - x_2^k) =: s^k, \\ \hat{w}^k &:= -(Ax_1^{k+1} - x_2^{k+1}) \in (-\partial_y L(x_1^{k+1}, x_2^{k+1}, \hat{y}^k)). \end{aligned} \tag{10}$$

Defining $\hat{v}^k = (\hat{u}_1^k, \hat{u}_2^k, \hat{w}^k) \in Z$, we have that $\hat{v}^k \in T(\hat{z}^k)$. Combining further (9) and (10), we conclude that (5) holds with $\varepsilon_k = 0$ and $\delta^k = (r_1^k, r_2^k, s^k) = (0, 0, s^k)$.

Taking $\tau_k = 1$ in the update formula (7) and using again (9) and (10), we have that

$$\begin{aligned} z^{k+1} &= (x_1^k - \alpha_k \hat{u}_1^k, x_2^k - \alpha_k \hat{u}_2^k, y^k - \alpha_k \hat{w}^k) \\ &= (x_1^{k+1}, x_2^{k+1}, y^{k+1}), \end{aligned}$$

where the right-hand side is the same as the objects in (3).

It remains to verify that the error tolerance condition (6) holds for appropriate choices of α_k, so that convergence is guaranteed. To this end, observe that

$$\begin{aligned} \|\delta^k\| = \|s^k\| &= \alpha_k \|A(x_1^k - x_1^{k+1}) + x_2^{k+1} - x_2^k\| \\ &\le 2\alpha_k \max\{\|A\|, 1\} \max\{\|x_1^{k+1} - x_1^k\|, \|x_2^{k+1} - x_2^k\|\} \\ &\le \sqrt{\alpha_k} \|\hat{z}^k - z^k\|, \end{aligned}$$

if we set $\alpha_k \le \sqrt{\sigma_k}(2\max\{\|A\|, 1\})^{-1}$, and take into account the definition of $\hat{z}^k$ and the monotonicity of the norm. Obviously, the above (rather conservative!) estimate implies that (6) is satisfied.

We have thus demonstrated that the method of Chen and Teboulle is a special case in the family of HIPPM. In Section 3.1, we shall further use this relation to set up a constructive rule for solving the two optimization problems in (3) approximately. The latter is important, as solving those problems exactly is in most cases impractical if not impossible. This is particularly true if f_1 and/or f_2 is nonsmooth, in which case bundle methods [4, Chap. 9] have to be used. We shall argue that the approximation criterion which we shall propose is well suited for coupling the decomposition scheme with bundle methods.

We note that HIPPM had proved to be a useful framework in a number of other works related to splitting and decomposition methods. For example, the modified forward–backward splitting of Ref. [38] can be derived from HIPPM, see Ref. [27]. The way HIPPM can be combined with the method of partial inverses [34] has been studied in Refs. [6,20]. To conclude this section,

we cite some further literature related to proximal-based decomposition, in addition to what has been already mentioned above: Refs. [1,2,15,22].

We next describe our notation. Given a space Z, $\mathcal{M}(Z)$ denotes the set of all subsets of Z. For an operator $T\colon Z \to \mathcal{M}(Z)$, $\operatorname{dom} T = \{z \in Z | T(z) \neq \emptyset\}$ is its domain, and for $v \in Z$, $T^{-1}(v) = \{z \in Z | T(z) = v\}$. By I we shall denote the identity operator (in a given space). By $\langle \cdot, \cdot \rangle$ we denote the inner product and by $\| \cdot \|$ the associated norm, where the space would always be clear from the context. For a proper closed convex function $f\colon Z \to \Re \cup \{+\infty\}$ and $\varepsilon \geq 0$, $\partial_\varepsilon f(z) = \{v \in Z | f(z') \geq f(z) + \langle v, z' - z \rangle - \varepsilon \ \forall z' \in Z\}$ is the ε-subdifferential mapping, and $\partial f = \partial_0 f$ is the usual subdifferential. For a convex set $C \subset Z$, $\operatorname{ri} C$ denotes its relative interior. For a closed convex set $C \subset Z$, $P_C(z) = \arg\min_{z' \in Z} \|z' - z\|$ is the orthogonal projection operator onto C, and $\operatorname{dist}(z, C) = \|z - P_C(z)\|$ is the distance from $z \in Z$ to C.

2 THE GENERAL DECOMPOSITION SCHEME

Consider problem (1) with

$$T(z) = F(x, y) \times (G(x, y) + H(y)), \quad z = (x, y) \in Z = X \times Y, \tag{11}$$

where X and Y are some Euclidean spaces, $F\colon X \times Y \to \mathcal{M}(X)$, $G\colon X \times Y \to Y$ and $H\colon Y \to \mathcal{M}(Y)$. We make the following standing assumptions:

(A1) G is (single-valued) continuous on $X \times Y$;
(A2) H is maximal monotone;
(A3) The mapping $(x, y) \to \mathcal{M}(F(x, y) \times G(x, y))$ is maximal monotone;
(A4) $\operatorname{dom} H \subset \operatorname{ri}\{y \in Y | \exists x \in X \text{ s.t. } F(x, y) \times G(x, y) \neq \emptyset\}$.

Under the stated assumptions, it follows from Ref. [23] that T is maximal monotone and further that the mapping $x \to F(x, y)$ is also maximal monotone for any fixed $y \in \operatorname{dom} H$ [37, Lemma 2.1].

We next describe our decomposition algorithm. Let $(x^k, y^k) \in Z$ be the current iterate. We first choose a continuous monotone function $G_1^k\colon Y \to Y$. We can choose $G_1^k \equiv 0$, but in general the choice depends on the structure of $G(x^k, \cdot)$ and $H(\cdot)$. The chosen G_1^k is employed for function-splitting of the operator $G(x^k, \cdot) + H(\cdot)$, which is computationally useful in some applications. In particular, we pass to the representation $G(x^k, \cdot) + H(\cdot) = (G(x^k, \cdot) - G_1^k(\cdot)) + (G_1^k(\cdot) + H(\cdot))$, and perform a forward–backward splitting step [10,18,21,35] for this representation of the operator, with the x-part fixed. The resulting $\hat{y}^k \in Y$ is used in the approximate proximal point step for the operator $F(\cdot, \hat{y}^k)$, with the y-part fixed. The iteration finishes with the update of (x^k, y^k), which makes use of the objects computed during the above steps (and which is derived from HIPPM). To simplify the statement of the algorithm, we shall leave out until a later discussion the detail of the choice of the regularization parameter α_k. Essentially, this parameter has to be sufficiently small (recall the method of Chen and Teboulle). Its value can either be determined by a suitable linesearch procedure or be set according to some heuristic considerations. We proceed to state the method.

ALGORITHM 2.1 Hybrid proximal decomposition method (HPDM) *Choose* $(x^0, y^0) \in X \times Y$, $0 < \underline{\sigma} \leq \bar{\sigma} < 1$, $0 < \underline{\theta} \leq \bar{\theta} < 2$. *Set* $k := 0$.

`Forward–Backward Splitting Step`. *Choose a continuous monotone function* $G_1^k\colon Y \to Y$ *and* $\alpha_k > 0$. *Compute* $\hat{y}^k \in Y$ *by*

$$\hat{y}^k = (I + \alpha_k(H(\cdot) + G_1^k(\cdot)))^{-1}((I - \alpha_k(G(x^k, \cdot) - G_1^k(\cdot)))y^k), \tag{12}$$

and define $\hat{h}^k$ *to be the element of* $H(\hat{y}^k)$ *computed within* (12).

`Inexact Proximal Step`. *Choose the error tolerance parameter* $\sigma_k \in [\underline{\sigma}, \bar{\sigma}]$.

Compute $\hat{x}^k \in X$, $\hat{u}^k \in X$ *and* $\varepsilon_k \geq 0$ *such that*

$$\begin{cases} \hat{u}^k \in F^{\varepsilon_k}(\hat{x}^k, \hat{y}^k), \\ r^k = \alpha_k \hat{u}^k + \hat{x}^k - x^k, \end{cases} \tag{13}$$

where $F^{\varepsilon_k}(\hat{x}^k, \hat{y}^k)$ *is the* ε_k*-enlargement of* $F(\cdot, \hat{y}^k)$ *at* $\hat{x}^k$, *with the* y*-part fixed, and*

$$\|r^k\|^2 + \|s^k\|^2 + 2\alpha_k\varepsilon_k \leq \sigma_k(\|\alpha_k\hat{u}^k\|^2 + \|\alpha_k\hat{w}^k\|^2 + \|\hat{x}^k - x^k\|^2 + \|\hat{y}^k - y^k\|^2), \tag{14}$$

with

$$\hat{w}^k := G(\hat{x}^k, \hat{y}^k) + \hat{h}^k, \quad s^k := \alpha_k\hat{w}^k + \hat{y}^k - y^k.$$

Once (14) *is satisfied, go to* `Iterates Update` (*if the proximal subproblem is solved to 'maximal' possible precision, but* (14) *is still not satisfied, decrease* α_k, *compute a new* $\hat{y}^k$ *by* (12) *and repeat the* `Inexact Proximal Step` *with the new* $\hat{y}^k$).

`Iterates Update`. *Stop if* $\hat{x}^k = x^k$ *and* $\hat{y}^k = y^k$. *Otherwise, define*

$$x^{k+1} = x^k - \tau_k\alpha_k\hat{u}^k, \tag{15}$$

$$y^{k+1} = y^k - \tau_k\alpha_k\hat{w}^k, \tag{16}$$

where

$$\tau_k = \theta_k \frac{\langle \hat{u}^k, x^k - \hat{x}^k\rangle + \langle \hat{w}^k, y^k - \hat{y}^k\rangle - \varepsilon_k}{\alpha_k(\|\hat{u}^k\|^2 + \|\hat{w}^k\|^2)}, \quad \theta_k \in [\underline{\theta}, \bar{\theta}]. \tag{17}$$

Set $k := k + 1$ *and go to* `Forward–Backward Splitting Step`.

Apart from our ability to satisfy condition (14), which would require a proof, the other parts of the method are easily seen to be well defined. Indeed, since G_1^k is a monotone continuous function on Y, it follows that $H + G_1^k$ is maximal monotone. Thus $\hat{y}^k$ in (12) is well defined, and further $\hat{y}^k \in \operatorname{dom} H$. As already noted above, for any $\hat{y}^k \in \operatorname{dom} H$, the mapping $x \to F(x, \hat{y}^k)$ is maximal monotone under the stated assumptions. Thus, the proximal point subproblem (13) is well posed and can be solved to any degree of precision, at least in theory (note that the exact solution corresponds to setting $r^k = 0$ and $\varepsilon_k = 0$). Furthermore, the stepsize choice (17) is well defined whenever $\|\hat{u}^k\| + \|\hat{w}^k\| \neq 0$. Now, if it were the case that $\hat{u}^k = 0$ and $\hat{w}^k = 0$, then it would follow that $r^k = \hat{x}^k - x^k$ and $s^k = \hat{y}^k - y^k$. But by (14), $\|r^k\|^2 + \|s^k\|^2 \leq \sigma_k(\|\hat{x}^k - x^k\|^2 + \|\hat{y}^k - y^k\|^2)$, where $\sigma_k \in [0, 1)$. From the latter inequality we conclude that in this case $\hat{x}^k = x^k$ and $\hat{y}^k = y^k$, so that the stopping rule would have been activated (as will be shown below, in this case (x^k, y^k) is a solution of the problem).

To clarify the nature of (12) and some options concerning the choice of G_k^1, suppose that H is the normal cone mapping associated to a closed convex set $C \subset Y$, i.e.

$$H(y) = N_C(y) = \begin{cases} h \in Y \text{ s.t } \langle h, y' - y\rangle \leq 0 \quad \forall y' \in C, & \text{if } y \in C, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{18}$$

In that case, (12) gives

$$\hat{y}^k = P_C(y^k - \alpha_k(G_1^k(\hat{y}^k) - G_1^k(y^k) + G(x^k, y^k))).$$

If we take $G_1^k \equiv 0$, then

$$\hat{y}^k = P_C(y^k - \alpha_k G(x^k, y^k)),$$

which is the standard projection step. If we take $G_1^k \equiv G(x^k, \cdot)$, then

$$\hat{y}^k = P_C(y^k - \alpha_k G(x^k, \hat{y}^k)),$$

which is an implicit (proximal) step. Obviously, the first choice requires the least amount of computational work to obtain $\hat{y}^k$ (for C with some special structure, e.g. the nonnegative orthant, an explicit formula can be used), while the second choice requires the most work. Inbetween there are various intermediate choices.

Which particular G_k^1 should be used depends on the structure of $G(x^k, \cdot)$ and of $H(\cdot)$, see Ref. [37] for a more detailed discussion.

Before proceeding with the convergence analysis, we make one final assumption:

(A5) It holds that

$$\left.\begin{array}{l} u \in F^\varepsilon(x, y) \\ w \in G(x, y) + H(y) \end{array}\right\} \Rightarrow (u, w) = v \in T^\varepsilon(z), \quad z = (x, y),$$

where $F^\varepsilon(x, y)$ is the ε-enlargement of $F(\cdot, y)$ at x with the y-part fixed.

We note that this assumption is redundant if in HPDM we set $\varepsilon_k = 0$ for all k. However, as will be argued in Section 3.1, having $\varepsilon_k \neq 0$ is useful, for example, for coupling the decomposition scheme with appropriate bundle methods for solving the subproblems. With $\varepsilon \neq 0$, the implication stated in the assumption above is not true in general. But it does hold for some important special cases – in particular those with the primal–dual structure, including the problem (2) of Chen and Teboulle. We shall show this for the more general variational inequality

Find $x \in D$ such that $\exists \phi \in \Phi(x)$ with $\langle \phi, x' - x \rangle \geq 0 \quad \forall\, x' \in D$,

$$D = \{x \in X | Bx = b\},$$

where $\Phi\colon X \to \mathcal{M}(X)$ is maximal monotone, $B\colon X \to Y$ is a linear operator (a matrix of appropriate dimensions) and $b \in Y$. This problem is equivalent to (1) with T given by (11), if we choose

$$F(x, y) = \Phi(x) + B^\top y, \quad G(x, y) = b - Bx, \quad H(y) = 0.$$

If $u \in F^\varepsilon(x, y)$ in the sense of Assumption (A5), then

$$\langle u - (\phi' + B^\top y), x - x' \rangle \geq -\varepsilon \quad \forall x' \in X, \quad \phi' \in \Phi(x').$$

Take $w \in G(x, y)$ and any $(x', y') \in Z$, $(u', w') \in T(x', y')$. We have that

$$\begin{aligned} &\langle (u, w) - (u', w'), (x, y) - (x', y') \rangle \\ &\quad = \langle u - (\phi' + B^\top y'), x - x' \rangle + \langle b - Bx - (b - Bx'), y - y' \rangle \\ &\quad = \langle u - (\phi' + B^\top y), x - x' \rangle + \langle B^\top (y - y'), x - x' \rangle \\ &\qquad + \langle B(x' - x), y - y' \rangle \\ &\quad = \langle u - (\phi' + B^\top y), x - x' \rangle \geq -\varepsilon, \end{aligned}$$

which verifies that Assumption (A5) holds.

To prove that HPDM converges, we shall show that for all α_k small enough, a sufficiently good approximate solution for the proximal subproblem (13) would guarantee that the condition (14) is satisfied. Convergence then would follow from relating HPDM to HIPPM and following the line of analysis for the latter. Note that as a by-product, we shall obtain a new linear rate of convergence result for decomposition methods of this type. Specifically, for the method of Ref. [37] the rate of convergence had been established for the case of the normal cone (18) with $C = Y$ only (so that $H(y) = \{0\}$). We obtain linear convergence for the case of (18) with a general C, and also for the case where H is not necessarily a normal cone but can be a general maximal monotone operator. We comment, however, that in the case of the normal cone with $C \neq Y$, our decomposition method does not include the method of Ref. [37] (the methods are slightly different is that case, see Section 3.2).

THEOREM 2.1 *Suppose that* $T^{-1}(0) \neq \emptyset$ *and that Assumptions* (A1)–(A4) *hold. Suppose further that either in Algorithm* 2.1 *we set* $\varepsilon_k = 0$ *for all* k, *or Assumption* (A5) *holds.*

If G *is Lipschitz-continuous on* $X \times Y$ *and* G_1^k *is* (*uniformly in* k) *Lipschitz-continuous on* Y, *then there exists* $\bar{\alpha} > 0$ *such that for* $\{\alpha_k\}$ *satisfying*

$$0 < \underline{\alpha} \leq \liminf_{k\to\infty} \alpha_k \leq \limsup_{k\to\infty} \alpha_k < \bar{\alpha},$$

the sequence $\{(x^k, y^k)\}$ *generated by Algorithm* 2.1 *is well defined and converges to an element of* $T^{-1}(0)$.

Moreover, if there exist $c_1 > 0$ *and* $c_2 > 0$ *such that*

$$\operatorname{dist}(z, T^{-1}(0)) \leq c_1 \min_{v\in T(z)} \|v\| \quad \forall z \in \{z' \in \operatorname{dom} T \mid \min_{v\in T(z')} \|v\| \leq c_2\}, \tag{19}$$

then the rate of convergence is linear.

Proof Let $k \geq 0$ be any iteration index. As is easily seen, (12) is equivalent to

$$\hat{y}^k + \alpha_k(\hat{h}^k + G_1^k(\hat{y}^k)) = y^k - \alpha_k(G(x^k, y^k) - G_1^k(y^k)), \quad \hat{h}^k \in H(\hat{y}^k), \tag{20}$$

from which we have that

$$\hat{y}^k - y^k = \alpha_k(G_1^k(y^k) - G_1^k(\hat{y}^k) - G(x^k, y^k) - \hat{h}^k).$$

We then further obtain

$$\begin{aligned} s^k &= \alpha_k(G(\hat{x}^k, \hat{y}^k) + \hat{h}^k) + \hat{y}^k - y^k \\ &= \alpha_k(G_1^k(y^k) - G_1^k(\hat{y}^k) + G(\hat{x}^k, \hat{y}^k) - G(x^k, y^k)). \end{aligned}$$

By the Lipschitz-continuity of G and G_1^k (with some modulus $L > 0$), it follows that

$$\begin{aligned} \|s^k\| &\leq \alpha_k L(\|(\hat{x}^k, \hat{y}^k) - (x^k, y^k)\| + \|\hat{y}^k - y^k\|) \\ &\leq 2L\alpha_k\|(\hat{x}^k, \hat{y}^k) - (x^k, y^k)\|, \end{aligned}$$

where the monotonicity of the norm was used in the last relation. Hence,

$$\|s^k\|^2 \leq \sigma_k(\|\hat{x}^k - x^k\|^2 + \|\hat{y}^k - y^k\|^2) \tag{21}$$

whenever

$$\alpha_k \leq \frac{\sqrt{\sigma_k}}{2L}. \tag{22}$$

The approximation condition (14) can be re-written as

$$\begin{aligned}\|r^k\|^2 + 2\alpha_k\varepsilon_k \leq \sigma_k(\|\alpha_k\hat{u}^k\|^2 + \|\alpha_k\hat{w}^k\|^2) \\ + (\sigma_k(\|\hat{x}^k - x^k\|^2 + \|\hat{y}^k - y^k\|^2) - \|s^k\|^2).\end{aligned} \tag{23}$$

Under (22), relation (21) holds and implies that the second term in the right-hand side of (23) is nonnegative. Hence, at the very least the exact solution of the proximal system (13) would satisfy (23) and thus (14) (for the exact solution, the left-hand side in (23) is zero). This completes the proof of the fact that the algorithm is well-defined. (In fact, if the inequality in (22) is strict then the right-hand side of (23) is guaranteed to be positive (unless $\hat{y}^k = y^k$ and $\hat{x}^k = x^k$), and then appropriate inexact solutions of the proximal system (13) with r^k and ε_k small enough would also satisfy (23).)

Suppose that for some k the stopping test is satisfied: $\hat{x}^k = x^k$ and $\hat{y}^k = y^k$. Then we have that $r^k = \alpha_k\hat{u}^k$ and $s^k = \alpha_k\hat{w}^k$. Hence, using (14), we deduce that $\hat{u}^k = 0$, $\hat{w}^k = 0$ and $\varepsilon_k = 0$. By (13), it immediately follows that $0 \in F(x^k, y^k)$. In addition, using (20), we also have that $0 \in G(x^k, y^k) + H(y^k)$. Thus $(x^k, y^k) = z^k \in T^{-1}(0)$, i.e. we terminate at a solution of the problem.

Suppose now that the method does not terminate and an infinite sequence $\{z^k\}$ is generated, $z^k = (x^k, y^k) \in Z$. Define $\hat{z}^k = (\hat{x}^k, \hat{y}^k) \in Z$, $\hat{v}^k = (\hat{u}^k, \hat{w}^k) \in Z$, $\delta^k = (r^k, s^k) \in Z$. By construction, $\hat{u}^k \in F^{\varepsilon_k}(\hat{x}^k, \hat{y}^k)$ and $\hat{w}^k \in G(\hat{x}^k, \hat{y}^k) + H(\hat{y}^k)$. By Assumption (A5), $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, with the inclusion being automatic if $\varepsilon_k = 0$. Combining this with condition (14), we conclude that all the relations in (5) and (6) hold.

The rest of the proof mostly follows the analysis of convergence of HIPPM. We include a streamlined proof for completeness (also, the condition (19) that we use here to obtain the linear rate of convergence result does not imply the uniqueness of the solution, unlike the previous analysis for HIPPM). Using (5), by re-arranging the terms in (6), it is easy to see that the latter condition is equivalent to

$$\langle\hat{v}^k, z^k - \hat{z}^k\rangle - \varepsilon_k \geq \frac{1-\sigma_k}{2\alpha_k}(\|\alpha_k\hat{v}^k\|^2 + \|\hat{z}^k - z^k\|^2). \tag{24}$$

If the right-hand side in (24) is zero, then (5) and (6) imply that $z^k \in T^{-1}(0)$, and in particular, the method would have stopped. We consider therefore the case when the right-hand side in (24) is positive. In that case,

$$z^k \notin S_k := \{z \in Z | \langle\hat{v}^k, z - \hat{z}^k\rangle - \varepsilon_k \leq 0\}.$$

On the other hand, for any $\bar{z} \in T^{-1}(0)$, it holds that $\langle\hat{v}^k - 0, \hat{z}^k - \bar{z}\rangle \geq -\varepsilon_k$ (since $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$), which means that $\bar{z} \in S_k$. By the basic properties of the projection operator (onto the closed half-space S_k), we have that

$$\begin{gathered}P_{S_k}(z^k) = z^k - \frac{\langle\hat{v}^k, z^k - \hat{z}^k\rangle - \varepsilon_k}{\|\hat{v}^k\|^2}\hat{v}^k = z^k - \frac{\tau_k\alpha_k}{\theta_k}\hat{v}^k, \\ \langle\bar{z} - P_{S_k}(z^k), \hat{v}^k\rangle \leq 0.\end{gathered}$$

Using those relations, we obtain that

$$\begin{aligned}\|z^{k+1}-\bar{z}\|^2 &= \|z^k-\bar{z}\|^2+\|z^{k+1}-z^k\|^2+2\langle z^{k+1}-z^k, z^k-\bar{z}\rangle\\ &= \|z^k-\bar{z}\|^2+\|\tau_k\alpha_k\hat{v}^k\|^2-2\tau_k\alpha_k\langle\hat{v}^k, z^k-P_{S_k}(z^k)\rangle\\ &\quad +2\tau_k\alpha_k\langle\hat{v}^k,\bar{z}-P_{S_k}(z^k)\rangle\\ &\le \|z^k-\bar{z}\|^2+\left(1-\frac{2}{\theta_k}\right)\|\tau_k\alpha_k\hat{v}^k\|^2.\end{aligned}$$

Furthermore,

$$\begin{aligned}\|\tau_k\alpha_k\hat{v}^k\| &= \theta_k\frac{\langle\hat{v}^k, z^k-\hat{z}^k\rangle-\varepsilon_k}{\|\hat{v}^k\|} \ge \frac{\theta_k(1-\sigma_k)\|\alpha_k\hat{v}^k\|^2}{2\alpha_k\|\hat{v}^k\|}\\ &= \frac{\theta_k(1-\sigma_k)\alpha_k\|\hat{v}^k\|}{2}\end{aligned}$$

where the inequality is by (24). Taking into account the choices of σ_k, θ_k and α_k, we then conclude that

$$\|z^{k+1}-\bar{z}\|^2 \le \|z^k-\bar{z}\|^2-c_3\|\hat{v}^k\|^2, \tag{25}$$

where $c_3 := (2/\bar{\theta}-1)(\underline{\theta}(1-\bar{\sigma})\underline{\alpha}/2)^2 > 0$. Hence, the sequence $\{\|z^k-\bar{z}\|\}$ is nonincreasing and convergent. [In fact, we have that the sequence $\{z^k\}$ is Fejér-monotone with respect to the solution set $T^{-1}(0)$, and convergence can be claimed by invoking certain facts about Fejér-monotone sequences, e.g. Refs. 3,11 and references therein. But we shall provide details for completeness.]

It follows from (25) that the sequence $\{z^k\}$ is bounded and further

$$0 = \lim_{k\to\infty}\|\hat{v}^k\|. \tag{26}$$

Recalling (22), define

$$\bar{\alpha} := \frac{\sqrt{\sigma}}{(2L)}, \tag{27}$$

and assume that $\alpha_k \le \bar{\alpha}$ for all k sufficiently large (so that (22) holds). Since by (24) we have that

$$\|\hat{v}^k\| \ge \frac{(1-\sigma_k)\|z^k-\hat{z}^k\|}{(2\alpha_k)} \ge \frac{(1-\bar{\sigma})\|z^k-\hat{z}^k\|}{(2\bar{\alpha})}, \tag{28}$$

it follows that

$$0 = \lim_{k\to\infty}(z^k-\hat{z}^k). \tag{29}$$

Then (24) further implies that

$$0 = \lim_{k\to\infty}\varepsilon_k. \tag{30}$$

Let z^* be any accumulation point of the bounded sequence $\{z^k\}$, and let $\{z^{k_j}\}$ be the subsequence converging to z^*. By (29), it follows that $\{\hat{z}^{k_j}\}$ also converges to z^*. Take any $z \in Z$ and $v \in T(z)$. Since $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, we have that

$$\langle\hat{v}^{k_j}-v, \hat{z}^{k_j}-z\rangle \ge -\varepsilon_{k_j}.$$

Passing onto the limit in the above relation as $j \to \infty$ and using (26) and (30), we obtain that

$$\langle 0 - v, z^* - z \rangle \geq 0.$$

The maximal monotonicity of T now implies that $0 \in T(z^*)$. We have thus established that every accumulation point of $\{z^k\}$ is a solution of the problem. We can therefore take $\bar{z} = z^* \in T^{-1}(0)$ in (25). The sequence $\{\|z^k - \bar{z}\|\}$ converges and it has a subsequence converging to zero (since $\bar{z} = z^*$ is an accumulation point of $\{z^k\}$). Thus $\{\|z^k - \bar{z}\|\}$ converges to zero, i.e. $\{z^k\}$ converges to $\bar{z} \in T^{-1}(0)$.

We proceed to establish the linear rate of convergence under the assumption (19). Define $\xi^k \in Z$ and $\psi^k \in Z$ as the exact solution of the proximal subproblem $0 \in \alpha_k T(z) + z - z^k$, that is

$$0 = \alpha_k \psi^k + \xi^k - z^k, \quad \psi^k \in T(\xi^k).$$

Since $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, by Ref. [29, Corollary 2.1], it holds that

$$\|\hat{z}^k - \xi^k\|^2 + \alpha_k^2 \|\hat{v}^k - \psi^k\|^2 \leq \|\alpha_k \hat{v}^k + \hat{z}^k - z^k\|^2 + 2\alpha_k \varepsilon_k.$$

Using further (6), we have that

$$\|\hat{z}^k - \xi^k\|^2 \leq \sigma_k (\|\alpha_k \hat{v}^k\|^2 + \|\hat{z}^k - z^k\|^2).$$

By (24), $\|\hat{z}^k - z^k\| \geq (1 - \sigma_k)\alpha_k \|\hat{v}^k\|/2$. Hence,

$$\|\hat{z}^k - \xi^k\| \leq \sigma_k \left(1 + \frac{4}{(1 - \sigma_k)^2}\right) \|\hat{z}^k - z^k\| \leq c_4 \|\hat{z}^k - z^k\|,$$

where $c_4 := \bar{\sigma}(1 + 4/(1 - \underline{\sigma})^2) > 0$. We further have that

$$\begin{aligned} \alpha_k \|\psi^k\| &= \|\xi^k - z^k\| \\ &\leq \|\hat{z}^k - \xi_k\| + \|\hat{z}^k - z^k\| \\ &\leq (1 + c_4)\|\hat{z}^k - z_k\|. \end{aligned} \tag{31}$$

By (29), it follows that $\psi^k \to 0$. Hence, by (19), for all k sufficiently large it holds that

$$\operatorname{dist}(\xi^k, T^{-1}(0)) \leq c_1 \|\psi^k\| = \frac{c_1 \|\xi^k - z^k\|}{\alpha_k} \leq \frac{c_1 \|\xi^k - z^k\|}{\underline{\alpha}}. \tag{32}$$

Let $\bar{\xi}^k = P_{T^{-1}(0)}(\xi^k)$. Then

$$\begin{aligned} \operatorname{dist}(z^k, T^{-1}(0)) &\leq \|z^k - \bar{\xi}^k\| \\ &\leq \operatorname{dist}(\xi^k, T^{-1}(0)) + \|\xi^k - z^k\| \\ &\leq \left(1 + \frac{c_1}{\underline{\alpha}}\right) \|\xi^k - z^k\| \\ &\leq \left(1 + \frac{c_1}{\underline{\alpha}}\right) (1 + c_4) \|\hat{z}^k - z^k\|, \end{aligned} \tag{33}$$

where the third inequality is by (32) and the last is by (31). Setting now $\bar{z}^k = P_{T^{-1}}(0)(z^k)$, from (25) we obtain that

$$\begin{aligned}\operatorname{dist}(z^{k+1}, T^{-1}(0))^2 &\le \|z^{k+1} - \bar{z}^k\|^2 \\ &\le \operatorname{dist}(z^k, T^{-1}(0))^2 - c_3\|\hat{v}^k\|^2 \\ &\le \operatorname{dist}(z^k, T^{-1}(0))^2 - \frac{c_3(1-\bar{\sigma})^2\|\hat{z}^k - z^k\|^2}{(2\bar{\alpha})^2} \\ &\le \left(1 - \frac{c_3(1-\bar{\sigma})^2}{(2\bar{\alpha}(1 + c_1/\underline{\alpha})(1+c_4))^2}\right)\operatorname{dist}(z^k, T^{-1}(0))^2,\end{aligned}$$

where the third inequality follows from (28) and the last from (33). The latter relation establishes the linear rate of convergence of $\{\operatorname{dist}(z^k, T^{-1}(0))\}$ to zero. For Fejér-monotone sequences this is equivalent to the linear rate of convergence of $\{z^k\}$ to its limit, e.g. Ref. [3]. ■

Remark 2.1 We note that the assumptions of Lipschitz-continuity of G on $X \times Y$ and of G_1^k on Y can be replaced by assuming the same properties on bounded sets. Indeed, once the iterates are well defined, the analysis above immediately shows that the generated sequence is bounded, and thus we can restrict the analysis to a bounded set.

Additionally, we could consider the case where G and G_1^k are assumed to be only continuous, similarly to some results in Ref. [37], although in that case some further technical details have to be worked out.

Remark 2.2 The choice of parameter α_k can be made according to (27), if the Lipschitz constant L is known or can be estimated (for example, in the problem (2) of Chen and Teboulle, $L = \|A\|$). Of course, (27) is a rather conservative estimate and one might expect that a larger value should still be acceptable for convergence. Appropriate α_k can be obtained by a suitable Armijo-like backtracking procedure. The proof of Theorem 2.1 establishes that such a procedure would be well defined, since there exists an interval of acceptable values $(0, \bar{\alpha}]$ with $\bar{\alpha} > 0$.

Remark 2.3 It can be seen that after computing x^{k+1} and y^{k+1} by (15) and (16), we could project the point (x^{k+1}, y^{k+1}) onto any closed convex set containing $T^{-1}(0)$ before proceeding to the next iteration. For example, if H is the normal cone operator defined in (18), then we can project onto the set $X \times C$. With this modification, the convergence analysis still applies, because the left-hand side in the key relation (25) can only become smaller after the projection.

Remark 2.4 The error bound condition (19) is equivalent to the following Lipschitzian property of T^{-1} at zero:

$$T^{-1}(v) \subset T^{-1}(0) + c_1\|v\|\mathcal{B} \quad \forall v \in c_2\mathcal{B},$$

where $\mathcal{B} = \{z \in Z \mid \|z\| \le 1\}$. We note that this condition does not imply that the solution set $T^{-1}(0)$ is a singleton, unlike in most rate of convergence results for proximal-like methods, e.g. Refs. [9,25,31] [there are some exceptions, however, e.g. Refs. 36,37].

3 SOME SPECIAL CASES

In this section, we apply our general decomposition scheme to problems where some further structure is assumed.

3.1 Coupling the Chen–Teboulle Method with Bundle Techniques

We now turn our attention back to problem (2) and decomposition scheme (3). At each iteration of the method we have to solve a problem of the form

$$\min_{\xi \in X_1 \times X_2} \left\{ \alpha_k L(\xi, \hat{y}^k) + \frac{\|\xi - x^k\|^2}{2} \right\},$$

which is decomposable as discussed in Section 1. In this subsection, we shall be talking about solving the above problem as a whole, with the understanding that the actual computational work is done separately for the two independent pieces. The above problem can be written as

$$\min_{\xi} \Psi_k(\xi) := \alpha_k (f(\xi) + \langle b^k, \xi \rangle) + \frac{\|\xi - x^k\|^2}{2}, \tag{34}$$

where

$$f(\xi) = f_1(\xi_1) + f_2(\xi_2), \quad b^k = (A^\top \hat{y}^k, -\hat{y}^k).$$

If the function f_1 and/or f_2 is nondifferentiable, then (34) is a nonsmooth convex minimization problem. The family of bundle methods [4, Chap. 9] is perhaps the most practical computational tool for nonsmooth optimization. In what follows, we shall argue that HPDM provides a convenient framework for coupling the decomposition scheme (3) with appropriate bundle techniques for solving the nonsmooth minimization subproblems (34).

The approach of (proximal form of) bundle methods consists in replacing the objective function with its regularized cutting-planes approximation in order to generate the next candidate point. Since the smooth part of the objective function in (34) is quadratic, it appears to make sense to insert it directly into the (quadratic programming) subproblems of the bundle method, using linearizations of the nonsmooth part only. We therefore arrive at the following scheme.

Suppose that the bundle method had already generated a sequence $\xi^0, \ldots, \xi^m$ of candidate points, $m \geq 0$ (ξ^0 is an arbitrary starting point). By $\bar{\xi}^m \in \{\xi^0, \ldots, \xi^m\}$ we denote the stability centre, which is (roughly speaking) the point with the lowest objective function value generated so far. At every candidate point ξ^i, we have available the values of the nonsmooth part f of the objective function Ψ_k and of *one* of the subgradients of f. This information defines the current cutting-planes approximation of f:

$$\check{f}_m(\xi) := \max_{i \leq m} \{ f(\xi^i) + \langle g^i, \xi - \xi^i \rangle \}, \quad g^i \in \partial f(\xi^i).$$

The next candidate point is generated by

$$\xi^{m+1} = \arg\min_{\xi} \check{\Psi}_m(\xi) := \alpha_k (\check{f}_m(\xi) + \langle b^k, \xi \rangle) + \frac{\|\xi - x^k\|^2}{2} + \frac{\mu_m \|\xi - \bar{\xi}^m\|^2}{2}, \tag{35}$$

where $0 < \underline{\mu} \leq \mu_m \leq \bar{\mu}$ is the stabilization parameter. We note that treating the smooth part of Ψ_k directly in the subproblem (35) (instead of approximating it by cutting planes) does not increase the complexity of the subproblem, since an obvious reformulation of (35) is still a quadratic programme. The new point ξ^{m+1} is accepted as the next stability centre $\bar{\xi}^{m+1}$ if $\Psi_k(\xi^{m+1})$ is sufficiently smaller than $\Psi_k(\bar{\xi}^m)$. Otherwise, the stability centre does not change ($\bar{\xi}^{m+1} := \bar{\xi}^m$), and the next candidate point is generated using the enriched approximation $\check{f}_{m+1}$ of f (and therefore the enriched approximation $\check{\Psi}_{m+1}$ of Ψ_k).

By optimality condition for (35), we have that

$$0 = \alpha_k(\check{g}^{m+1} + b^k) + \xi^{m+1} - x^k + \mu_m(\xi^{m+1} - \bar{\xi}^m), \quad \check{g}^{m+1} \in \partial \check{f}_m(\xi^{m+1}). \tag{36}$$

By the subgradient inequality, for all ξ it holds that

$$\check{f}_m(\xi) \geq \check{f}_m(\xi^{m+1}) + \langle \check{g}^{m+1}, \xi - \xi^{m+1} \rangle.$$

Since by construction $f(\xi) \geq \check{f}_m(\xi)$, we have that for all ξ,

$$\begin{aligned} f(\xi) &\geq \check{f}_m(\xi^{m+1}) + \langle \check{g}^{m+1}, \xi - \xi^{m+1} \rangle \\ &= f(\bar{\xi}^m) + \langle \check{g}^{m+1}, \xi - \bar{\xi}^m \rangle - e_m^1, \end{aligned} \tag{37}$$

where

$$e_m^1 = f(\bar{\xi}^m) - \check{f}_m(\xi^{m+1}) - \langle \check{g}^{m+1}, \bar{\xi}^m - \xi^{m+1} \rangle \geq 0,$$

with the nonnegativity of e_m^1 following from setting $\xi = \bar{\xi}^m$ in (37). We therefore have that

$$\check{g}^{m+1} \in \partial_{e_m^1} f(\bar{\xi}^m). \tag{38}$$

It can be also seen by direct verification that

$$(\xi^{m+1} - x^k) \in \partial_{e_m^2} \left(\frac{\| \cdot - x^k \|}{2} \right) (\bar{\xi}^m), \tag{39}$$

$$e_m^2 = \frac{\|\bar{\xi}^m - x^k\|^2}{2} - \frac{\|\xi^{m+1} - x^k\|^2}{2} - \langle \xi^{m+1} - x^k, \bar{\xi}^m - \xi^{m+1} \rangle \geq 0.$$

Using (36), we then obtain

$$\xi^{m+1} = \bar{\xi}^m - \frac{1}{\mu_m} \check{G}^m, \tag{40}$$

where

$$\begin{aligned} \check{G}^m &= \alpha_k(\check{g}^{m+1} + b^k) + \xi^{m+1} - x^k \\ &\in \alpha_k \partial_{e_m^1} f(\bar{\xi}^m) \\ &\quad + \alpha_k b^k + \partial_{e_m^2} \left(\frac{\| \cdot - x^k \|}{2} \right) (\bar{\xi}^m) \\ &\subset \partial_{e_m} \Psi_k(\bar{\xi}^m), \end{aligned} \tag{41}$$

where

$$e_m = e_m^1 + e_m^2 \geq 0,$$

and the first inclusion is by (38) and (39), while the last inclusion follows from the fact that $(\partial_{\varepsilon_1} \varphi_1 + \partial_{\varepsilon_2} \varphi_2) \subset \partial_{\varepsilon_1 + \varepsilon_2}(\varphi_1 + \varphi_2)$ for any convex functions φ_1 and φ_2 and any $\varepsilon_1, \varepsilon_2 \geq 0$.

The standard (implementable) stopping test for the bundle method applied to (34) [e.g. Ref. 4, Chap. 9] is

$$e_m + \frac{1}{2\mu_m} \|\check{G}^m\|^2 \leq \texttt{TOL}, \tag{42}$$

where all the quantities involved are readily available, as exhibited above. We next show that an appropriate choice of $\texttt{TOL} > 0$ guarantees that an approximate solution of the minimization subproblems in (3) computed by the bundle method satisfies the criteria of HPDM.

To decide whether the current $\bar{\xi}^m$ can be accepted as $\hat{x}^k$ within HPDM, we shall be checking the criterion (14) (or, equivalently, (23)). So, given $\bar{\xi}^m$, define

$$\bar{u}^m = \check{g}^{m+1} + b^k \in \partial_{e_m^1} f(\bar{\xi}^m) + b^k \subset (\partial_x L)^{e_m^1}(\bar{\xi}^m, \hat{y}^k),$$

where the second inclusion is by the fact that $\partial_\varepsilon \varphi \subset (\partial \varphi)^\varepsilon$ for any convex function φ and any $\varepsilon \geq 0$ (recall that the enlargement of the operator is understood here with respect to the ξ-part, with the y-part fixed). Define further

$$\bar{r}^m = \alpha_k \bar{u}^m + \bar{\xi}^m - x^k = \alpha_k(\check{g}^{m+1} + b^k) + \bar{\xi}^m - x^k = \check{G}^m + \bar{\xi}^m - \xi^{m+1},$$

where (41) was used. We then have that

$$\|\bar{r}^m\| \leq \|\check{G}^m\| + \|\bar{\xi}^m - \xi^{m+1}\| = \left(1 + \frac{1}{\mu_m}\right)\|\check{G}^m\|,$$

where the equality is by (40). Hence, if (42) holds then

$$\begin{aligned}\|\bar{r}^m\|^2 + 2\alpha_k e_m^1 &\leq \left(1 + \frac{1}{\mu_m}\right)^2 \|\check{G}^m\|^2 + 2\alpha_k e_m \\ &\leq 2\left(\mu_m\left(1 + \frac{1}{\mu_m}\right)^2 + \alpha_k\right)\texttt{TOL}. \end{aligned} \tag{43}$$

If we further define $\bar{w}^m$ and $\bar{s}^m$ as in HPDM but with $\hat{x}^k$ substituted by the current trial point $\bar{\xi}^m$, and set

$$\texttt{TOL}_m := \frac{\sigma_k(\|\alpha_k \bar{u}^m\|^2 + \|\alpha_k \bar{w}^m\|^2 + \|\bar{\xi}^m - x^k\|^2 + \|\hat{y}^k - y^k\|^2) - \|\bar{s}^m\|^2}{2\mu_m(1 + 1/\mu_m)^2 + \alpha_k)},$$

then whenever the bundle method stopping test (42) would be satisfied with this $\texttt{TOL}_m$, it would imply (by virtue of (43)) that

$$\|\bar{r}^m\|^2 + 2\alpha_k e_m^1 \leq \sigma_k(\|\alpha_k \bar{u}^m\|^2 + \|\alpha_k \bar{w}^m\|^2 + \|\bar{\xi}^m - x^k\|^2 + \|\hat{y}^k - y^k\|^2) - \|\bar{s}^m\|^2,$$

which means that the HPDM condition (14) is satisfied for $\hat{x}^k = \bar{\xi}^m$, $\hat{u}^k = \bar{u}^m$, $r^k = \bar{r}^m$, $s^k = \bar{s}^m$ and $\varepsilon_k = e_m^1$.

It remains to note that $\texttt{TOL}_m$ stays bounded away from zero, while by the convergence properties of the bundle method,

$$0 = \lim_{m\to\infty} \check{G}^m \quad \text{and} \quad 0 = \lim_{m\to\infty} e_m.$$

This means that the bundle method stopping test (42) (with $\texttt{TOL}_m$ in the right-hand side) would be satisfied after a finite number of iterations, yielding an acceptable approximate solution for the proximal point subproblem in HPDM. To see that $\texttt{TOL}_m$ stays bounded away from zero, observe that by the fact that the bundle iterates converge to the minimizer of (34), it holds that

$$\lim_{m\to\infty} \alpha_k \bar{u}^m = \lim_{m\to\infty} (x^k - \bar{\xi}^m).$$

If we are not at a solution of our problem (2), then it must be the case that either the limit above is nonzero (equivalently, $0 \notin \partial_x L(x^k, \hat{y}^k)$) or $\hat{y}^k \neq y^k$ (equivalently, $0 \notin \partial_y L(x^k, \hat{y}^k)$), or both. In either case, it is easily seen that $\mathtt{TOL}_m$ is bounded away from zero as $m \to \infty$.

3.2 Relations with Tseng's Method

Consider now problem (1) with T given by (11), where H is the normal cone mapping (18) for some closed convex set $C \subset Y$. Given $(x^k, y^k) \in X \times Y$, the decomposition method of Tseng [37] performs the following updates:

$$\begin{aligned} \hat{y}^k &= P_C(y^k - \alpha_k(G_1(\hat{y}^k) - G_1(y^k) + G(x^k, y^k))), \\ x^{k+1} &= (I + \alpha_k F(\cdot, \hat{y}^k))^{-1} x^k, \\ y^{k+1} &= P_C(y^k - \alpha_k G(x^{k+1}, \hat{y}^k), \end{aligned} \tag{44}$$

where G_1 and α_k are chosen in a manner analogous to HPDM, except that G_1 is fixed over the iterations.

We first note that the step to compute $\hat{y}^k$ in (44) is precisely the forward–backward step (12) in HPDM for the choice of $H = N_C$. Furthermore, the step to compute x^{k+1} in (44) amounts to demanding that $r^k = 0$ and $\varepsilon_k = 0$ in (13) of HPDM. In the latter case, if we choose $\tau_k = 1$ in (15) (which is admissible), we obtain in HPDM $x^{k+1} = x^k - \alpha_k \hat{u}^k = \hat{x}^k - r^k = \hat{x}^k$, with $\hat{x}^k$ in that case being the exact solution of the proximal sub-problem. Therefore x^{k+1} from (44) can be obtained within HPDM. But the steps to compute y^{k+1} in (44) and HPDM are in general different, except when $C = Y$. If $C = Y$ then $H = N_C = 0$, and thus $\hat{h}^k = 0$ in HPDM, so that $\hat{w}^k = G(\hat{x}^k, \hat{y}^k)$. If we solve the proximal subproblem in HPDM exactly then $x^{k+1} = \hat{x}^k$, as noted above. Taking further $\tau_k = 1$ in (16), we obtain $y^{k+1} = y^k - \alpha_k \hat{w}^k = y^k - \alpha_k G(x^{k+1}, \hat{y}^k)$, which is the same as in (44) when $C = Y$.

We emphasize that even for the case of $C = Y$, HPDM presents an important improvement over (44) in allowing the proximal subproblems to be solved approximately, according to a constructive criterion (e.g. recall Section 3.1). It should be noted that the analysis of Ref. [37] could be extended to handle approximate solutions, but using the standard summability-type error conditions only (which are not very suitable for computational use). Some other potentially useful features of HPDM, which add more flexibility, are the following. The splitting term G_1^k is not fixed and can be adjusted along the iterations (the structure of $G(x^k, \cdot)$ may differ depending on x^k). In addition, an extra stepsize parameter τ_k is allowed for accelerating convergence.

We next comment on the case of $C \neq Y$. As already noted, in that case the resulting y^{k+1} is in general different for the two methods. One advantage of HPDM is that we were able to establish the linear rate of convergence result, while in Ref. [37] the rate of convergence was obtained for $C = Y$ only (the case of $C \neq Y$ was posed as an open question). To further compare the different updates of y, consider the case where (11) has no x-part, so that the problem becomes the variational inequality

$$0 \in G(y) + N_C(y).$$

Choosing $G_1 \equiv 0$, (44) gives

$$\hat{y}^k = P_C(y^k - \alpha_k G(y^k)), \quad y^{k+1} = P_C(y^k - \alpha_k G(\hat{y}^k)),$$

which is the extragradient iteration [14,16]. Choosing $G_1^k \equiv 0$ and $\tau_k = 1$ in HPDM, we have

$$\begin{aligned} \hat{y}^k &= P_C(y^k - \alpha_k G(y^k)) = y^k - \alpha_k(G(y^k) + \hat{h}^k), \quad \hat{h}^k \in N_C(\hat{y}^k), \\ y^{k+1} &= y^k - \alpha_k(G(\hat{y}^k) + \hat{h}^k) = \hat{y}^k - \alpha_k(G(\hat{y}^k) - G(y^k)). \end{aligned} \tag{45}$$

The latter is the projection method derived from the modified forward–backward splitting scheme [38, Example 2], which is different from the extragradient method. It is also related to the modified projection-type methods of Ref. [33]. In particular, Algorithm 3.1 in Ref. [33] computes

$$y^{k+1} = y^k - \gamma_k(y^k - \hat{y}^k - \alpha_k G(y^k) + \alpha_k G(\hat{y}^k)),$$

with a certain $\gamma_k > 0$. The latter is the same as (45) if $\gamma_k = 1$. It should be noted, however, that it is not clear whether $\gamma_k = 1$ is admissible in Ref. [33].

Finally, we point out that HPDM is applicable beyond the case where H is the normal cone operator. Consider, for example, the min–max problem

$$\min_{x\in X} \max_{y\in Y}\{f(x) - g(y) + \langle y, Bx - b\rangle\}, \tag{46}$$

where f and g are closed proper convex functions on the Euclidean spaces X and Y, respectively, $B\colon X \to Y$ is linear, $b \in Y$. This problem is equivalent to (1) with T defined in (11), if we set

$$F(x, y) = \partial f(x) + B^\top y, \quad G(x, y) = b - Bx, \quad H(y) = \partial g(y).$$

It can be seen that applying HPDM (with $G_k^1 \equiv 0$, $\varepsilon_k = 0$, $r^k = 0$, $\tau_k = 1$), we obtain

$$\begin{aligned}
\hat{y}^k &= \arg\min_{y\in Y}\left\{\alpha_k(g(y) + \langle y, b - Bx^k\rangle) + \frac{\|y - y^k\|^2}{2}\right\},\\
x^{k+1} &= \arg\min_{x\in X}\left\{\alpha_k(f(x) + \langle x, B^\top \hat{y}^k\rangle) + \frac{\|x - x^k\|^2}{2}\right\},\\
y^{k+1} &= \hat{y}^k - \alpha_k B(x^k - x^{k+1}).
\end{aligned}$$

The two nonsmooth convex optimization problems above further decompose according to the structure of f and g, and can be solved approximately by an appropriate application of bundle methods, as discussed in Section 3.1.

We note that for problem (46) the method of Ref. [37] can handle only the case where $g(y) = h(y) + I_C(y)$, where h is continuously differentiable and I_C is the indicator function of a closed convex set C in Y (then we could define $G(x, y) = h'(y) + b - Bx$ and $H(y) = \partial I_C(y) = N_C(y)$). In particular, within the framework of Ref. [37] g cannot have any nonsmoothness except when induced by the constraints on the variable y.

Acknowledgments

The author is supported in part by CNP_q Grants 300734/95-6(RN) and 471780/2003-0, by PRONEX-Optimization and by FAPERJ.

References

[1] A. Auslender (1985). Two general methods for computing saddle points with applications for decomposing convex programming problems. *Applied Mathematics and Optimization*, **13**, 79–95.

[2] A. Auslender and M. Teboulle (2001). Entropic proximal decomposition methods for convex programs and variational inequalities. *Mathematical Programming*, **91**, 33–47.

[3] H.H. Bauschke (2001). Projection algorithms: results and open problems. newblock In: D. Butnariu, Y. Censor and S. Reich (Eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, Vol. 8 of *Studies in Computational Mathematics*, pp. 11–22. Elsevier Science B.V., Amsterdam.
[4] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal and C. Sagastizábal (2003). *Numerical Optimization: Theoretical and Practical Aspects*. Springer-Verlag, Berlin, Germany.
[5] R.S. Burachik, A.N. Iusem and B.F. Svaiter (1997). Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, **5**, 159–180.
[6] R.S. Burachik, C.A. Sagastizábal and S. Scheimberg (2002). An inexact method of partial inverses. Application to splitting methods. IMPA preprint A144, Rio de Janeiro, Brazil.
[7] R.S. Burachik, C.A. Sagastizábal and B. F. Svaiter (1999). Bundle methods for maximal monotone operators. In: R. Tichatschke and M. Théra (Eds.), *Ill-posed variational problems and ragularization techniques*, Lecture Notes in Economics and Mathematical Systems, No. 477, pp. 49–64. Springer-Verlag, Berlin.
[8] R.S. Burachik, C.A. Sagastizábal and B.F. Svaiter (1999). ε-Enlargements of maximal monotone operators: Theory and applications. In: M. Fukushima and L. Qi (Eds.), *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 25–44. Kluwer Academic Publishers, Dordrecht.
[9] G. Chen and M. Teboulle (1994). A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, **64**, 81–101.
[10] H.-G. Chen and R.T. Rockafellar (1997). Convergence rates in forward-backward splitting. *SIAM Journal on Optimization*, **7**, 421–444.
[11] P.L. Combettes (2001). Quasi-Fejérian analysis of some optimization algorithms. In: D. Butnariu, Y. Censor and S. Reich (Eds.), *Inherantly Parallel Algorithms in Feasibility and Optimization and their Applications*, Vol. 8 of *Studies in Computational Mathematics*, pp. 115–152. Elsevier Science B.V., Amsterdam.
[12] J. Eckstein and D.P. Bertsekas (1992). On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, **55**, 293–318.
[13] D. Gabay (1983). Applications of the method of multipliers to variational inequalities. In: M. Fortin and R. Glowinski (Eds.), *Augmented Lagrangian Methods : Applications to the Solution of Boundary-Value Problems*, pp. 299–331. North–Holland, Amsterdam.
[14] E.N. Khobotov (1987). A modification of the extragradient method for the solution of variational inequalities and some optimization problems. *USSR Computational Mathematics and Mathematical Physics*, **27**, 1462–1473.
[15] K.C. Kiwiel, C.H. Rosa and A. Ruszczyński (1999). Proximal decomposition via alternating linearization. *SIAM Journal on Optimization*, **9**, 668–689.
[16] G.M. Korpelevich (1976). The extragradient method for finding saddle points and other problems. *Matecon*, **12**, 747–756.
[17] B. Lemaire (1989). The proximal algorithm. In: J.-P. Penot (Ed.), *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics 87*, pp. 73–87. Birkhauser, Basel.
[18] P.L. Lions and B. Mercier (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, **16**, 964–979.
[19] B. Martinet (1970). Regularisation d'inequations variationelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationelle*, **4**, 154–159.
[20] A. Ouorou (2004). Epsilon-proximal decomposition method. *Mathematical Programming*, **99**, 89–108.
[21] G.B. Passty (1979). Ergodic convergence to a zero of the sum of monotone operators in Hubert space. *Journal of Mathematical Analysis and Applications*, **72**, 383–390.
[22] T. Pennanen (2002). A splitting method for composite mappings. *Numerical Functional Analysis and Optimization*, **23**, 875–890.
[23] R.T. Rockafellar (1970). On the maximality of sums of nonlinear monotone operators. *Transactions of the American Mathematical Society*, **149**, 75–88.
[24] R.T. Rockafellar (1976). Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, **1**, 97–116.
[25] R.T. Rockafellar (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, **14**, 877–898.
[26] C.A. Sagastizábal and MV. Solodov (2001). On the relation between bundle methods for maximal monotone inclusions and hybrid proximal point algorithms. In: D. Butnariu, Y. Censor and S. Reich (Eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, Vol. 8 of *Studies in Computational Mathematics*, pp. 441–455. Elsevier Science B.V., Amsterdam.
[27] M.V. Solodov and B.F. Svaiter (1999). A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, **7**, 323–345.
[28] M.V. Solodov and B.F. Svaiter (1999). A hybrid projection–proximal point algorithm. *Journal of Convex Analysis*, **6**, 59–70.
[29] M.V. Solodov and B.F. Svaiter (2000). Error bounds for proximal point subproblems and associated inexact proximal point algorithms. *Mathematical Programming*, **88**, 371–389.
[30] M.V. Solodov and B.F. Svaiter (2000). A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM Journal on Optimization*, **10**, 605–625.
[31] M.V. Solodov and B.F. Svaiter (2001). A unified framework for some inexact proximal point algorithms. *Numerical Functional Analysis and Optimization*, **22**, 1013–1035.
[32] M.V. Solodov and B.F. Svaiter (2002). A new proximal-based globalization strategy for the Josephy-Newton method for variational inequalities. *Optimization Methods and Software*, **17**, 965–983.

[33] M.V. Solodov and P. Tseng (1996). Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, **34**, 1814–1830.
[34] J.E. Spingarn (1985). Applications of the method of partial inverses to convex programming. *Mathematical Programming*, **32**, 199–223.
[35] P. Tseng (1991). Applications of a splitting algorithm to decomposition in convex programmning and variational inequalities. *SIAM Journal on Control and Optimization*, **29**, 119–138.
[36] P. Tseng (1995). On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, **60**, 237–252.
[37] P. Tseng (1997). Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM Journal on Optimization*, **7**, 951–965.
[38] P. Tseng (2000). A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, **38**, 431–446.