Iteration-complexity of block-decomposition algorithms and the alternating minimization augmented Lagrangian method

Renato D. C. Monteiro* B. F. Svaiter†

August 1, 2010

Abstract

In this paper, we consider the monotone inclusion problem consisting of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable two-block structure and introduce a framework of block-decomposition prox-type algorithms for solving it which allows for each one of the single-block proximal subproblems to be solved in an approximate sense. Moreover, by showing that any method in this framework is also a special instance of the hybrid proximal extragradient (HPE) method introduced by Solodov and Svaiter, we derive corresponding convergence rate results. We also describe some instances of the framework based on specific and inexpensive schemes for solving the single-block proximal subproblems. Finally, we consider some applications of our methodology to: i) propose new algorithms for the monotone inclusion problem consisting of the sum of two maximal monotone operators, and; ii) study the complexity of the classical alternating minimization augmented Lagrangian method for a class of linearly constrained convex programming problems with proper closed convex objective functions.

1 Introduction

A broad class of optimization, saddle point, equilibrium and variational inequality (VI) problems can be posed as the *monotone inclusion problem*, namely: finding x such that $0 \in T(x)$, where T is a maximal monotone point-to-set operator. The proximal point method, proposed by Rockafellar [20], is a classical iterative scheme for solving the monotone inclusion problem which generates a sequence $\{x_k\}$ according to

$$x_k = (\lambda_k T + I)^{-1}(x_{k-1}).$$

It has been used as a generic framework for the design and analysis of several implementable algorithms. The classical inexact version of the proximal point method allows for the presence of a sequence of summable errors in the above iteration, i.e.:

$$\|x_k - (\lambda_k T + I)^{-1}(x_{k-1})\| \leq e_k, \qquad \sum_{k=1}^{\infty} e_k < \infty.$$

Convergence results under the above error condition have been established in [20] and have been used in the convergence analysis of other methods that can be recast in the above framework.

New inexact versions of the proximal point method with relative error tolerance were proposed by Solodov and Svaiter [22, 23, 24, 25]. Iteration complexity results for one of these inexact versions of the proximal point

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grants CCF-0808863 and CMMI-0900094 and ONR Grant ONR N00014-08-1-0033.

†IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (emal: benar@impa.br). The work of this author was partially supported by CNPq grants 480101/2008-6, 303583/2008-8, FAPERJ grant E-26/102.821/2008 and PRONEX-Optimization

method introduced in [22], namely the hybrid proximal extragradient (HPE) method, were established in [14]. As a consequence, iteration complexity results for Korpelevich's extragradient method for mixed variational inequalities with Lipschitz continuous monotone operators, and a variant of Tseng's modified forward-backward splitting (MF-BS) method (see [27]) for finding a zero of the sum of a Lipschitz continuous monotone map with a maximal monotone operator whose resolvent is assumed to be easily computable, were also derived in [14] and [15] by showing that both methods are special cases of the HPE method. Moreover, iteration-complexity results for these methods are developed in [15] in the context of a generalized saddle point problem and a large class of linearly constrained convex programming problems, including for example cone programming and problems whose objective functions converge to infinity as the boundaries of their domain are approached. A nice feature of the analysis in [14] and [15] is that, by working with some suitable termination criteria, it is shown that its complexity results, as opposed to the ones in [16], also apply to variational inequality and/or monotone inclusion problems with unbounded feasible sets.

In this paper, we continue along the same line of investigation as in our previous papers [14] and [15], which is to use the HPE method as a general framework to derive iteration-complexity results for specific algorithms for solving various types of structured monotone inclusion problems. More specifically, we consider the monotone inclusion problem consisting of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable two-block structure. We introduce a general block-decomposition HPE (BD-HPE) framework in the context of this inclusion problem, which allows for each one of the single-block proximal subproblems to be solved in an approximate sense. Moreover, by showing that any method in this framework is also a special instance of the HPE method, we derive convergence rate results for the BD-HPE framework based on the ones developed in [14] for the HPE method. Subsequently, we describe some ways of implementing the BD-HPE framework based on specific and inexpensive schemes for solving the single-block proximal subproblems. We also consider some applications of our methodology introduced here to: i) propose new algorithms for the monotone inclusion problem consisting of the sum of two maximal monotone operators, and; ii) study the complexity of the classical alternating minimization augmented Lagrangian (AMAL) method for a class of linearly constrained convex programming problems with proper closed convex objective functions.

The AMAL (also called alternating direction) method was first introduced in [10, 11]. Recently, there has been some growing interest in the AMAL method for solving large scale linear cone programming (see for example [5, 4, 18, 12, 13]). However, to the best of our knowledge, no iteration-complexity analysis for the AMAL method have been established so far. Development and analysis of splitting and block-decomposition methods is by now a well-developed area, although algorithms which allow a relative error tolerance in the solution of the proximal subproblems have been studied in just a few papers. In particular, Ouorou [17] discusses an ε-proximal decomposition using the ε-subdifferential and a relative error criterion on ε. Projection splitting methods for the sum of arbitrary maximal monotone operators using a particular case of the HPE error tolerance for solving the proximal subproblems were presented in [7, 8]. The use of the HPE method for studying block-decomposition methods was first presented in [21]. We observe however that none of these works deal with the derivation of iteration-complexity bounds. More recently, Chambolle and Pock [6] have developed and established iteration-complexity bounds for a block-decomposition method, which solves the proximal subproblems exactly, in the context of saddle-point problems with a bilinear coupling.

This paper is organized as follows. Section 2 contains two subsections. Subsection 2.1 reviews some basic definitions and facts on convex functions and the definition and some basic properties of the ε-enlargement of a point-to-set maximal monotone operator. Subsection 2.2 reviews the HPE method and the global convergence rate results obtained for it in [14]. Section 3 introduces the BD-HPE framework for solving a special type of monotone inclusion problem mentioned above and shows that any instance of the framework can be viewed as a special case of the HPE method. As a consequence, global convergence rate results for the BD-HPE framework are also obtained in this section using the general theory outlined in Subsection 2.2. Section 4 describes specific schemes for solving the single-block proximal subproblems based on a small number (one or two) of resolvent evaluations. Section 5 describes some instances of the BD-HPE framework which, are not only interesting in their own right, but also illustrate the use of the different schemes for solving the single-block proximal subproblems. It contains three subsections as follows. Subsection 5.1 discusses a specific instance of the BD-HPE framework where both single-block proximal subproblems are solved exactly. Subsection 5.2 gives another instance of the BD-HPE framework in which both single-block proximal subproblems are approximately solved by means of a Tseng's type scheme. Subsection 5.3 studies a block-decomposition method for a large class of linearly constrained convex optimization problems, which includes cone programs and problems whose objective functions converge to infinity as the relative boundaries of their domain are approached. Section 6 considers the monotone inclusion problem consisting of the sum of two maximal monotone operators and show how it can be transformed to an equivalent monotone inclusion problem with two-block structure of the aforementioned type, which can then be solved by any instance of the BD-HPE framework. Section 7 considers the AMAL method for solving a class of linearly constrained convex programming problems with proper closed convex objective functions and shows that it can be interpreted as a specific instance of the BD-HPE framework applied to a two-block monotone inclusion problem.

2 Technical background

This section contains two subsections. In the first one, we review some basic definitions and facts about convex functions and ε-enlargement of monotone multi-valued maps. This subsection also reviews the weak transportation formula for the ε-subdifferentials of closed convex functions and the ε-enlargements of maximal monotone operators. The second subsection reviews the HPE method and the global convergence rate results obtained for it in [14].

2.1 The ε-subdifferential and ε-enlargement of monotone operators

Let $\mathcal{Z}$ denote a finite dimensional inner product space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. A point-to-set operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is a relation $T \subset \mathcal{Z} \times \mathcal{Z}$ and

$$T(z) = \{v \in \mathcal{Z} \mid (z, v) \in T\}.$$

Alternatively, one can consider T as a multi-valued function of $\mathcal{Z}$ into the family $\wp(\mathcal{Z}) = 2^{(\mathcal{Z})}$ of subsets of $\mathcal{Z}$. Regardless of the approach, it is usual to identify T with its graph defined as

$$Gr(T) = \{(z, v) \in \mathcal{Z} \times \mathcal{Z} \mid v \in T(z)\}.$$

The domain of T, denoted by $\operatorname{Dom} T$, is defined as

$$\operatorname{Dom} T := \{z \in \mathcal{Z} : T(z) \neq \emptyset\}.$$

An operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is *affine* if its graph is an affine manifold. Clearly, if T is affine, then the following implication holds:

$$\left.\begin{array}{c} \alpha_i \geq 0, \ i = 1, \ldots, k \\ \alpha_1 + \ldots + \alpha_k = 1 \\ v_i \in T(z_i), \ 1, \ldots, k \end{array}\right\} \Longrightarrow \sum_{i=1}^{k} \alpha_i v_i \in T\left(\sum_{i=1}^{k} \alpha_i v_i\right). \tag{1}$$

Moreover, $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is *monotone* if

$$\langle v - \tilde{v}, z - \tilde{z} \rangle \geq 0, \qquad \forall (z, v), (\tilde{z}, \tilde{v}) \in Gr(T),$$

and T is *maximal monotone* if it is monotone and maximal in the family of monotone operators with respect to the partial order of inclusion, i.e., $S : \mathcal{Z} \rightrightarrows \mathcal{Z}$ monotone and $Gr(S) \supset Gr(T)$ implies that $S = T$.

For a scalar $\varepsilon \geq 0$, the ε-subdifferential of a function $f : \mathcal{Z} \to \bar{\mathbb{R}}$ is the operator $\partial_\varepsilon f : \mathcal{Z} \rightrightarrows \mathcal{Z}$ defined as

$$\partial_\varepsilon f(z) = \{v \mid f(\tilde{z}) \geq f(z) + \langle \tilde{z} - z, v \rangle - \varepsilon, \ \forall \tilde{z} \in \mathcal{Z}\}, \quad \forall z \in \mathcal{Z}. \tag{2}$$

When $\varepsilon = 0$, the operator $\partial_\varepsilon f$ is simply denoted by ∂f and is referred to as the subdifferential of f. The operator ∂f is trivially monotone if f is proper. If f is a proper lower semi-continuous convex function, then ∂f is maximal monotone [19].

The conjugate f^* of f is the function $f^* : \mathcal{Z} \to \bar{\mathbb{R}}$ defined as

$$f^*(v) = \sup_{z\in\mathcal{Z}} \langle v, z\rangle - f(z), \quad \forall v \in \mathcal{Z}.$$

The following result lists some useful properties about the ε-subdifferential of a proper convex function.

Proposition 2.1. *Let $f : \mathcal{Z} \to \bar{\mathbb{R}}$ be a proper convex function. Then,*

a) $\partial_\varepsilon f(z) \subset (\partial f)^\varepsilon(z)$ for any $\varepsilon \geq 0$ and $z \in \mathcal{Z}$;

b) $\partial_\varepsilon f(z) = \{v \mid f(z) + f^*(v) \leq \langle z, v\rangle + \varepsilon\}$ for any $\varepsilon \geq 0$ and $z \in \mathcal{Z}$;

c) if $v \in \partial f(z)$ and $f(\tilde{z}) < \infty$, then $v \in \partial_\varepsilon f(\tilde{z})$, where $\varepsilon := f(\tilde{z}) - [f(z) + \langle \tilde{z} - z, v\rangle]$.

The *indicator function* of a closed convex set $Z \subset \mathcal{Z}$ is the function $\delta_Z : \mathcal{Z} \to \bar{\mathbb{R}}$ defined as

$$\delta_Z(z) = \begin{cases} 0, & z \in Z, \\ \infty, & \text{otherwise,} \end{cases}$$

and the *normal cone operator* of Z is the point-to-set map $N_Z : \mathcal{Z} \rightrightarrows \mathcal{Z}$ given by

$$N_Z(z) = \begin{cases} \emptyset, & z \notin Z, \\ \{v \in \mathcal{Z}, \mid \langle \tilde{z} - z, v\rangle \leq 0, \ \forall \tilde{z} \in Z\}, & z \in Z. \end{cases} \tag{3}$$

Clearly, the normal cone operator N_Z of Z can be expressed in terms of δ_Z as $N_Z = \partial\delta_Z$.

In [1], Burachik, Iusem and Svaiter introduced the ε-enlargement of maximal monotone operators. In [14] this concept was extended to a generic point-to-set operator in $\mathcal{Z}$ as follows. Given $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ and a scalar ε, define $T^\varepsilon : \mathcal{Z} \rightrightarrows \mathcal{Z}$ as

$$T^\varepsilon(z) = \{v \in \mathcal{Z} \mid \langle z - \tilde{z}, v - \tilde{v}\rangle \geq -\varepsilon, \quad \forall \tilde{z} \in \mathcal{Z}, \ \forall \tilde{v} \in T(\tilde{z})\}, \quad \forall z \in \mathcal{Z}. \tag{4}$$

We now state a few useful properties of the operator T^ε that will be needed in our presentation.

Proposition 2.2. *Let $T, T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$. Then,*

a) if $\varepsilon_1 \leq \varepsilon_2$, then $T^{\varepsilon_1}(z) \subset T^{\varepsilon_2}(z)$ for every $z \in \mathcal{Z}$;

b) $T^\varepsilon(z) + (T')^{\varepsilon'}(z) \subset (T + T')^{\varepsilon+\varepsilon'}(z)$ for every $z \in \mathcal{Z}$ and $\varepsilon, \varepsilon' \in \mathbb{R}$;

c) T is monotone if, and only if, $T \subset T^0$;

d) T is maximal monotone if, and only if, $T = T^0$;

Note that, due to the definition of T^ε, the verification of the inclusion $v \in T^\varepsilon(z)$ requires checking an infinite number of inequalities. This verification is feasible only for specially-structured instances of operators T. However, it is possible to compute points in the graph of T^ε using the following *weak transportation formula* [2].

Theorem 2.3 ([2, Theorem 2.3])**.** *Suppose that $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is maximal monotone. Let $z_i, v_i \in \mathcal{Z}$ and $\varepsilon_i, \alpha_i \in \mathbb{R}_+$, for $i = 1, \ldots, k$, be such that*

$$v_i \in T^{\varepsilon_i}(z_i), \quad i = 1, \ldots, k, \qquad \sum_{i=1}^{k} \alpha_i = 1,$$

and define

$$z^a := \textstyle\sum_{i=1}^k \alpha_i z_i, \qquad v^a := \sum_{i=1}^k \alpha_i v_i,$$

$$\varepsilon^a := \textstyle\sum_{i=1}^k \alpha_i[\varepsilon_i + \langle z_i - z^a, v_i - v^a\rangle] = \sum_{i=1}^k \alpha_i[\varepsilon_i + \langle z_i - z^a, v_i\rangle].$$

Then, the following statements hold:

a) $\varepsilon^a \geq 0$ *and* $v^a \in T^{\varepsilon^a}(z^a)$*;*

b) if, in addition, $T = \partial f$ *for some proper lower semi-continuous convex function* f *and* $v_i \in \partial_{\varepsilon_i} f(z_i)$ *for* $i = 1, \ldots, k$*, then* $v^a \in \partial_{\varepsilon^a} f(z^a)$*.*

Finally, we refer the reader to [3, 26] for further discussion on the ε-enlargement of a maximal monotone operator.

2.2 The hybrid proximal extragradient method

This subsection reviews the HPE method and corresponding global convergence rate results obtained in [14].

Let $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ be maximal monotone operator. The monotone inclusion problem for T consists of finding $z \in \mathcal{Z}$ such that

$$0 \in T(z)\,.$$

We also assume throughout this section that this problem has a solution, that is, $T^{-1}(0) \neq \emptyset$.

We next review the hybrid proximal extragradient method introduced in [22] for solving the above problem and state the iteration-complexity results obtained for it in [14].

Hybrid Proximal Extragradient Method:

0) Let $z_0 \in \mathcal{Z}$ and $0 \leq \sigma \leq 1$ be given and set $k = 1$;

1) choose $\lambda_k > 0$ and find $\tilde{z}_k, \tilde{v}_k \in \mathcal{Z}$, $\sigma_k \in [0, \sigma]$ and $\varepsilon_k \geq 0$ such that

$$\tilde{v}_k \in T^{\varepsilon_k}(\tilde{z}_k), \qquad \|\lambda_k \tilde{v}_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma_k^2 \|\tilde{z}_k - z_{k-1}\|^2\,; \tag{5}$$

2) define $z_k = z_{k-1} - \lambda_k \tilde{v}_k$, set $k \leftarrow k + 1$, and go to step 1.

end

We now make several remarks about the HPE method. First, the HPE method does not specify how to choose λ_k and how to find $\tilde{z}_k$, $\tilde{v}_k$ and ε_k as in (5). The particular choice of λ_k and the algorithm used to compute $\tilde{z}_k, \tilde{v}_k$ and ε_k will depend on the particular implementation of the method and the properties of the operator T. Second, if $\tilde{z} := (\lambda_k T + I)^{-1} z_{k-1}$ is the *exact* proximal point iterate, or equivalently

$$\tilde{v} \in T(\tilde{z}), \tag{6}$$

$$\lambda_k \tilde{v} + \tilde{z} - z_{k-1} = 0, \tag{7}$$

for some $\tilde{v} \in \mathcal{Z}$, then $(\tilde{z}_k, \tilde{v}_k) = (\tilde{z}, \tilde{v})$ and $\varepsilon_k = 0$ satisfies (5). Therefore, the error criterion (5) relaxes the inclusion (6) to $\tilde{v} \in T^{\varepsilon}(\tilde{z})$ and relaxes equation (7) by allowing a small error relative to $\|\tilde{z}_k - z_{k-1}\|$.

We now state a few results about the convergence behaviour of the HPE method. The proof of the following result can be found in Lemma 4.2 of [14].

Proposition 2.4. *For any* $z^* \in T^{-1}(0)$*, the sequence* $\{\|z^* - z_k\|\}$ *is non-increasing and*

$$\|z^* - z_0\|^2 \geq \|z^* - z_k\|^2 + \sum_{i=1}^{k} \left[\|\tilde{z}_i - z_{i-1}\|^2 - \left(\|\lambda_i \tilde{v}_i + \tilde{z}_i - z_{i-1}\|^2 + 2\lambda_i \varepsilon_i\right)\right] \tag{8}$$

$$\geq \|z^* - z_k\|^2 + (1 - \sigma^2) \sum_{i=1}^{k} \|\tilde{z}_i - z_{i-1}\|^2\,. \tag{9}$$

The proof of the following result which establishes the convergence rate of the residual $(\tilde{v}_k, \varepsilon_k)$ of z_k can be found in Theorem 4.4 of [14].

Theorem 2.5. *Assume that $\sigma < 1$ and let d_0 be the distance of z_0 to $T^{-1}(0)$. Then, for every $k \in \mathbb{N}$, $\tilde{v}_k \in T^{\varepsilon_k}(\tilde{z}_k)$ and there exists an index $i \le k$ such that*

$$\|\tilde{v}_i\| \le d_0 \sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{1}{\sum_{j=1}^k \lambda_j^2}\right)}, \qquad \varepsilon_i \le \frac{\sigma^2 d_0^2 \lambda_i}{2(1-\sigma^2)\sum_{j=1}^k \lambda_j^2}. \tag{10}$$

Theorem 2.5 estimate the quality of the best among the iterates $\tilde{z}_1, \ldots, \tilde{z}_k$. We will refer to these estimates as the *pointwise* complexity bounds for the HPE method.

We will now describe alternative estimates for the HPE method which we refer to as the *ergodic* complexity bounds. The next result describes the convergence properties of an ergodic sequence associated with $\{\tilde{z}_k\}$.

Theorem 2.6. *For every $k \in \mathbb{N}$, define*

$$\tilde{z}_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \tilde{z}_i, \quad \tilde{v}_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \tilde{v}_i, \quad \varepsilon_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i(\varepsilon_i + \langle \tilde{z}_i - \tilde{z}_k^a, \tilde{v}_i\rangle), \tag{11}$$

where $\Lambda_k := \sum_{i=1}^k \lambda_i$. Then, for every $k \in \mathbb{N}$,

$$\tilde{v}_k^a = \frac{1}{\Lambda_k}(z_0 - z_k) \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \qquad \|\tilde{v}_k^a\| \le \frac{2d_0}{\Lambda_k},$$

and

$$0 \le \varepsilon_k^a \le \frac{1}{2\Lambda_k}\left[2\langle \tilde{z}_k^a - z_0, z_k - z_0\rangle - \|z_k - z_0\|^2\right] \le \frac{2d_0^2}{\Lambda_k}(1+\rho_k), \tag{12}$$

where d_0 is the distance of z_0 to $T^{-1}(0)$, and

$$\rho_k := \frac{1}{d_0}\|\tilde{z}_k^a - z_k^a\| \le \frac{1}{d_0}\max_{i=1,\ldots,k}\|\tilde{z}_i - z_i\|, \qquad \textit{where} \;\; z_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i z_i. \tag{13}$$

Moreover, if $\sigma < 1$ then

$$\rho_k \le \frac{\sigma\sqrt{\tau_k}}{\sqrt{(1-\sigma^2)}}, \qquad \textit{where} \;\; \tau_k = \max_{i=1,\ldots,k}\frac{\lambda_i}{\Lambda_k} \le 1. \tag{14}$$

Proof. This result follows immediately from Proposition 4.6 and the proof of Theorem 4.7 of [14]. □

Note that the rate of convergence (12) in Theorem 2.6 is only useful if we know how to bound ρ_k, which is the case when $\sigma < 1$. When $\sigma = 1$, we do not know how to bound ρ_k in the general setting of the HPE method. However, we will show that ρ_k can be bounded in special cases of the HPE method when $\sigma = 1$. Our interest in this extreme case is due to the fact that the AMAL method (see Section 7) can be viewed as a special implementation of the HPE method with $\sigma = 1$.

3 The BD-HPE framework

In this section, we introduce the BD-HPE framework for solving a special type of monotone inclusion problem consisting of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable block-structure. We also show that the BD-HPE framework can be viewed as a special case of the HPE method, and as a consequence, obtain global convergence rate results for it using the general theory outlined in Subsection 2.2.

In this section, we make the following assumptions:

A.1 $A : \mathcal{X} \rightrightarrows \mathcal{X}$ and $B : \mathcal{Y} \rightrightarrows \mathcal{Y}$ are maximal monotone operators;

A.2 $F : \operatorname{Dom} F \subset \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y}$ is a continuous map such that $\operatorname{Dom} F \supset \operatorname{cl}(\operatorname{Dom} A) \times \mathcal{Y}$;

A.3 F is monotone on $\operatorname{Dom} A \times \operatorname{Dom} B$;

A.4 there exists $L_{xy} > 0$ such that

$$\|F_x(x, y') - F_x(x, y)\| \le L_{xy}\|y' - y\|, \quad \forall x \in \operatorname{Dom} A, \ \forall y, y' \in \mathcal{Y}. \tag{15}$$

It is trivial to check that the operator $A \otimes B : \mathcal{X} \times F \rightrightarrows \mathcal{X} \times \mathcal{Y}$ defined as

$$(A \otimes B)(x, y) = A(x) \times B(y), \quad \forall (x, y) \in \mathcal{X} \times F,$$

is maximal monotone.

Our problem of interest in this section is the monotone inclusion problem of finding (x, y) such that

$$(0, 0) \in [F + (A \otimes B)](x, y), \tag{16}$$

or equivalently,

$$0 \in F_x(x, y) + A(x), \quad 0 \in F_y(x, y) + B(y),$$

where $F(x, y) = (F_x(x, y), F_y(x, y)) \in \mathcal{X} \times F$. In view of the proof of Proposition A.1 of [15], it follows that $F + (A \otimes B)$ is maximal monotone.

The exact proximal point iteration for this problem is: given $(x_{k-1}, y_{k-1}) \in \mathcal{X} \times \mathcal{Y}$, let (x_k, y_k) be the solution of the proximal inclusion subproblem:

$$\begin{aligned} 0 &\in \lambda[F_x(x, y) + A(x)] + x - x_{k-1}, \\ 0 &\in \lambda[F_y(x, y) + B(y)] + y - y_{k-1}. \end{aligned}$$

In this section, we are interested in block-decomposition methods for solving (16), where the k-th iteration consists of finding an approximate solution $\tilde{x}_k$ of the subproblem

$$0 \in \lambda[F_x(x, y_{k-1}) + A(x)] + x - x_{k-1}, \tag{17}$$

then computing an approximate solution $\tilde{y}_k$ of

$$0 \in \lambda[F_y(\tilde{x}_k, y) + B(y)] + y - y_{k-1}, \tag{18}$$

and finally using the pair $(\tilde{x}_k, \tilde{y}_k)$ to obtain the next iterate (x_k, y_k).

To formalize the above idea, we now state the BD-HPE framework that we are interested in studying in this section.

Block-decomposition HPE framework:

0) Let $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, $\sigma_x, \sigma_y \in [0,1)$ and $\sigma \in [0,1]$ be given and set $k = 1$;

1) choose $\lambda_k > 0$ such that

$$\sigma_k := \lambda_{\max}\left(\begin{bmatrix} \sigma_x^2 & \lambda_k \sigma_x L_{xy} \\ \lambda_k \sigma_x L_{xy} & \sigma_y^2 + \lambda_k^2 L_{xy}^2 \end{bmatrix}\right)^{1/2} \leq \sigma; \tag{19}$$

2) compute $\tilde{x}_k, \tilde{a}_k \in \mathcal{X}$ and $\varepsilon_k^x \geq 0$ such that

$$\tilde{a}_k \in A^{\varepsilon_k^x}(\tilde{x}_k), \quad \|\lambda_k[F_x(\tilde{x}_k, y_{k-1}) + \tilde{a}_k] + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k^x \leq \sigma_x^2 \|\tilde{x}_k - x_{k-1}\|^2; \tag{20}$$

3) compute $\tilde{y}_k, \tilde{b}_k \in \mathcal{Y}$ and $\varepsilon_k^y \geq 0$ such that

$$\tilde{b}_k \in B^{\varepsilon_k^y}(\tilde{y}_k), \quad \|\lambda_k[F_y(\tilde{x}_k, \tilde{y}_k) + \tilde{b}_k] + \tilde{y}_k - y_{k-1}\|^2 + 2\lambda_k \varepsilon_k^y \leq \sigma_y^2 \|\tilde{y}_k - y_{k-1}\|^2; \tag{21}$$

4) set

$$(x_k, y_k) = (x_{k-1}, y_{k-1}) - \lambda_k[F(\tilde{x}_k, \tilde{y}_k) + (\tilde{a}_k, \tilde{b}_k)], \tag{22}$$

$k \leftarrow k + 1$, and go to step 1.

end

Instead of using a constant σ_x and σ_y in (20) and (21), respectively, we could use variable factors $\sigma_{k,x} \leq \sigma_x$ and $\sigma_{k,y} \leq \sigma_y$, respectively, just like in the HPE method. However, for sake of simplicity, we will deal only with the case where these factors are constant.

The following result shows, under some suitable conditions on F, that any instance of the BD-HPE framework is also an instance of the HPE method of Section 2.2 applied to the monotone inclusion (16).

Proposition 3.1. *Consider the sequences $\{(x_k, y_k)\}$, $\{(\tilde{x}_k, \tilde{y}_k)\}$, $\{(\tilde{a}_k, \tilde{b}_k)\}$, $\{\lambda_k\}$ and $\{(\varepsilon_k^x, \varepsilon_k^y)\}$ generated by the BD-HPE framework. Then, for every $k \in \mathbb{N}$,*

$$F(\tilde{x}_k, \tilde{y}_k) + (\tilde{a}_k, \tilde{b}_k) \in [F + (A \otimes B)^{\varepsilon_k^x + \varepsilon_k^y}](\tilde{x}_k, \tilde{y}_k) \subset (F + A \otimes B)^{\varepsilon_k^x + \varepsilon_k^y}(\tilde{x}_k, \tilde{y}_k) \tag{23}$$

and

$$\left\|\lambda_k[F(\tilde{x}_k, \tilde{y}_k) + (\tilde{a}_k, \tilde{b}_k)] + (\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\right\|^2 + 2\lambda_k(\varepsilon_k^x + \varepsilon_k^y) \leq \sigma_k^2 \|(\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\|^2.$$

As a consequence, any instance of the BD-HPE framework is a special case of the HPE method for the inclusion problem (16) with $\tilde{v}_k = F(\tilde{x}_k, \tilde{y}_k) + (\tilde{a}_k, \tilde{b}_k)$ and $\varepsilon_k = \varepsilon_k^x + \varepsilon_k^y$ for every $k \in \mathbb{N}$.

Proof. Using the inclusions in (20) and (21), definition (4) and the definition of ε_k, we have for every $(a, b) \in (A \otimes B)(x, y)$ that

$$\langle (\tilde{x}_k, \tilde{y}_k) - (x, y), (\tilde{a}_k, \tilde{b}_k) - (a, b) \rangle = \langle \tilde{x}_k - x, \tilde{a}_k - a \rangle + \langle \tilde{y}_k - y, \tilde{b}_k - b \rangle \geq -(\varepsilon_k^x + \varepsilon_k^y),$$

which shows that $(\tilde{a}_k, \tilde{b}_k) \in (A \otimes B)^{\varepsilon_k^x + \varepsilon_k^y}(\tilde{x}_k, \tilde{y}_k)$, and hence that (23) holds, in view of statements b) and c) of Proposition 2.2. Let

$$r_k^x := \lambda_k(F_x(\tilde{x}_k, y_{k-1}) + \tilde{a}_k) + \tilde{x}_k - x_{k-1}, \qquad r_k^y := \lambda_k(F_y(\tilde{x}_k, \tilde{y}_k) + \tilde{b}_k) + \tilde{y}_k - y_{k-1} \tag{24}$$

Then,

$$\lambda_k[F(\tilde{x}_k, \tilde{y}_k) + (\tilde{a}_k, \tilde{b}_k)] + (\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1}) = (r_k^x + \lambda_k(F_x(\tilde{x}_k, \tilde{y}_k) - F_x(\tilde{x}_k, y_{k-1})), r_k^y),$$

which, together with (15), (19) and (24), and the inequalities in (20) and (21), imply

$$
\begin{aligned}
&\left\|\lambda_k[F(\tilde{x}_k,\tilde{y}_k)+(\tilde{a}_k,\tilde{b}_k)]+(\tilde{x}_k,\tilde{y}_k)-(x_{k-1},y_{k-1})\right\|^2+2\lambda_k(\varepsilon_k^x+\varepsilon_k^y)\\
&\le \|r_k^x+\lambda_k(F_x(\tilde{x}_k,\tilde{y}_k)-F_x(\tilde{x}_k,y_{k-1})\|^2+\|r_k^y\|^2+2\lambda_k(\varepsilon_k^x+\varepsilon_k^y)\\
&\le \left(\|r_k^x\|+\lambda_k\|F_x(\tilde{x}_k,\tilde{y}_k)-F_x(\tilde{x}_k,y_{k-1})\|\right)^2+\|r_k^y\|^2+2\lambda_k(\varepsilon_k^x+\varepsilon_k^y)\\
&\le \left(\|r_k^x\|+\lambda_k L_{xy}\|\tilde{y}_k-y_{k-1}\|\right)^2+\|r_k^y\|^2+2\lambda_k(\varepsilon_k^x+\varepsilon_k^y)\\
&\le \lambda_k^2L_{xy}^2\|\tilde{y}_k-y_{k-1}\|^2+2\lambda_kL_{xy}\|r_k^x\|\|\tilde{y}_k-y_{k-1}\|+(\|r_k^x\|^2+2\lambda_k\varepsilon_k^x)+(\|r_k^y\|^2+2\lambda_k\varepsilon_k^y)\\
&\le \lambda_k^2L_{xy}^2\|\tilde{y}_k-y_{k-1}\|^2+2\lambda_k\sigma_xL_{xy}\|\tilde{x}_k-x_{k-1}\|\|\tilde{y}_k-y_{k-1}\|+\sigma_x^2\|\tilde{x}_k-x_{k-1}\|^2+\sigma_y^2\|\tilde{y}_k-y_{k-1}\|^2\\
&\le \sigma_k^2\left(\|\tilde{x}_k-x_{k-1}\|^2+\|\tilde{y}_k-y_{k-1}\|^2\right)=\sigma_k^2\|(\tilde{x}_k,\tilde{y}_k)-(x_{k-1},y_{k-1})\|^2. \qquad \square
\end{aligned}
$$

We now state two iteration-complexity results for the BD-HPE framework which are direct consequences of Proposition 3.1 and Theorems 2.5 and 2.6. The first (pointwise) one is about the behaviour of the sequence $\{(\tilde{x}_k,\tilde{y}_k)\}$ and the second (ergodic) one is in regards to an ergodic sequence associated with $\{(\tilde{x}_k,\tilde{y}_k)\}$.

Theorem 3.2. *Assume that $\sigma<1$ and consider the sequences $\{(\tilde{x}_k,\tilde{y}_k)\}$, $\{(\tilde{a}_k,\tilde{b}_k)\}$, $\{\lambda_k\}$ and $\{(\varepsilon_k^x,\varepsilon_k^y)\}$ generated by the BD-HPE framework and let d_0 denote the distance of the initial point $(x_0,y_0)\in\mathcal{X}\times\mathcal{Y}$ to the solution set of (16). Then, for every $k\in\mathbb{N}$,*

$$(\tilde{a}_k,\tilde{b}_k)\in A^{\varepsilon_k^x}(\tilde{x}_k)\times B^{\varepsilon_k^y}(\tilde{y}_k),$$

and there exists $i\le k$ such that

$$\left\|F(\tilde{x}_i,\tilde{y}_i)+(\tilde{a}_i,\tilde{b}_i)\right\|\le d_0\sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{1}{\sum_{j=1}^k\lambda_j^2}\right)},\qquad \varepsilon_i^x+\varepsilon_i^y\le\frac{\sigma^2d_0^2\lambda_i}{2(1-\sigma^2)\sum_{j=1}^k\lambda_j^2}.$$

Proof. This result follows immediately from Proposition 3.1 and Theorem 2.5. $\square$

Theorem 3.3. *Consider the sequences $\{(\tilde{x}_k,\tilde{y}_k)\}$, $\{(\tilde{a}_k,\tilde{b}_k)\}$ and $\{(\varepsilon_k^x,\varepsilon_k^y)\}$ generated by the BD-HPE framework and define for every $k\in\mathbb{N}$:*

$$(\tilde{x}_k^a,\tilde{y}_k^a)=\frac{1}{\Lambda_k}\sum_{i=1}^k\lambda_i(\tilde{x}_i,\tilde{y}_i),\quad (\tilde{a}_k^a,\tilde{b}_k^a)=\frac{1}{\Lambda_k}\sum_{i=1}^k\lambda_i(\tilde{a}_i,\tilde{b}_i),\quad \tilde{F}_k^a:=\frac{1}{\Lambda_k}\sum_{i=1}^k\lambda_iF(\tilde{x}_i,\tilde{y}_i), \tag{25}$$

and

$$\varepsilon_{k,F}^a := \frac{1}{\Lambda_k}\sum_{i=1}^k\lambda_i\left\langle(\tilde{x}_i,\tilde{y}_i)-(\tilde{x}_k^a,\tilde{y}_k^a)\,,\,F(\tilde{x}_i,\tilde{y}_i)\right\rangle\ge 0, \tag{26}$$

$$\varepsilon_{k,A}^a := \frac{1}{\Lambda_k}\sum_{i=1}^k\lambda_i\left(\varepsilon_i^x+\langle\tilde{x}_i-\tilde{x}_k^a\,,\,\tilde{a}_i\rangle\right)\ge 0, \tag{27}$$

$$\varepsilon_{k,B}^a := \frac{1}{\Lambda_k}\sum_{i=1}^k\lambda_i\left(\varepsilon_i^y+\left\langle\tilde{y}_i-\tilde{y}_k^a\,,\,\tilde{b}_i\right\rangle\right)\ge 0. \tag{28}$$

Let d_0 denote the distance of the initial point $(x_0,y_0)\in\mathcal{X}\times\mathcal{Y}$ to the solution set of (16). Then, for every $k\in\mathbb{N}$,

$$\tilde{F}_k^a\in F^{\varepsilon_{k,F}^a}(\tilde{x}_k^a,\tilde{y}_k^a),\quad (\tilde{a}_k^a,\tilde{b}_k^a)\in A^{\varepsilon_{k,A}^a}(\tilde{x}_k^a)\times B^{\varepsilon_{k,B}^a}(\tilde{y}_k^a), \tag{29}$$

and

$$\left\|\tilde{F}_k^a+(\tilde{a}_k^a,\tilde{b}_k^a)\right\|\le\frac{2d_0}{\Lambda_k},\qquad \varepsilon_{k,F}^a+\varepsilon_{k,A}^a+\varepsilon_{k,B}^a\le\frac{2d_0^2}{\Lambda_k}(1+\eta_k), \tag{30}$$

where

$$\eta_k := \frac{2}{1-\sigma_{xy}}\left(1+\frac{1}{(1-\sigma_y)^2}\right)^{1/2}\sqrt{\sigma_x^2+\sigma_y^2+\lambda_k^2 L_{xy}^2} \;\le\; \frac{2\sqrt{2}\,\sigma}{1-\sigma_{xy}}\left(1+\frac{1}{(1-\sigma_y)^2}\right)^{1/2}. \tag{31}$$

Moreover, if F is affine, then $\tilde{F}_k^a = F(\tilde{x}_k^a, \tilde{y}_k^a)$. Also, if A (resp., B) is affine and $\varepsilon_k^x = 0$ (resp., $\varepsilon_k^y = 0$) for every $k \in \mathbb{N}$, then $\tilde{a}_k^a \in A(\tilde{x}_k^a)$ (resp., $\tilde{b}_k^a \in B(\tilde{y}_k^a)$).

Proof. First, note that (29) follows immediately from the definitions in (25), (26), (27) and (28), the inclusions in (20) and (21), and Theorem 2.3. In view of Proposition 3.1, any instance of the BD-HPE framework is a special case of the HPE method with $\tilde{v}_k = F(\tilde{x}_k, \tilde{y}_k) + (\tilde{a}_k, \tilde{b}_k)$ and $\varepsilon_k = \varepsilon_k^x + \varepsilon_k^y$ for every $k \in \mathbb{N}$. Since, in this case, the quantities $\tilde{z}_k^a$, $\tilde{v}_k^a$ and ε_k^a defined in Theorem 2.6 are equal to $(\tilde{x}_k^a, \tilde{y}_k^a)$, $F_k^a + (\tilde{a}_k^a, \tilde{b}_k^a)$ and $\varepsilon_{x,F}^a + \varepsilon_{x,F}^a + \varepsilon_{x,F}^a$, respectively, it follows from the conclusions of this theorem that the first inequality in (30) holds and

$$\varepsilon_{x,F}^a + \varepsilon_{x,F}^a + \varepsilon_{x,F}^a \le \frac{2d_0^2}{\Lambda_k}(1+\rho_k), \tag{32}$$

where

$$\rho_k := \frac{1}{d_0}\max_{i=1,\ldots,k} \|(\tilde{x}_i, \tilde{y}_i) - (x_i, y_i)\|. \tag{33}$$

Noting the definition of η_k in (31), we now claim that $\rho_k \le \eta_k$, which, together with (32), clearly implies the second inequality in (30). Indeed, let (x^*, y^*) be a solution of (16) such that $\|(x_0, y_0) - (x^*, y^*)\| = d_0$. Due to Proposition 2.4, we know that

$$\begin{aligned}\|(x_k, y_k) - (x_{k-1}, y_{k-1})\| &\le \|(x_k, y_k) - (x^*, y^*)\| + \|(x_{k-1}, y_{k-1}) - (x^*, y^*)\| \\ &\le \|(x_0, y_0) - (x^*, y^*)\| + \|(x_0, y_0) - (x^*, y^*)\| = 2d_0.\end{aligned} \tag{34}$$

It follows from (21) and (22) that

$$\|\tilde{y}_k - y_{k-1}\| \le \|\tilde{y}_k - y_k\| + \|y_k - y_{k-1}\| \le \sigma_y\|\tilde{y}_k - y_{k-1}\| + \|y_k - y_{k-1}\|,$$

and hence that

$$\|\tilde{y}_k - y_{k-1}\| \le \frac{\|y_k - y_{k-1}\|}{1-\sigma_y}, \qquad \|\tilde{y}_k - y_k\| \le \sigma_y\|\tilde{y}_k - y_{k-1}\| \le \frac{\sigma_y\,\|y_k - y_{k-1}\|}{1-\sigma_y}. \tag{35}$$

Also, it follows from (15), (20) and (22) that

$$\begin{aligned}\|\tilde{x}_k - x_k\| - \lambda_k L_{xy}\|\tilde{y}_k - y_{k-1}\| &\le \|\tilde{x}_k - x_k + \lambda_k[F_x(\tilde{x}_k, y_{k-1}) - F_x(\tilde{x}_k, \tilde{y}_k)]\| \\ &\le \sigma_x\|\tilde{x}_k - x_{k-1}\| \le \sigma_x(\|\tilde{x}_k - x_k\| + \|x_k - x_{k-1}\|),\end{aligned}$$

and hence that

$$\|\tilde{x}_k - x_k\| \le \frac{\sigma_x\,\|x_k - x_{k-1}\| + \lambda_k L_{xy}\|\tilde{y}_k - y_{k-1}\|}{1-\sigma_x}. \tag{36}$$

Adding the second inequality in (35) to inequality (36) and noting (34), we conclude that

$$\begin{aligned}\|(\tilde{x}_k, \tilde{y}_k) - (x_k, y_k)\| &\le \|\tilde{x}_k - x_k\| + \|\tilde{y}_k - y_k\| \\ &\le \frac{1}{1-\sigma_{xy}}\left(\sigma_x\,\|x_k - x_{k-1}\| + \sigma_y\,\|y_k - y_{k-1}\| + \lambda_k L_{xy}\|\tilde{y}_k - y_{k-1}\|\right) \\ &\le \frac{1}{1-\sigma_{xy}}\sqrt{\sigma_x^2+\sigma_y^2+\lambda_k^2 L_{xy}^2}\,\sqrt{\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2 + \|\tilde{y}_k - y_{k-1}\|^2} \\ &\le \frac{1}{1-\sigma_{xy}}\sqrt{\sigma_x^2+\sigma_y^2+\lambda_k^2 L_{xy}^2}\,\sqrt{4d_0^2 + (1-\sigma_y)^{-2}\|y_k - y_{k-1}\|^2} \\ &\le \frac{2d_0}{1-\sigma_{xy}}\left(1+\frac{1}{(1-\sigma_y)^2}\right)^{1/2}\sqrt{\sigma_x^2+\sigma_y^2+\lambda_k^2 L_{xy}^2}.\end{aligned}$$

The last estimate together with (31) and (33) clearly imply our claim that $\rho_k \le \eta_k$. Finally, note that the second inequality in (31) follows from (19) and the assumption that $\sigma_k \le \sigma$, and that the last assertion of the theorem follows from implication (1) about an affine map T. $\square$

We observe that, if we had assumed in Theorem 3.3 that $\sigma < 1$, then its proof would be much simpler since in this case we could have used the last assertion of Theorem 2.6. However, as observed earlier, our interest in the case where $\sigma = 1$ is due to the fact that the AMAL method (see Section 7) can be viewed as an instance of the HPE method with $\sigma = 1$. To handle the case $\sigma = 1$, the proof of Theorem 3.3 uses inequality (12), which depends on the quantity ρ_k. As shown in this proof, all that is required to bound ρ_k is the assumption that $\sigma \le 1$ and $\max\{\sigma_x, \sigma_y\} < 1$.

4 Approximate solutions of the proximal subproblems

In this section, we describe some specific procedures for finding approximate solutions $(\tilde{x}_k, \tilde{a}_k, \varepsilon_k^x)$ and $(\tilde{y}_k, \tilde{b}_k, \varepsilon_k^y)$ of (17) and (18) according to steps 1 and 2, respectively, of the BD-HPE framework. We should emphasize that such solutions can be found by other procedures which are not discussed below.

The problem of finding approximate solutions as above can be cast in the following general form. Throughout this section, we assume that

B.1) $C : \mathcal{X} \rightrightarrows \mathcal{X}$ is a maximal monotone operator;

B.2) $G : \operatorname{Dom} G \subset \mathcal{X} \to \mathcal{X}$ is a continuous map which is monotone on $\operatorname{cl}(\operatorname{Dom} C) \subset \operatorname{Dom} G$.

Given a point $x \in \mathcal{X}$, $\lambda > 0$ and $\sigma \in [0, 1)$, our goal is to describe specific procedures for computing a pair $(\tilde{x}, \tilde{c}) \in \mathcal{X} \times \mathcal{X}$ and a scalar $\varepsilon \ge 0$ such that

$$\tilde{c} \in C^{\varepsilon}(\tilde{x}), \quad \|\lambda(G(\tilde{x}) + \tilde{c}) + \tilde{x} - x\|^2 + 2\lambda\varepsilon \le \sigma^2 \|\tilde{x} - x\|^2. \tag{37}$$

We note that conditions B.1 and B.2 imply that $G + C$ is a maximal monotone operator (see the proof of Proposition A.1 of [15]). This implies that, for any $\lambda > 0$, the resolvent of $G + C$, namely the map $[I + \lambda(G + C)]^{-1}$ is a single-valued map defined over the whole $\mathcal{X}$. The following simple result shows that when the resolvent of $G + C$ is computable, (37) can be solved exactly.

Proposition 4.1. *Let $x \in \mathcal{X}$ and $\lambda > 0$ be given. Then,*

$$\tilde{x} := [\lambda(G + C) + I]^{-1}(x), \qquad \tilde{c} := \frac{1}{\lambda}(x - \tilde{x}) - G(\tilde{x}), \qquad \varepsilon = 0 \tag{38}$$

satisfy (37) with $\sigma = 0$.

Proof. Using the three identities in (38), we easily see that $\tilde{c} \in C(\tilde{x})$ and

$$\|\lambda(G(\tilde{x}) + \tilde{c}) + \tilde{x} - x\|^2 + 2\lambda\varepsilon = 0 \le \sigma^2 \|\tilde{x} - x\|^2.$$

$\square$

Proposition 4.2. *Assume that $C = \partial f + T$, where $T : \mathcal{X} \rightrightarrows \mathcal{X}$ is maximal monotone and $f : \mathcal{X} \to \bar{\mathbb{R}}$ is a proper closed convex function such that f is differentiable on $\operatorname{cl}(\operatorname{dom} T) \subset \operatorname{int}(\operatorname{dom} f)$, ∇f is L-Lipschitz continuous on $\operatorname{cl}(\operatorname{dom} T)$. Also, let $\sigma > 0$, $x \in \operatorname{Dom} T$ and $\lambda \in (0, \sigma^2/L]$ be given. Then,*

$$\tilde{x} = [I + \lambda(G + T)]^{-1}(x - \lambda \nabla f(x)), \quad \tilde{c} = \frac{1}{\lambda}(x - \tilde{x}) - G(\tilde{x}), \quad \varepsilon = \frac{L}{2}\|\tilde{x} - x\|^2. \tag{39}$$

satisfy (37). Moreover, $\tilde{c} \in (\partial_{\varepsilon} f + T)(\tilde{x})$.

Proof. First observe that the last two identities in (39) and the assumption that $\lambda \in (0, 2\sigma^2/L]$ imply that

$$\|\lambda(G(\tilde{x}) + \tilde{c}) + \tilde{x} - x\|^2 + 2\lambda\varepsilon = 2\lambda\varepsilon \le 2\left(\frac{\sigma^2}{L}\right)\left(\frac{L}{2}\|\tilde{x} - x\|^2\right) = \sigma^2\|x - \tilde{x}\|^2.$$

It remains to show that $\tilde{c} \in C^{\varepsilon}(\tilde{x})$. Using the definition of $\tilde{x}$, we have

$$\frac{1}{\lambda}(x - \tilde{x}) - \nabla f(x) \in (G + T)(\tilde{x}),$$

and hence, $\tilde{c} \in \nabla f(x) + T(\tilde{x})$, due to the definition of $\tilde{c}$. We now claim that $\nabla f(x) \in \partial_{\varepsilon} f(\tilde{x})$, from which we conclude that

$$\tilde{c} \in (\partial_{\varepsilon} f + T)(\tilde{x}) \subset [(\partial f)^{\varepsilon} + T](\tilde{x}) \subset (\partial f + T)^{\varepsilon}(\tilde{x}) = C^{\varepsilon}(\tilde{x}),$$

where the second and third inclusions follow from Proposition 2.1(a) and Proposition 2.2, and the equality follows from the definition of C. To prove the claim, note that Proposition 2.1(c) with $v = \nabla f(x)$ implies that $\nabla f(x) \in \partial_{\varepsilon'} f(\tilde{x})$, where

$$\varepsilon' := f(\tilde{x}) - f(x) - \langle \nabla f(x), \tilde{x} - x\rangle \le \frac{L}{2}\|\tilde{x} - x\|^2 =: \varepsilon,$$

where the inequality is due to the fact that ∇f is L-Lipschitz continuous on $\mathrm{cl}(\mathrm{Dom}\, T) \supset \mathrm{Dom}\, T \ni \tilde{x}, x$, and $\mathrm{cl}(\mathrm{Dom}\, T)$ is convex. □

Proposition 4.3. *Assume that $G : \mathcal{X} \to \mathcal{X}$ is L-Lipschitz continuous on a closed convex set Ω such that $\mathrm{Dom}\, C \subset \Omega \subset \mathrm{Dom}\, G$. Let $\sigma > 0$, $x \in \mathcal{X}$ and $\lambda \in (0, \sigma/L]$ be given. Then,*

$$\tilde{x} = (I + \lambda C)^{-1}(x - \lambda G(P_{\Omega}(x))), \quad \tilde{c} = \frac{1}{\lambda}(x - \tilde{x}) - G(P_{\Omega}(x)), \quad \varepsilon = 0 \tag{40}$$

satisfy (37).

Proof. First note that the first two identities in (40) imply that $\tilde{c} \in C(\tilde{x})$. Also, (40) and the assumption that $\lambda \le \sigma/L$ and G is L-Lipschitz continuous on $\Omega \ni \tilde{x}$ imply that

$$\begin{aligned}\|\lambda(G(\tilde{x}) + \tilde{c}) + \tilde{x} - x\|^2 + 2\lambda\varepsilon &= \|\lambda(G(\tilde{x}) + \tilde{c}) + \tilde{x} - x\|^2 = \lambda^2\|G(\tilde{x}) - G(P_{\Omega}(x))\|^2 \\ &\le \lambda^2 L^2\|\tilde{x} - P_{\Omega}(x)\|^2 \le \sigma^2\|P_{\Omega}(\tilde{x}) - P_{\Omega}(x)\|^2 \le \sigma^2\|\tilde{x} - x\|^2.\end{aligned}$$

We have thus shown that $(\tilde{x}, \tilde{c}, \lambda)$ and $\varepsilon = 0$ satisfy (37). □

Note that the above formula for $\tilde{x}$ is in terms of the resolvent $(I + \lambda C)^{-1}$ of C, which must be easily computable so that $\tilde{x}$ can be obtained. Observe also that for the case where $C = N_X$ for some closed convex set $X \subset \mathcal{X}$, we have $(I + \lambda N_X)^{-1} = P_X$ and the above expression for $\tilde{x}$ reduces to

$$\tilde{x} = P_X(x - \lambda G(P_{\Omega}(x))).$$

Proposition 4.4. *Assume that $C = \partial g$, where $g : \mathcal{X} \to (-\infty, \infty]$ is a closed proper convex function and G is L-Lipschitz continuous on $\mathrm{dom}\, g$. Let $\sigma > 0$, $x \in \mathrm{dom}\, g$ and $\lambda \in (0, \sigma/L]$ be given. Then,*

$$\tilde{x} = (I + \lambda\partial g)^{-1}(x - \lambda G(x)), \quad \tilde{c} = \frac{1}{\lambda}[x - x^+] - G(\tilde{x}), \quad \varepsilon := g(\tilde{x}) - [g(x^+) + \langle \tilde{x} - x^+, \tilde{c}\rangle], \tag{41}$$

where

$$x^+ := (I + \partial g)^{-1}(x - \lambda G(\tilde{x})), \tag{42}$$

satisfy (37).

Proof. First observe that ε is well-defined since $\tilde{x}, x^+ \in \operatorname{dom} g$, in view of their definition in (41) and (42), respectively. We first prove that $\tilde{c} \in C^\varepsilon(\tilde{x})$. Indeed, the definition of $\tilde{c}$ and x^+ in (41) and (42), respectively, imply that $\tilde{c} \in \partial g(x^+)$. Hence, it follows from Proposition 2.1(c) and the definition of ε in (41) that $\tilde{c} \in \partial_\varepsilon g(\tilde{x})$, and hence that $\tilde{c} \in (\partial g)^\varepsilon(\tilde{x}) = C^\varepsilon(\tilde{x})$, in view of Proposition 2.1(a).

To show the inequality in (37), define

$$p = \frac{1}{\lambda}\left[x - \lambda G(x) - \tilde{x}\right]. \tag{43}$$

Using the definition of $\tilde{x}$ and p in (41) and (43), respectively, we conclude that $p \in \partial g(\tilde{x})$. This fact and the last identity in (41) then imply that

$$\varepsilon = g(\tilde{x}) - g(x^+) - \langle \tilde{c}, \tilde{x} - x^+\rangle = -[g(x^+) - g(\tilde{x}) - \langle p, x^+ - \tilde{x}\rangle] + \langle p - \tilde{c}, \tilde{x} - x^+\rangle \le \langle p - \tilde{c}, \tilde{x} - x^+\rangle.$$

This, together with the second identity in (41), then imply that

$$\begin{aligned}\|\lambda(G(\tilde{x}) + \tilde{c}) + \tilde{x} - x\|^2 + 2\lambda\varepsilon &= \|\tilde{x} - x^+\|^2 + 2\lambda\varepsilon \le \|\tilde{x} - x^+\|^2 + 2\lambda\langle p - \tilde{c}, \tilde{x} - x^+\rangle \\ &= \|\lambda(p - \tilde{c}) + \tilde{x} - x^+\|^2 - \lambda^2\|p - \tilde{c}\|^2 \le \|\lambda(p - \tilde{c}) + \tilde{x} - x^+\|^2 \\ &= \|\lambda(G(x) - G(\tilde{x}))\|^2 \le (\lambda L\|x - \tilde{x}\|) \le \sigma^2\|x - \tilde{x}\|^2,\end{aligned}$$

where the last equality follows from (43) and the second identity in (41), and the last two inequalities are due to the assumption that $\lambda \le \sigma/L$ and G is L-Lipschitz continuous on $\operatorname{dom} g \ni x, \tilde{x}$. □

5 Specific examples of BD-HPE methods

The goal of this section is to illustrate how the different procedures discussed in Section 4 for constructing triples $(\tilde{x}_k, \tilde{a}_k, \varepsilon_k^x)$ and $(\tilde{y}_k, \tilde{b}_k, \varepsilon_k^y)$ satisfying (20) and (21), respectively, can be used to obtain specific instances of the BD-HPE framework presented in Section 3. This section is divided into three subsections. In the first one, we discuss a specific instance of the the BD-HPE framework in which (17) and (18) are both solved exactly (see Proposition 4.1). In the second subsection, we give another instance of BD-HPE framework in which these two proximal subproblems are approximately solved by means of Tseng's scheme presented in Proposition 4.3. In the third subsection we study a block-decomposition method for a large class of linearly constrained convex optimization problems, which includes cone programs whose objective functions converge to infinity as the relative boundaries of their domain are approached.

5.1 Exact BD-HPE method

In this subsection, we consider a special case of the general BD-HPE framework where the subproblems (17) and (18) are solved exactly and specialize the iteration-complexity bounds of Theorems 3.2 and 3.3 to the current setting.

In this subsection, we assume that we know how to solve the proximal subproblems (17) and (18) exactly. More precisely, we consider the following special case of the BD-HPE framework.

Exact BD-HPE method:

0) Let $(x_0, \tilde{y}_0) \in \mathcal{X} \times \mathcal{Y}$ and $\sigma \in (0, 1]$ be given and set $\lambda = \sigma/L_{xy}$ and $k = 1$;

1) compute $(\tilde{x}_k, \tilde{y}_k) \in \mathcal{X} \times \mathcal{Y}$ as

$$\tilde{x}_k = [I + \lambda(F_x(\cdot, \tilde{y}_{k-1}) + A)]^{-1}(x_{k-1}), \qquad \tilde{y}_k = [I + \lambda(F_y(\tilde{x}_k, \cdot) + B)]^{-1}(\tilde{y}_{k-1}). \tag{44}$$

2) set $x_k = \tilde{x}_k - \lambda[F_x(\tilde{x}_k, \tilde{y}_k) - F_x(\tilde{x}_k, \tilde{y}_{k-1})]$ and $k \leftarrow k + 1$, and go to step 1;

end

The following result shows that the above algorithm is indeed a special case of the BD-HPE framework in which subproblems (17) and (18) are solved exactly (see Proposition 4.1).

Lemma 5.1. *For each $k \in \mathbb{N}$, define*

$$\tilde{a}_k = \frac{1}{\lambda}(x_{k-1} - \tilde{x}_k) - F_x(\tilde{x}_k, \tilde{y}_{k-1}), \qquad \tilde{b}_k = \frac{1}{\lambda}(\tilde{y}_{k-1} - \tilde{y}_k) - F_y(\tilde{x}_k, \tilde{y}_k), \qquad y_{k-1} = \tilde{y}_{k-1} \tag{45}$$

Then, the sequences $\{(x_k, y_k)\}$, $\{(\tilde{x}_k, \tilde{y}_k)\}$ and $\{(\tilde{a}_k, \tilde{b}_k)\}$ satisfy (20), (21) and (22) with $\varepsilon_k^x = \varepsilon_k^y = 0$, $\lambda_k = \lambda$ and $\sigma_x = \sigma_y = 0$. As a consequence, the exact BD-HPE method is a special case of the BD-HPE framework.

Proof. Using the fact that $y_{k-1} = \tilde{y}_{k-1}$ and applying Proposition 4.1 twice to the pairs $(G, C) = (F(\cdot, \tilde{y}_{k-1}), A)$ and $(G, C) = (F(\tilde{x}_k, \cdot), B)$, we conclude that $(\tilde{x}_k, \tilde{a}_k)$ and $(\tilde{y}_k, \tilde{b}_k)$ satisfy (20) and (21) with $\varepsilon_k^x = \varepsilon_k^y = 0$, $\lambda_k = \lambda$ and $\sigma_x = \sigma_y = 0$. Moreover, (22) follows from (45) and the update rule of x_k in step 2 of the exact BD-HPE method. ◻

The following result, which is an immediate consequence of the previous result and Theorem 3.3, establishes the iteration-complexity of the exact BD-HPE method.

Theorem 5.2. *Consider the sequences $\{x_k\}$ and $\{(\tilde{x}_k, \tilde{y}_k)\}$ generated by the exact BD-HPE method, and define the sequence $\{(\tilde{a}_k, \tilde{b}_k)\}$ according to (45). Moreover, for each $k \in \mathbb{N}$, define:*

$$(\tilde{x}_k^a, \tilde{y}_k^a) = \frac{1}{k}\sum_{i=1}^{k}(\tilde{x}_i, \tilde{y}_i), \quad (\tilde{a}_k^a, \tilde{b}_k^a) = \frac{1}{k}\sum_{i=1}^{k}(\tilde{a}_i, \tilde{b}_i), \quad \tilde{F}_k^a := \frac{1}{k}\sum_{i=1}^{k} F(\tilde{x}_i, \tilde{y}_i),$$

and

$$\varepsilon_{k,F}^a := \tfrac{1}{k}\textstyle\sum_{i=1}^{k} \langle (\tilde{x}_i, \tilde{y}_i) - (\tilde{x}_k^a, \tilde{y}_k^a)\,,\, F(\tilde{x}_i, \tilde{y}_i)\rangle \geq 0,$$
$$\varepsilon_{k,A}^a := \tfrac{1}{k}\textstyle\sum_{i=1}^{k} (\langle \tilde{x}_i - \tilde{x}_k^a\,,\, \tilde{a}_i\rangle) \geq 0, \qquad \varepsilon_{k,B}^a := \tfrac{1}{k}\textstyle\sum_{i=1}^{k} \left(\left\langle \tilde{y}_i - \tilde{y}_k^a\,,\, \tilde{b}_i\right\rangle\right) \geq 0.$$

Let d_0 denote the distance of the initial point $(x_0, \tilde{y}_0) \in \mathcal{X} \times \mathcal{Y}$ to the solution set of (16). Then, for every $k \in \mathbb{N}$, the following statements hold:

a) $(\tilde{a}_k, \tilde{b}_k) \in A(\tilde{x}_k) \times B(\tilde{y}_k)$, *and if $\sigma < 1$, there exists $i \leq k$ such that*

$$\left\| F(\tilde{x}_i, \tilde{y}_i) + (\tilde{a}_i, \tilde{b}_i) \right\| \leq \frac{L_{xy} d_0}{\sigma\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}};$$

b) we have

$$\tilde{F}_k^a \in F^{\varepsilon_{k,F}^a}(\tilde{x}_k^a, \tilde{y}_k^a), \quad (\tilde{a}_k^a, \tilde{b}_k^a) \in A^{\varepsilon_{k,A}^a}(\tilde{x}_k^a) \times B^{\varepsilon_{k,B}^a}(\tilde{y}_k^a),$$

and

$$\left\| \tilde{F}_k^a + (\tilde{a}_k^a, \tilde{b}_k^a) \right\| \leq \frac{2L_{xy} d_0}{k\sigma}, \qquad \varepsilon_{k,F}^a + \varepsilon_{k,A}^a + \varepsilon_{k,B}^a \leq \frac{2L_{xy} d_0^2}{k\sigma}\left(1 + 2\sqrt{2}\sigma\right).$$

Also, if F is affine, then $\tilde{F}_k^a = F(\tilde{x}_k^a, \tilde{y}_k^a)$. In addition, if A (resp., B) is affine, then $\tilde{a}_k^a \in A(\tilde{x}_k^a)$ (resp., $\tilde{b}_k^a \in B(\tilde{y}_k^a)$).

Proof. This result follows immediately from Lemma 5.1 and Theorems 3.2 and 3.3 by specializing the latter two results to the case where $\sigma_x = \sigma_y = 0$, $\lambda_k = \lambda := \sigma/L_{xy}$ and $\varepsilon_k^x = \varepsilon_k^y = 0$. ◻

5.2 An inexact BD-HPE method based on Tseng's procedure

In this subsection, we describe an inexact BD-HPE method based on Tseng's procedure described in Proposition 4.3.

We start by describing the general assumptions of this subsection. In addition to conditions A.1) to A.4) of Subsection 3, we also impose the following condition:

A.5 there exist scalars $L_{xx}, L_{yy} \geq 0$ and a closed convex set Ω_x such that $\operatorname{Dom} A \subset \Omega_x$, $\Omega_x \times \mathcal{Y} \subset \operatorname{dom} F$, and:

- $F_x(\cdot, y)$ is L_{xx}-Lipschitz continuous on Ω_x for every $y \in \mathcal{Y}$;
- $F_y(x, \cdot)$ is L_{yy}-Lipschitz continuous on $\mathcal{Y}$ for every $x \in \Omega_x$.

Tseng's based inexact BD-HPE method:

0) Let $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, $\sigma \in (0, 1]$ and $\lambda \in (0, \sigma/\tilde{L}]$ be given, where

$$\tilde{L} := \left(\lambda_{\max}\begin{bmatrix} L_{xx}^2 & L_{xx}L_{xy} \\ L_{xx}L_{xy} & L_{yy}^2 + L_{xy}^2 \end{bmatrix}\right)^{1/2}, \tag{46}$$

and set $k = 1$;

1) set $x'_{k-1} := P_{\Omega_x}(x_{k-1})$ and compute $(\tilde{x}_k, \tilde{y}_k) \in \mathcal{X} \times \mathcal{Y}$ as

$$\tilde{x}_k = [I + \lambda A]^{-1}(x_{k-1} - \lambda F_x(x'_{k-1}, y_{k-1})), \qquad \tilde{y}_k = [I + \lambda B]^{-1}(y_{k-1} - F_y(\tilde{x}_k, y_{k-1})); \tag{47}$$

2) compute (x_k, y_k) as

$$x_k := \tilde{x}_k - \lambda[F_x(\tilde{x}_k, \tilde{y}_k) - F_x(x'_k, y_{k-1})], \qquad y_k := \tilde{y}_k - \lambda[F_y(\tilde{x}_k, \tilde{y}_k) - F_y(\tilde{x}_k, y_{k-1})], \tag{48}$$

set $k \leftarrow k + 1$, and go to step 1.

end

It is easy to see that (46) and the assumption that $L_{xy} > 0$ imply that

$$\theta := \frac{1}{\tilde{L}} \max\{L_{xx}, L_{yy}\} < 1, \tag{49}$$

Proposition 5.3. *Tseng's based inexact BD-HPE method is a special instance of the BD-HPE framework, where $\sigma_x = \lambda L_{xx}$ and $\sigma_y = \lambda L_{yy}$, and for every $k \in \mathbb{N}$,*

$$\lambda_k = \lambda, \qquad \varepsilon_k^x = \varepsilon_k^y = 0,$$

and

$$\tilde{a}_k = \frac{1}{\lambda}(x_{k-1} - \tilde{x}_k) - F_x(x'_{k-1}, y_{k-1}), \quad \tilde{b}_k = \frac{1}{\lambda}(y_{k-1} - \tilde{y}_k) - F_y(\tilde{x}_k, y_{k-1}). \tag{50}$$

Proof. Applying Proposition 4.3 to the quintuple $(G, C, \Omega, x, \sigma) = (F_x(\cdot, y_{k-1}), A, \Omega_x, x_{k-1}, \lambda L_{xx})$, and also to the quintuple $(G, C, \Omega, x, \sigma) = (F_y(\tilde{x}_k, \cdot), B, \mathcal{Y}, y_{k-1}, \lambda L_{yy})$, and noting (48) and the definition of $\tilde{a}_k$ and $\tilde{b}_k$, we conclude that (22) holds and that $(\tilde{x}_k, \tilde{a}_k)$ and $(\tilde{y}_k, \tilde{b}_k)$ satisfy (20) and (21), respectively, with $\sigma_x = \lambda L_{xx}$, $\sigma_y = \lambda L_{yy}$ and $\varepsilon_k^x = \varepsilon_k^y = 0$. It remains to show that $\lambda_k = \lambda$ satisfies (19) and that $\max\{\sigma_x, \sigma_y\} < 1$. Indeed, using the fact that $\sigma_x = \lambda L_{xx}$ and $\sigma_y = \lambda L_{yy}$, the definition of $\tilde{L}$ in (46), the assumption that $\lambda \leq \sigma/\tilde{L}$ and $\sigma \leq 1$, and (49), we easily see that (19) holds and that

$$\max\{\sigma_x, \sigma_y\} = \lambda \max\{L_{xx}, L_{yy}\} \leq \frac{\sigma}{\tilde{L}} \max\{L_{xx}, L_{yy}\} \leq \frac{1}{\tilde{L}} \max\{L_{xx}, L_{yy}\} = \theta < 1. \qquad \square$$

The following convergence rate result now follows as an immediate consequence of Proposition 5.3 and Theorems 3.2 and 3.3.

Theorem 5.4. *Consider the sequences* $\{(x_k, y_k)\}$, $\{(\tilde{x}_k, \tilde{y}_k)\}$ *generated by Tseng's based inexact BD-HPE method with* $\lambda = \sigma/\tilde{L}$, *where* $\tilde{L}$ *is given by (46). Define the sequence* $\{(\tilde{a}_k, \tilde{b}_k)\}$ *according to (50) and the sequences* $\{(\tilde{x}_k^a, \tilde{y}_k^a)\}$, $\{(\tilde{a}_k^a, \tilde{b}_k^a)\}$, $\{\tilde{F}_k^a\}$, $\{\varepsilon_{k,F}^a\}$, $\{\varepsilon_{k,A}^a\}$ *and* $\{\varepsilon_{k,B}^a\}$ *as in Theorem 5.2. Let* d_0 *denote the distance of the initial point* $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$ *to the solution set of (16). Then, for every* $k \in \mathbb{N}$, *the following statements hold:*

a) $(\tilde{a}_k, \tilde{b}_k) \in A(\tilde{x}_k) \times B(\tilde{y}_k)$, *and if* $\sigma < 1$, *there exists* $i \leq k$ *such that*

$$\left\|F(\tilde{x}_i, \tilde{y}_i) + (\tilde{a}_i, \tilde{b}_i)\right\| \leq \frac{\tilde{L}d_0}{\sigma\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}};$$

b) we have

$$\tilde{F}_k^a \in F^{\varepsilon_{k,F}^a}(\tilde{x}_k^a, \tilde{y}_k^a), \quad (\tilde{a}_k^a, \tilde{b}_k^a) \in A^{\varepsilon_{k,A}^a}(\tilde{x}_k^a) \times B^{\varepsilon_{k,B}^a}(\tilde{y}_k^a),$$

and

$$\left\|\tilde{F}_k^a + (\tilde{a}_k^a, \tilde{b}_k^a)\right\| \leq \frac{2\tilde{L}d_0}{k\sigma}, \qquad \varepsilon_k^a \leq \frac{2\tilde{L}d_0^2}{k\sigma}(1+\eta),$$

where

$$\eta := \frac{2\sqrt{2}\sigma}{1-\theta\sigma}\left(1 + \frac{1}{(1-\theta\sigma)^2}\right)^{1/2}$$

and θ *is defined in (49).*

Also, if F *is affine, then* $\tilde{F}_k^a = F(\tilde{x}_k^a, \tilde{y}_k^a)$. *In addition, if* A *(resp.,* B*) is affine, then* $\tilde{a}_k^a \in A(\tilde{x}_k^a)$ *(resp.,* $\tilde{b}_k^a \in B(\tilde{y}_k^a)$*).*

Proof. This result follows immediately from Proposition 5.3 and Theorems 3.2 and 3.3 by specializing the latter two results to the case where $\lambda_k = \lambda := \sigma/\tilde{L}$, $\sigma_x = \lambda L_{xx}$, $\sigma_y = \lambda L_{yy}$, and $\varepsilon_k^x = \varepsilon_k^y = 0$, and using the fact that $\max\{\sigma_x, \sigma_y\} \leq \sigma\theta$ (see the proof of Proposition 5.3). □

We observe that it is possible to transform the bounds in terms of $\tilde{L}$ in the above result to bounds in terms of the quantities L_{xx}, L_{yy} and L_{xy} by using the estimate

$$\tilde{L} \leq \sqrt{L_{xx}^2 + L_{xy}^2 + L_{yy}^2} \leq \sqrt{2}\,\tilde{L},$$

which follows immediately from the definition of $\tilde{L}$ in (46).

5.3 An inexact block-decomposition for convex optimization

In this subsection, we are interested in developing a specific instance of the BD-HPE framework for solving a class of linearly constrained convex optimization.

In this subsection, we consider the following optimization problem:

$$\min\{f(y) + h(y) : Cy = d\}, \tag{51}$$

where the following assumptions are made:

O.1) $C : \mathcal{Y} \to \mathcal{X}$ is a nonzero linear map and $d \in \mathcal{X}$;

O.2) $f, h : \mathcal{Y} \to \bar{\mathbb{R}}$ are proper closed convex functions;

O.3) $\text{dom}(g) \subset \text{dom}(f)$ and there exist a point $\hat{y} \in \text{ri}(\text{dom}\, g) \cap \text{ri}(\text{dom}\, f)$ such that $C\hat{y} = d$;

O.4) the solution set of (51) is non-empty;

O.5) f is differentiable on $\mathrm{cl}(\mathrm{dom}\, h)$ and ∇f is L-Lipschitz continuous on $\mathrm{cl}(\mathrm{dom}\, h)$.

We now make some observations. First, under the above assumptions, y^* is an optimal solution if, and only if, it satisfies the condition

$$0 \in \partial f(y) + \partial h(y) + N_{\mathcal{M}}(y), \tag{52}$$

where $\mathcal{M} := \{y \in \mathcal{Y} : Cy = d\}$. Second, the above assumptions also guarantee that $\partial f + \partial g + N_{\mathcal{M}}$ is maximal monotone.

Clearly, y^* is an optimal solution if, and only if, there exists $x^* \in \Re^m$ such that the pair $(y, x) = (y^*, x^*)$ satisfies

$$Cy - d = 0, \qquad \nabla f(y) + \partial g(y) + C^* x \ni 0, \tag{53}$$

or equivalently, to the inclusion problem (16) with x and y swapped, where

$$F(x, y) := \begin{pmatrix} Cy - d \\ C^* x \end{pmatrix}, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \qquad A(\cdot) = 0, \qquad B(\cdot) = \partial f(\cdot) + \partial g(\cdot).$$

We now state the algorithm that we are interested in studying in this subsection.

An inexact block-decomposition method for (51):

0) Let $(x_0, \tilde{y}_0) \in \mathcal{X} \times \mathcal{Y}$ and $0 < \sigma \leq 1$ be given, let $\lambda > 0$ be such that

$$\lambda L + \lambda^2 \|C\|^2 = \sigma^2. \tag{54}$$

and set $k = 1$;

1) compute

$$\tilde{x}_k = x_{k-1} - \lambda(C\tilde{y}_{k-1} - d), \qquad \tilde{y}_k = (I + \lambda \partial h)^{-1}[\tilde{y}_{k-1} - \lambda(\nabla f(\tilde{y}_{k-1}) - C^* \tilde{x}_k)], \tag{55}$$

2) set $x_k = \tilde{x}_k + \lambda C(\tilde{y}_{k-1} - \tilde{y}_k)$ and $k \leftarrow k + 1$, and go to step 1.

end

Define

$$\bar{\theta} := (\lambda L)^{1/2} < 1, \tag{56}$$

where the inequality is due to (54) and the assumption that $C \neq 0$ (see O.1).

Proposition 5.5. *The above inexact block-decomposition method for (51) is a special instance of the BD-HPE framework, where $\sigma_x = 0$, $\sigma_y = (\lambda L)^{1/2}$, and for every $k \in \mathbb{N}$,*

$$\lambda_k = \lambda, \qquad \varepsilon_k^x = 0, \qquad \varepsilon_k^y = \frac{L}{2}\|\tilde{y}_k - \tilde{y}_{k-1}\|^2, \tag{57}$$

and

$$\tilde{a}_k = 0, \qquad \tilde{b}_k = \frac{1}{\lambda}(\tilde{y}_{k-1} - \tilde{y}_k) + C^* \tilde{x}_k \in [\partial_{\varepsilon_k^y} f + \partial h](\tilde{y}_k), \qquad y_{k-1} = \tilde{y}_{k-1}, \tag{58}$$

Proof. Applying Proposition 4.1 with $G \equiv C\tilde{y}_{k-1} - d$, $C \equiv 0$ and $x = x_{k-1}$, and noting the definition of $\tilde{a}_k$, we conclude that $(\tilde{x}_k, \tilde{a}_k)$ satisfies (21) with $\sigma_x = 0$. Applying Proposition 4.2 with $T = \partial h$, $G \equiv -C^* \tilde{x}_k$, $x = \tilde{y}_{k-1}$ and $\sigma = \lambda L$, and noting the definition of $\tilde{b}_k$, we conclude that $(\tilde{y}_k, \tilde{b}_k)$ satisfies (21) with $\sigma_y = (\lambda L)^{1/2}$ and that the inclusion in (58) holds. Moreover, (58) and the update rule for x_k in step 2 of the algorithm imply that (22) also holds. Also, by (56), we have

$$\max\{\sigma_x, \sigma_y\} = (\lambda L)^{1/2} = \bar{\theta} < 1, \tag{59}$$

Moreover, using the fact that $\sigma_x^2 = 0$, $\sigma_y^2 = \lambda L$ and (54), we easily see that that λ satisfies (19). $\square$

We are now ready to state the convergence rate result for the inexact block-decomposition method for (51).

Theorem 5.6. *Consider the sequences* $\{x_k\}$ *and* $\{(\tilde{x}_k, \tilde{y}_k)\}$ *generated by the inexact block-decomposition method for (51), and the sequences* $\{\tilde{b}_k\}$ *and* $\{\varepsilon_k^y\}$ *defined as in (58) and (57). Moreover, for each* $k \in \mathbb{N}$, *define*

$$(\tilde{x}_k^a, \tilde{y}_k^a) = \frac{1}{k}\sum_{i=1}^{k}(\tilde{x}_i, \tilde{y}_i), \quad \tilde{b}_k^a = \frac{1}{k}\sum_{i=1}^{k}\tilde{b}_i,$$

and

$$\varepsilon_{k,B}^a := \frac{1}{k}\sum_{i=1}^{k}\left(\varepsilon_k^y + \left\langle \tilde{y}_i - \tilde{y}_k^a\,,\, \tilde{b}_i\right\rangle\right) \geq 0.$$

Let d_0 *denote the distance of the initial point* $(x_0, \tilde{y}_0) \in \mathcal{X} \times \mathcal{Y}$ *to the solution set of (53). Then, for every* $k \in \mathbb{N}$, *the following statements hold:*

a) $\tilde{b}_k \in (\partial_{\varepsilon_k^y} f + \partial h)(\tilde{y}_k)$, *and if* $\sigma < 1$, *there exists* $i \leq k$ *such that*

$$\left\|(C\tilde{y}_i - d, -C^*\tilde{x}_i + \tilde{b}_i)\right\| \leq \frac{d_0}{\lambda\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}}, \qquad \varepsilon_i^y \leq \frac{\sigma^2 d_0^2}{(1-\sigma^2)\lambda k};$$

b) we have

$$\tilde{b}_k^a \in \partial_{\varepsilon_{k,B}^a}(f+h)(\tilde{y}_k^a), \tag{60}$$

and

$$\left\|(C\tilde{y}_k^a - d, -C^*\tilde{x}_k^a + \tilde{b}_k^a)\right\| \leq \frac{2d_0}{k\lambda}, \qquad \varepsilon_{k,B}^a \leq \frac{2d_0^2}{k\lambda}(1+\bar{\eta}),$$

where

$$\bar{\eta} := \frac{2\sqrt{2}\sigma}{1-\bar{\theta}}\left(1 + \frac{1}{(1-\bar{\theta})^2}\right)^{1/2}$$

and $\bar{\theta}$ *is defined in (56).*

Proof. This result follows immediately from Proposition 5.5 and Theorems 3.2 and 3.3 by specializing the latter two results to the case where $\lambda_k = \lambda := \sigma\bar{\lambda}$, $\sigma_x = 0$, $\sigma_y = (\lambda L)^{1/2}$, and ε_k^x and ε_k^y given by (29), and using the fact that, by (59), $\max\{\sigma_x, \sigma_y\} = \bar{\theta}$. Observe also that (60) follows Theorem 2.3(b) and the fact that $\tilde{b}_k \in \partial_{\varepsilon_k^y}(f+h)(\tilde{y}_k)$, in view of statement a). ☐

Finally, note that it is possible to replace λ in the above estimates using its explicit formula:

$$\frac{1}{\lambda} = \frac{L + \sqrt{L^2 + 4\sigma^2\|C\|^2}}{2\sigma^2} \leq \frac{L}{\sigma^2} + \frac{\|C\|}{\sigma}.$$

6 Sum of two maximal monotone operators

In this section, we consider the monotone inclusion problem consisting of the sum of two maximal monotone operators and show how it can be transformed to a problem of form (16), which can then be solved by any instance of the BD-HPE framework.

Consider the problem

$$0 \in (A+B)(x) \tag{61}$$

where $A, B : \mathcal{X} \rightrightarrows \mathcal{X}$ are maximal monotone. We assume that the resolvents of A and B are easily computable. Note that (61) is equivalent to the existence of $b \in \mathcal{X}$ such that

$$-b \in A(x), \quad b \in B(x),$$

or equivalently,

$$0 \in b + A(x), \quad 0 \in -x + B^{-1}(b).$$

Hence, the inclusion problem (61) is equivalent to the monotone inclusion problem

$$0 \in [F + (A \otimes B^{-1})](x, b), \tag{62}$$

where

$$F(x, b) = (b, -x). \tag{63}$$

Hence, we can apply any instance of the BD-HPE framework, and in particular the exact BD-HPE method of Subsection 5.1, to the inclusion problem (62) in order to compute an approximate solution of (61). In the following two subsections, we will discuss in detail these approaches, stating a general BD-HPE framework for (62) in Subsection 6.1, and an exact BD-HPE method for (62) in Subsection 6.2.

6.1 Block-decomposition HPE framework for (61)

We start by stating a general BD-HPE framework for solving (61).

Inexact BD-HPE framework for (61):

0) Let $(x_0, b_0) \in \mathcal{X} \times \mathcal{X}$, $\sigma_x, \sigma_y \in [0, 1)$ and $\sigma \in [0, 1]$ be given and set $k = 1$;

1) choose $\lambda_k > 0$ such that (19) holds with $L_{xy} = 1$;

2) compute $\tilde{x}_k, \tilde{a}_k \in \mathcal{X}$ and $\varepsilon_k^x \geq 0$ such that

$$\tilde{a}_k \in A^{\varepsilon_k^x}(\tilde{x}_k), \quad \|\lambda_k[b_{k-1} + \tilde{a}_k] + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k^x \leq \sigma_x^2 \|\tilde{x}_k - x_{k-1}\|^2; \tag{64}$$

3) compute $\tilde{y}_k, \tilde{b}_k \in \mathcal{X}$ and $\varepsilon_k^y \geq 0$ such that

$$\tilde{y}_k \in (B^{-1})^{\varepsilon_k^y}(\tilde{b}_k), \quad \|\lambda_k[-\tilde{x}_k + \tilde{y}_k] + \tilde{b}_k - b_{k-1}\|^2 + 2\lambda_k \varepsilon_k^y \leq \sigma_y^2 \|\tilde{b}_k - b_{k-1}\|^2; \tag{65}$$

3) set

$$(x_k, b_k) = (x_{k-1}, b_{k-1}) - \lambda_k[(\tilde{b}_k, -\tilde{x}_k) + (\tilde{a}_k, \tilde{y}_k)] = (x_{k-1}, b_{k-1}) - \lambda_k(\tilde{b}_k + \tilde{a}_k, \tilde{y}_k - \tilde{x}_k), \tag{66}$$

$k \leftarrow k + 1$ and go to step 1.

end

Clearly, the above framework is nothing else but the BD-HPE framework of Section 3 for the monotone inclusion problem (62). Note that it is stated in terms of B^{-1}. It is possible to state a version of it which replaces condition (65) based on B^{-1} by a sufficient condition based on B as described by the following result.

Proposition 6.1. *If $\tilde{x}_k, \tilde{y}_k, b_{k-1}, \tilde{b}_k \in \mathcal{X}$, $\lambda_k > 0$ and $\varepsilon_k^y \geq 0$ satisfy*

$$\tilde{b}_k \in (B)^{\varepsilon_k^y}(\tilde{y}_k), \quad \|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}] + \tilde{y}_k - \tilde{x}_k\|^2 + 2\lambda_k^{-1}\varepsilon_k^y \leq \left(\frac{\sigma_y}{1 + \sigma_y}\right)^2 \|\tilde{y}_k - \tilde{x}_k\|^2, \tag{67}$$

then (65) holds.

Proof. Since $(B^{-1})^\varepsilon = (B^\varepsilon)^{-1}$, the inclusion on (67) implies that $\tilde{y}_k \in (B^{-1})^{\varepsilon_k^y}$. Using the inequality in (67) and the triangle inequality for norms, we have

$$\begin{aligned} \|\tilde{y}_k - \tilde{x}_k\| &\leq \|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}]\| + \|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}] + \tilde{y}_k - \tilde{x}_k\| \\ &\leq \|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}]\| + \frac{\sigma_y}{1 + \sigma_y}\|\tilde{y}_k - \tilde{x}_k\| \end{aligned}$$

Hence, $\|\tilde{y}_k - \tilde{x}_k\| \le (1+\sigma_y)\|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}]\|$, which combined with the inequality in (67) yields

$$\|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}] + \tilde{y}_k - \tilde{x}_k\|^2 + 2\lambda_k^{-1}\varepsilon_k^y \le \sigma_y^2\|\lambda_k^{-1}[\tilde{b}_k - b_{k-1}]\|^2.$$

To end the proof, multiply the above inequality by λ_k^2. □

Note that the pair $(\tilde{y}, \tilde{b}) = (\tilde{y}_k, \tilde{b}_k)$ in the above result is an approximate solution of the proximal point equation $\lambda_k^{-1}[\tilde{b} - b_{k-1}] + (\tilde{y} - \tilde{x}_k) = 0$ and $\tilde{y} \in B(\tilde{b})$ in the sense described in the paragraph after the statement of the HPE method in Section 2.2.

The specialization of Theorems 3.2 and 3.3 for the above method are as follows.

Theorem 6.2. *Consider the sequences $\{\lambda_k\}$, $\{(\varepsilon_k^x, \varepsilon_k^y)\}$, $\{(\tilde{x}_k, \tilde{b}_k)\}$ and $\{(\tilde{a}_k, \tilde{y}_k)\}$ generated by the BD-HPE method for (61). Moreover, for every $k \in \mathbb{N}$, define:*

$$(\tilde{x}_k^a, \tilde{y}_k^a, \tilde{a}_k^a, \tilde{b}_k^a) = \tfrac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i(\tilde{x}_i, \tilde{y}_i, \tilde{a}_i, \tilde{b}_i),$$

$$\varepsilon_{k,A}^a = \tfrac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i\left(\varepsilon_i^x + \langle \tilde{x}_i - \tilde{x}_k^a, \tilde{a}_i\rangle\right), \qquad \varepsilon_{k,B}^a = \tfrac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i\left(\varepsilon_i^y + \langle \tilde{b}_i - \tilde{b}_k^a, \tilde{y}_i\rangle\right).$$

Let d_0 denote the distance of the initial point $(x_0, b_0) \in \mathcal{X}\times\mathcal{X}$ to the solution set of (62), i.e.:

$$d_0 := \min\left\{\left(\|x - x_0\|^2 + \|b - b_0\|^2\right)^{1/2} : -b \in A(x),\, b \in B(x)\right\}.$$

Then, for every $k \in \mathbb{N}$, the following statements hold:

a) $(\tilde{a}_k, \tilde{b}_k) \in A^{\varepsilon_k^x}(\tilde{x}_k) \times B^{\varepsilon_k^y}(\tilde{y}_k)$, and, if $\sigma < 1$, then there exists $i \le k$ such that

$$\left\|(\tilde{b}_i + \tilde{a}_i, -\tilde{x}_i + \tilde{y}_i)\right\| \le d_0\sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{1}{\sum_{j=1}^k \lambda_j^2}\right)}, \qquad \varepsilon_i^x + \varepsilon_i^y \le \frac{\sigma^2 d_0^2 \lambda_i}{2(1-\sigma^2)\sum_{j=1}^k \lambda_j^2};$$

b) $(\tilde{a}_k^a, \tilde{b}_k^a) \in A^{\varepsilon_{k,A}^a}(\tilde{x}_k^a) \times B^{\varepsilon_{k,B}^a}(\tilde{y}_k^a)$ and

$$\left\|(\tilde{b}_k^a + \tilde{a}_k^a, -\tilde{x}_k^a + \tilde{y}_k^a)\right\| \le \frac{2d_0}{\Lambda_k}, \quad \varepsilon_{k,A}^a + \varepsilon_{k,B}^a \le \frac{2d_0^2}{\Lambda_k}(1+\eta_k),$$

where η_k is defined in (31) with $L_{xy} = 1$.

Proof. This result follows as an immediate consequence of Theorems 3.2 and 3.3 applied to (62)-(63) and noting that in this case F is affine and $L_{xy} = 1$. □

6.2 Exact BD-HPE method for (61)

In this subsection, we state an exact BD-HPE method for (61) and corresponding convergence rate results.

Exact BD-HPE method for (61):

0) Let $x_0, \tilde{b}_0 \in \mathcal{X}$ and $\lambda \in (0, 1]$ be given, and set $k = 1$;

1) compute $\tilde{x}_k, \tilde{b}_k \in \mathcal{X}$ as

$$\tilde{x}_k = (I + \lambda A)^{-1}(x_{k-1} - \lambda\tilde{b}_{k-1}), \quad \tilde{b}_k = (I + \lambda B^{-1})^{-1}(\tilde{b}_{k-1} + \lambda\tilde{x}_k);$$

2) set $x_k = \tilde{x}_k - \lambda(\tilde{b}_k - \tilde{b}_{k-1})$ and go to step 1.

end

The method above is nothing else but the exact BD-HPE method of Subsection 5.1 applied to (62)-(63) with variable b replacing variable y and vice-versa.

The following result shows that the resolvent of B^{-1} used in step 1) can be computed using the resolvent of B.

Lemma 6.3. *Let $b, u \in \mathcal{X}$ be given. Then,*

$$b = (I + \lambda B^{-1})^{-1}(u) \Leftrightarrow b = u - \lambda(I + \lambda^{-1}B)^{-1}(\lambda^{-1}u).$$

Proof. We have

$$\begin{aligned} & b = (I + \lambda B^{-1})^{-1}(u) \;\Leftrightarrow\; u \in b + \lambda B^{-1}(b) \;\Leftrightarrow\; b \in B(\lambda^{-1}(u-b)) \\ \Leftrightarrow\; & \lambda^{-1}u \in (\lambda^{-1}B + I)(\lambda^{-1}(u-b)) \;\Leftrightarrow\; \lambda^{-1}(u-b) = (I + \lambda^{-1}B)^{-1}(\lambda^{-1}u), \end{aligned}$$

from which the conclusion of the lemma immediately follows. □

In view of the above result, $\tilde{b}_k$ may be computed as

$$\tilde{b}_k = \tilde{b}_{k-1} + \lambda\tilde{x}_k - \lambda(I + \lambda^{-1}B)^{-1}(\lambda^{-1}\tilde{b}_{k-1} + \tilde{x}_k).$$

The following iteration-complexity bounds for solving the inclusion problem (61) can now be stated as an immediate consequence of Theorem 5.2.

Theorem 6.4. *Consider the sequences $\{x_k\}$ and $\{(\tilde{x}_k, \tilde{b}_k)\}$ generated by the BD-HPE method for (61) and define the sequence $\{(\tilde{a}_k, \tilde{y}_k)\}$ as*

$$\tilde{a}_k = \frac{1}{\lambda}(x_{k-1} - \tilde{x}_k) - \tilde{b}_{k-1}, \quad \tilde{y}_k = \frac{1}{\lambda}(\tilde{b}_{k-1} - \tilde{b}_k) + \tilde{x}_k. \tag{68}$$

Moreover, for every $k \in \mathbb{N}$, define:

$$\begin{aligned} (\tilde{x}^a_k, \tilde{y}^a_k, \tilde{a}^a_k, \tilde{b}^a_k) &= \tfrac{1}{k}\textstyle\sum_{i=1}^k (\tilde{x}_i, \tilde{y}_i, \tilde{a}_i, \tilde{b}_i), \\ \varepsilon^a_{k,A} = \tfrac{1}{k}\textstyle\sum_{i=1}^k \langle \tilde{x}_i - \tilde{x}^a_k, \tilde{a}_i\rangle, &\qquad \varepsilon^a_{k,B} = \tfrac{1}{k}\textstyle\sum_{i=1}^k \langle \tilde{b}_i - \tilde{b}^a_k, \tilde{y}_i\rangle. \end{aligned}$$

Let d_0 denote the distance of the initial point $(x_0, \tilde{b}_0) \in \mathcal{X} \times \mathcal{X}$ to the solution set of (62), i.e.:

$$d_0 := \min\left\{ \left(\|x - x_0\|^2 + \|b - \tilde{b}_0\|^2\right)^{1/2} : -b \in A(x),\, b \in B(x) \right\}.$$

Then, for every $k \in \mathbb{N}$, the following statements hold:

a) $(\tilde{a}_k, \tilde{b}_k) \in A(\tilde{x}_k) \times B(\tilde{y}_k)$, and if $\lambda < 1$, there exists $i \le k$ such that

$$\left\| (\tilde{b}_i + \tilde{a}_i, -\tilde{x}_i + \tilde{y}_i) \right\| \le \frac{d_0}{\lambda\sqrt{k}}\sqrt{\frac{1+\lambda}{1-\lambda}};$$

b) $(\tilde{a}^a_k, \tilde{b}^a_k) \in A^{\varepsilon^a_{k,A}}(\tilde{x}^a_k) \times B^{\varepsilon^a_{k,B}}(\tilde{y}^a_k)$ and

$$\left\| (\tilde{b}^a_k + \tilde{a}^a_k, -\tilde{x}^a_k + \tilde{y}^a_k) \right\| \le \frac{2d_0}{k\lambda}, \quad \varepsilon^a_{k,A} + \varepsilon^a_{k,B} \le \frac{2d_0^2}{k\lambda}\left(1 + 2\sqrt{2}\lambda\right).$$

Proof. This result follows as an immediate consequence of Theorem 5.2 applied to (62)-(63) and noting that in this case F is affine and $L_{xy} = 1$, and hence $\lambda = \sigma$. □

We end this section by discussing the special case of the exact BD-HPE method for (61) in which $\lambda = 1$. It can be easily shown that this algorithm is equivalent to the Douglas-Rachford splitting method (see for example [9]). Hence, Theorem 6.4(b) for $\lambda = 1$ gives an ergodic complexity estimation of the Douglas-Rachford method.

7 Alternating minimization augmented Lagrangian methods

In this section, we consider the AMAL method for solving a large class of linearly constrained convex programming problems with proper closed convex objective functions and show that it can be interpreted as a specific instance of the BD-HPE framework applied to a two-block monotone inclusion problem.

We assume in this section that $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are inner product spaces whose inner products and associated norms are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. Consider the problem

$$\min\{f(y) + g(s) : Cy + Ds = c\} \tag{69}$$

where $c \in \mathcal{X}$, $C : \mathcal{Y} \to \mathcal{X}$ and $D : \mathcal{S} \to \mathcal{X}$ are linear operators, and $f : \mathcal{Y} \to \bar{\mathbb{R}}$ and $g : \mathcal{S} \to \bar{\mathbb{R}}$ are proper closed convex functions. Throughout this section, we also assume that the resolvent of ∂f and ∂g can be computed exactly.

The Lagrangian function $L : (\mathcal{Y} \times \mathcal{S}) \times \mathcal{X} \to \bar{\mathbb{R}}$ for problem (69) is defined as

$$\mathcal{L}(y, s; x) = f(y) + g(s) - \langle x, Cy + Ds - c\rangle. \tag{70}$$

We make the following assumptions throughout this section:

C.1) there exists a saddle point of $\mathcal{L}$, i.e., a point $(y^*, s^*; x^*)$ such that $\mathcal{L}(y^*, s^*; x^*)$ is finite and

$$\min_{(y,s)\in\mathcal{Y}\times\mathcal{S}} \mathcal{L}(y, s; x^*) = \mathcal{L}(y^*, s^*; x^*) = \max_{x\in\mathcal{X}} \mathcal{L}(y^*, s^*; x); \tag{71}$$

C.2) $\mathrm{ri}(\mathrm{dom}\, g^*) \cap \mathrm{range}\, \mathrm{D}^* \neq \emptyset$;

C.3) C is injective.

For a scalar $\rho \geq 0$, the ρ-augmented Lagrangian function $\mathcal{L}_\rho : (\mathcal{Y} \times \mathcal{S}) \times \mathcal{X} \to \bar{\mathbb{R}}$ associated with (69) is defined as

$$\mathcal{L}_\rho(y, s; x) := f(y) + g(s) + \langle x, c - Cy - Ds\rangle + \frac{\rho}{2}\|Cy + Ds - c\|^2.$$

We next state the alternating minimization augmented Lagrangian (AMAL) method applied for problem (69).

AMAL method:

0) Let $\rho > 0$ and $(x_0, \tilde{y}_0) \in \mathcal{X} \times \mathcal{Y}$ be given and set $k = 1$;

1) compute $\tilde{s}_k \in \mathcal{S}$ as

$$\tilde{s}_k = \mathrm{argmin}_s\{\mathcal{L}_\rho(\tilde{y}_{k-1}, s; x_{k-1})\} = \mathrm{argmin}_s \left\{g(s) - \langle D^* x_{k-1}, s\rangle + \frac{\rho}{2}\|C\tilde{y}_{k-1} + Ds - c\|^2\right\}. \tag{72}$$

and $\tilde{y}_k \in \mathcal{Y}$ as

$$\tilde{y}_k = \mathrm{argmin}_y\{\mathcal{L}_\rho(y, \tilde{s}_k; x_{k-1})\} = \mathrm{argmin}_y \left\{f(y) - \langle C^* x_{k-1}, y\rangle + \frac{\rho}{2}\|Cy + D\tilde{s}_k - c\|^2\right\}. \tag{73}$$

2) set $x_k = x_{k-1} - \rho(C\tilde{y}_k + D\tilde{s}_k - c)$ and $k \leftarrow k + 1$, and go to step 1;

end

Our goal in the remaining part of this section is to show that the AMAL method is a special case of the exact BD-HPE method of Subsection 5.1 for a specific monotone inclusion problem of the form (16). As a by-product, we also derive convergence rate results for the AMAL problem.

We start by giving a preliminary technical result about the AMAL method.

Proposition 7.1. *Let $(x_{k-1}, \tilde{y}_{k-1}) \in \mathcal{X} \times \mathcal{Y}$ be given. Then, the following statements hold:*

a) $\tilde{s}_k \in \mathcal{S}$ is an optimal solution of (72) if, and only if, the point

$$\tilde{x}_k := x_{k-1} - \rho(C\tilde{y}_{k-1} + D\tilde{s}_k - c), \tag{74}$$

satisfies

$$\tilde{s}_k \in \partial g^*(D^*\tilde{x}_k); \tag{75}$$

b) if $(\tilde{s}_k, \tilde{y}_k, x_k)$ are computed according to the k-iteration of the AMAL method, then

$$0 \in \partial(g^* \circ D^*)(\tilde{x}_k) + C\tilde{y}_{k-1} - c + \frac{1}{\rho}(\tilde{x}_k - x_{k-1}), \tag{76}$$

$$0 \in \partial f(\tilde{y}_k) - C^*\tilde{x}_k + \rho C^*C(\tilde{y}_k - \tilde{y}_{k-1}), \tag{77}$$

$$x_k = \tilde{x}_k - \rho C(\tilde{y}_k - \tilde{y}_{k-1}). \tag{78}$$

Proof. By (74) and the optimality conditions of (72), we have that $\tilde{s}_k$ is an optimal solution of (72) if, and only if,

$$D^*\tilde{x}_k = D^*[x_{k-1} - \rho(C\tilde{y}_{k-1} + D\tilde{s}_k - c)] \in \partial g(\tilde{s}_k),$$

which in turn is equivalent to (75). On the other hand, (75) implies that $D\tilde{s}_k \in D\partial g^*(D^*\tilde{x}_k) \subset \partial(g^* \circ D^*)(\tilde{x}_k)$. Combining the latter inclusion with (74), we obtain (76). Moreover, (73) implies that

$$0 \in \partial f(\tilde{y}_k) - C^*x_{k-1} + \rho C^*(C\tilde{y}_k + D\tilde{s}_k - c).$$

Combining the above equation with (74), we obtain (77). Finally, (78) follows immediately (74) and the update rule for x_k in step 3 of the AMAL method. □

Proposition 7.2. *Given $(x_{k-1}, \tilde{y}_{k-1}) \in \mathcal{X} \times \mathcal{Y}$, define*

$$\hat{x}_k := (\rho\partial(g^* \circ D^*) + I)^{-1}[x_{k-1} + \rho(c - C\tilde{y}_{k-1})], \qquad \hat{w}_k := \frac{1}{\rho}(x_{k-1} - \hat{x}_k) + c - C\tilde{y}_{k-1}.$$

Then, the following statements hold:

a) $\tilde{s}_k \in \mathcal{S}$ is an optimal solution of (72) if, and only if, $\tilde{s}_k \in D^{-1}(\hat{w}_k) \cap \partial g^*(D^*\hat{x}_k)$;

b) if condition C.2) holds, then $D^{-1}(\hat{w}_k) \cap \partial g^*(D^*\hat{x}_k) \neq \emptyset$, and hence the set of optimal solutions of (72) is nonempty;

c) if condition C.3) holds, then the set of optimal solutions of (73) is nonempty.

Proof. First, observe that $(\tilde{x}, \tilde{w}) = (\hat{x}_k, \hat{w}_k)$ is the unique solution of

$$0 = \rho\,(C\tilde{y}_{k-1} - c + \tilde{w}) + \tilde{x} - x_{k-1}, \qquad \tilde{w} \in (g^* \circ D^*)(\tilde{x}). \tag{79}$$

a) Assume first that $\tilde{s}_k$ is an optimal solution of (72). By Proposition 7.1, we conclude that $\tilde{x}_k$ given by (74) satisfies (75). Then,

$$D\tilde{s}_k \in D\partial g^*(D^*\tilde{x}_k) \subset \partial(g^* \circ D^*)(\tilde{x}_k),$$

which, together with (74), imply that $(\tilde{x}, \tilde{w}) = (\tilde{x}_k, D\tilde{s}_k)$ satisfies (79). Hence, in view of the observation made at the beginning of this proof, we conclude that $\tilde{x}_k = \hat{x}_k$ and $D\tilde{s}_k = \hat{w}_k$. These identities and inclusion (75) then imply that $\tilde{s}_k \in D^{-1}(\hat{w}_k) \cap \partial g^*(D^*\hat{x}_k)$.

Conversely, assume that $\tilde{s}_k \in D^{-1}(\hat{w}_k) \cap \partial g^*(D^*\hat{x}_k)$. Then, $\hat{w}_k = D\tilde{s}_k$, and hence $(\hat{x}_k, D\tilde{s}_k) = (\hat{x}_k, \hat{w}_k)$ satisfies (79). In particular,

$$\hat{x}_k = x_{k-1} - \rho\,(C\tilde{y}_{k-1} - c + D\tilde{s}_k) = \tilde{x}_k,$$

where the latter equality is due to (74). Since $\tilde{x}_k = \hat{x}_k$ and, by assumption, $\tilde{s}_k \in \partial g^*(D^*\hat{x}_k)$, we conclude that (75) holds, and hence that $\tilde{s}_k$ is an optimal solution of (72).

b) Using the fact that $(\hat{x}_k, \hat{w}_k)$ satisfies (79), we conclude that

$$\hat{w}_k \in \partial(g^* \circ D^*)(\hat{x}_k) = D(\partial g^*(D^*\hat{x}_k)).$$

where the latter equality is due to Assumption C.2). The latter inclusion clearly implies that $D^{-1}(\hat{w}_k) \cap \partial g^*(D^*\hat{x}_k) \neq \emptyset$.

c) This statement follows from the fact that, under Assumption C.3, the objective function of (73) is strongly convex. $\square$

We will now derive the aforementioned monotone inclusion problem of the form (16). Let $X \times Y := \operatorname{dom} f \times (D^*)^{-1}(\operatorname{dom} g^*)$ and $\Psi : X \times Y \to \mathbb{R}$ be defined as

$$\begin{aligned}\Psi(y,x) &= \min_s \mathcal{L}(y,s;x) = f(y) + \langle x, c - Cy\rangle + \Big(\min_s g(s) - \langle D^*x, s\rangle\Big)\\ &= f(y) + \langle x, c - Cy\rangle - g^*(D^*x).\end{aligned}$$

It is easy to see that the pair (y^*, x^*) as in Assumption C.1) satisfies

$$\max_{x\in\mathcal{X}} \min_{(y,s)\in\mathcal{Y}\times\mathcal{S}} \mathcal{L}(y,s;x) = \max_{x\in X}\min_{y\in Y} \Psi(y,x) = \Psi(y^*,x^*) = \min_{y\in Y}\max_{x\in X}\Psi(y,x) = \min_{(y,s)\in\mathcal{Y}\times\mathcal{S}}\max_{x\in\mathcal{X}} \mathcal{L}(y,s;x) \in \Re.$$

The latter condition is in turn equivalent to (y^*, x^*) being a solution of the inclusion problem:

$$0 \in \partial f(y) - C^*x, \qquad 0 \in \partial(g^* \circ D^*)(x) + Cy - c. \tag{80}$$

Under the assumption that C^*C is nonsingular, the latter inclusion problem is clearly equivalent to

$$0 \in (\rho C^*C)^{-1}(\partial f(y) - C^*x), \qquad 0 \in \rho\left[\partial(g^* \circ D^*)(x) + Cy - c\right]. \tag{81}$$

and hence to inclusion problem (16) with

$$F(x,y) = \begin{bmatrix} \rho(Cy - c) \\ -(\rho C^*C)^{-1}C^*x \end{bmatrix}, \qquad A(x) = \rho\partial(g^* \circ D^*)(x), \qquad B(y) = (\rho C^*C)^{-1}\partial f(y). \tag{82}$$

If $\mathcal{U}$ is an inner product space with inner product also denoted by $\langle\cdot,\cdot\rangle$, then a symmetric positive definite operator $M : \mathcal{U} \to \mathcal{U}$ defines another inner product, denoted by $\langle\cdot,\cdot\rangle_M$, as follows:

$$\langle u, u'\rangle_M = \langle u, Mu'\rangle, \quad \forall u, u' \in \mathcal{U}.$$

We will denote the norm associated with the above inner product by $\|\cdot\|_M$. Moreover, when $M = \tau I$ for some $\tau > 0$, where I denotes the identity operator, we denote the norm $\|\cdot\|_M$ simply by $\|\cdot\|_\tau$.

Proposition 7.3. *Assume that C^*C is nonsingular,* $\operatorname{ri}(\operatorname{dom} g^*) \cap \operatorname{Range}(D^*)$ *is non-empty and consider the inner products $\langle\cdot,\cdot\rangle_{\rho^{-1}}$ in $\mathcal{X}$ and $\langle\cdot,\cdot\rangle_{\tilde{C}_\rho}$ in $\mathcal{Y}$, where $\tilde{C}_\rho := \rho C^*C$. Then, the following statements hold:*

- *a) the map F defined in (82) is monotone with respect to the inner product $\langle\cdot,\cdot\rangle_{\rho^{-1}} + \langle\cdot,\cdot\rangle_{\tilde{C}_\rho}$ in $\mathcal{X}\times\mathcal{Y}$ and the operators A and B defined in (82) are maximal monotone with respect to $\langle\cdot,\cdot\rangle_{\rho^{-1}}$ and $\langle\cdot,\cdot\rangle_{\tilde{C}_\rho}$, respectively;*
- *b) F satisfies (15) with $L_{xy} = 1$;*
- *c) the sequence $\{(x_k, \tilde{y}_k)\}$ generated by the AMAL method together with the sequence $\{\tilde{x}_k\}$ defined in (74) correspond to the same sequence $\{(x_k, \tilde{x}_k, \tilde{y}_k)\}$ generated by the exact BD-HPE method applied to the inclusion problem (16) with (F, A, B) given by (82) and with $\sigma = 1$ (or equivalently, $\lambda = 1$).*

Proof. Monotonicity of F in the rescaled space $\mathcal{X} \times \mathcal{Y}$ holds trivially. Maximal monotonicity of A (resp., B) in $\mathcal{X}$ (resp., $\mathcal{Y}$) endowed with the norm $\|\cdot\|_{\rho^{-1}}$ (resp., $\|\cdot\|_{\tilde{C}_\rho}$) follows from the fact that this operator is the subdifferential of $g^* \circ D^*$ (resp., f) in the rescaled space. For b), observe that

$$\|F_x(x,y) - F_x(x,y')\|^2_{\rho^{-1}} = \|\rho C(y-y')\|^2_{\rho^{-1}} = \rho\|C(y-y')\|^2 = \langle y - y', (\rho C^* C)(y-y')\rangle = \|y-y'\|^2_{\tilde{C}_\rho}.$$

Statement c) follows immediately from Proposition 7.1 by noting that relations (76)-(78) reduce to the recursive formulas for obtaining $(x_k, \tilde{x}_k, \tilde{y}_k)$ in the the exact BD-HPE method applied to the inclusion problem (16) with (F, A, B) given by (82) and with $\lambda = 1$. □

As a consequence of the previous proposition, we obtain the following convergence rate result.

Theorem 7.4. *Consider the sequence $\{(x_k, \tilde{y}_k)\}$ generated by the AMAL method and the sequence $\{\tilde{x}_k\}$ defined according to (74). Consider also the sequence $\{(\tilde{a}_k, \tilde{b}_k)\}$ defined as*

$$\tilde{a}_k = x_{k-1} - \tilde{x}_k - \rho(C\tilde{y}_{k-1} - c), \qquad \tilde{b}_k = \tilde{y}_{k-1} - \tilde{y}_k + \tilde{C}_\rho^{-1} C^* \tilde{x}_k, \tag{83}$$

where $\tilde{C}_\rho := \rho C^* C$. Moreover, define the sequences $\{(\tilde{x}^a_k, \tilde{y}^a_k)\}$, $\{(\tilde{a}^a_k, \tilde{b}^a_k)\}$ and $\{(\varepsilon^a_{k,x}, \varepsilon^a_{k,y})\}$ as

$$(\tilde{x}^a_k, \tilde{y}^a_k) = \frac{1}{k}\sum_{i=1}^k (\tilde{x}_i, \tilde{y}_i), \qquad (\tilde{a}^a_k, \tilde{b}^a_k) = \frac{1}{k}\sum_{i=1}^k (\tilde{a}_i, \tilde{b}_i), \tag{84}$$

and

$$\varepsilon^a_{k,x} = \frac{1}{k}\sum_{i=1}^k \left\langle \tilde{x}_i - \tilde{x}^a_k, \frac{1}{\rho}\tilde{a}_i \right\rangle, \qquad \varepsilon^a_{k,y} = \frac{1}{k}\sum_{i=1}^k \left\langle \tilde{y}_i - \tilde{y}^a_k, \tilde{C}_\rho \tilde{b}_i \right\rangle. \tag{85}$$

Then, for every $k \in \mathbb{N}$,

$$\frac{1}{\rho}\tilde{a}^a_k \in \partial_{\varepsilon^a_{k,x}}(g^* \circ D^*)(\tilde{x}^a_k), \qquad \tilde{C}_\rho \tilde{b}^a_k \in \partial_{\varepsilon^a_{k,y}} f(\tilde{y}^a_k), \tag{86}$$

and

$$\left(\rho\left\|C\tilde{y}^a_k - c + \frac{1}{\rho}\tilde{a}^a_k\right\|^2 + \frac{1}{\rho}\left\|-C^*\tilde{x}^a_k + \tilde{C}_\rho \tilde{b}^a_k\right\|^2_{(C^*C)^{-1}}\right)^{1/2} \le \frac{2d_0}{k}, \qquad \varepsilon^a_{k,x} + \varepsilon^a_{k,y} \le \frac{2(1+2\sqrt{2})d_0^2}{k}, \tag{87}$$

where d_0 is the distance of the initial point $(x_0, \tilde{y}_0) \in \mathcal{X} \times \mathcal{Y}$ to the solution set of (80) with respect to the inner product $\langle \cdot,\cdot\rangle_{\rho^{-1}} + \langle\cdot,\cdot\rangle_{\tilde{C}_\rho}$ in $\mathcal{X} \times \mathcal{Y}$, namely d_0 is the infimum of

$$\sqrt{\frac{1}{\rho}\|x_0 - x^*\|^2 + \rho\|\tilde{y}_0 - y^*\|^2_{C^*C}}$$

over the set of all solutions (x^*, y^*) of (80).

Proof. By Proposition 7.3(c), we know that $\{(x_k, \tilde{x}_k, \tilde{y}_k)\}$ is the sequence generated by applying the exact BD-HPE method to (81) with $\sigma = 1$ and $L_{xy} = 1$, and hence $\lambda = 1$. Hence, it follows from Lemma 5.1 and Theorem 5.2, relations (82), (83), and the definition of F and the fact that F is affine, that the last inequality in (87) holds,

$$\tilde{a}_k \in \rho\partial(g^* \circ B^*)(\tilde{x}_k), \qquad \tilde{b}_k \in \tilde{C}_\rho^{-1}\partial f(\tilde{y}_k), \tag{88}$$

and

$$\left(\|\rho(C\tilde{y}^a_k - c) + \tilde{a}_k\|^2_{\rho^{-1}} + \|\tilde{b}^a_k - \tilde{C}_\rho^{-1}C^*\tilde{x}^a_k\|^2_{\tilde{C}_\rho}\right)^{1/2} \le \frac{2d_0}{k}.$$

Now, using the definition of $\tilde{C}_\rho$ and the norm induced by this operator, we easily see that the latter inequality is equivalent to the first inequality in (87). Moreover, (86) follows from (85) and (88) and Theorem 2.3(b). □

We now translate the above result stated more in the context of the inclusion problem (80) and the exact BD-HPE method to the context of the original optimization problem (69) and algorithm AMAL.

Theorem 7.5. *Consider the sequence* $\{(x_k, \tilde{y}_k, \tilde{s}_k)\}$ *generated by the AMAL method and the sequences* $\{\tilde{x}_k\}$ *and* $\{\tilde{b}_k\}$ *defined according to (74) and (83). Moreover, consider the sequences* $\{(\tilde{x}^a_k, \tilde{y}^a_k)\}$, $\{\tilde{b}^a_k\}$, $\{(\varepsilon^a_{k,x}, \varepsilon^a_{k,y})\}$ *defined in (84) and (85), and define for every* $k \in \mathbb{N}$:

$$\tilde{s}^a_k := \frac{1}{k}\sum_{i=1}^{k} \tilde{s}_i, \qquad r^x_k := C\tilde{y}^a_k + D\tilde{s}^a_k - c, \qquad r^y_k := \rho C^* C\tilde{b}^a_k - C^*\tilde{x}^a_k. \tag{89}$$

Then,

$$0 \in \partial g_{\varepsilon^a_{k,x}}(\tilde{s}^a_k) - D^*\tilde{x}^a_k, \qquad r^y_k \in \partial_{\varepsilon^a_{k,y}} f(\tilde{y}^a_k) - C^*\tilde{x}^a_k, \tag{90}$$

and

$$\left(\rho\|r^x_k\|^2 + \frac{1}{\rho}\|r^y_k\|^2_{(C^*C)^{-1}}\right)^{1/2} \leq \frac{2d_0}{k}, \qquad \varepsilon^a_{k,x} + \varepsilon^a_{k,y} \leq \frac{2(1+2\sqrt{2})d_0^2}{k}, \tag{91}$$

where d_0 *is defined as in Theorem 7.4.*

Proof. First note that (74) and the definition of $\tilde{a}_k$ in (83) imply that

$$\tilde{a}_k = \rho D\tilde{s}_k, \tag{92}$$

which together with (85) imply

$$\varepsilon^a_{k,x} = \frac{1}{k}\sum_{i=1}^{k}\left\langle \tilde{x}_i - \tilde{x}^a_k, \frac{1}{\rho}\tilde{a}_i\right\rangle = \frac{1}{k}\sum_{i=1}^{k}\langle \tilde{x}_i - \tilde{x}^a_k, D\tilde{s}_i\rangle = \frac{1}{k}\sum_{i=1}^{k}\langle D^*\tilde{x}_i - D^*\tilde{x}^a_k, \tilde{s}_i\rangle.$$

This identity, (75) and Theorem 2.3(b) then imply that

$$\tilde{s}^a_k \in \partial_{\varepsilon^a_{k,x}} g^*(D^*\tilde{x}^a_k),$$

from which the first inclusion in (90) follows. The second inclusion in (90) follows from the definition of $\tilde{C}_\rho$ in Theorem 7.4, the second inclusion in (86) and the definition of r^y_k in (89). In addition, the estimates in (91) follow from (87), the definition of $\tilde{C}_\rho$, (89), (92), and the definition of $\tilde{a}^a_k$ in (84). □

8 Concluding remarks

In this paper, we have presented a general framework, namely the BD-HPE framework, of block-decomposition algorithms and obtained broad convergence rate results for it. As a consequence, we have derived for the first time convergence rate results for the classical AMAL method by showing that it can be viewed as a special instance of the BD-HPE framework. We have also proposed new block-decomposition algorithms and derived their respective convergence rate results. These include a new splitting method for finding a zero of the sum of two maximal monotone operators and a new block-decomposition method based on Tseng's modified forward-backward splitting procedure. The analysis of the latter uses of an important feature of the BD-HPE framework, i.e., that it allows the one-block subproblems to be solved only approximately.

Finally, we make some remarks about a recent work of Chambolle and Pock [6] in light of the development in this paper. They have studied the monotone inclusion problem

$$0 \in K^*y + \partial g(x), \quad 0 \in -Kx + \partial f^*(y),$$

where K is a linear map and f, g are proper closed convex functions, and analyzed the convergence rate of an algorithm based on the exact evaluation of the resolvents of ∂g and ∂f^* (or ∂f). Their analysis, in contrast to ours, is heavily based on the fact that the above monotone inclusion problem is the optimality condition associated with the saddle point problem

$$\min_x \max_y \langle Kx, y\rangle + g(x) - f^*(y).$$

It can be shown, by means of a rescaling procedure, that their method and assumptions coincide with the exact BD-HPE method for the above inclusion problem (see Section 5.1) with the assumption that $\sigma < 1$. It should be noted however that, in contrast to our analysis, theirs does not deal with the extreme case of $\sigma = 1$ which, as mentioned earlier, is crucial to the analysis of the AMAL method.

References

[1] Regina S. Burachik, Alfredo N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.

[2] Regina S. Burachik, Claudia A. Sagastizábal, and B. F. Svaiter. ϵ-enlargements of maximal monotone operators: theory and applications. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 25–43. Kluwer Acad. Publ., Dordrecht, 1999.

[3] Regina Sandra Burachik and B. F. Svaiter. Maximal monotone operators, convex functions and a special family of enlargements. *Set-Valued Anal.*, 10(4):297–316, 2002.

[4] Samuel Burer. Optimizing a polyhedral-semidefinite relaxation of completely positive programs. *Math. Program. Comput.*, 2(1):1–19, 2010.

[5] Samuel Burer and Dieter Vandenbussche. Solving lift-and-project relaxations of binary integer programs. *SIAM J. Optim.*, 16(3):726–750, 2006.

[6] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Optimization online e-prints, CMPA, Ecole Polytechnique, Palaiseau, France, june 2010.

[7] Jonathan Eckstein and B. F. Svaiter. A family of projective splitting methods for the sum of two maximal monotone operators. *Math. Program.*, 111(1-2, Ser. B):173–199, 2008. Published as IMPA Tech. Rep. A 238/2003 in 2003.

[8] Jonathan Eckstein and B. F. Svaiter. General projective splitting methods for sums of maximal monotone operators. *SIAM J. Control Optim.*, 48(2):787–811, 2009.

[9] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems, Volume II.* Springer-Verlag, New York, 2003.

[10] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2:17–40, 1976.

[11] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. 1975.

[12] Florian Jarre and Franz Rendl. An augmented primal-dual method for linear conic programs. *SIAM J. Optim.*, 19(2):808–823, 2008.

[13] Jérôme Malick, Janez Povh, Franz Rendl, and Angelika Wiegele. Regularization methods for semidefinite programming. *SIAM J. Optim.*, 20(1):336–356, 2009.

[14] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal projection method for the iterates and the ergodic mean. Manuscript, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA, March 2009. To appear in *SIAM Journal on Optimization*.

[15] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of tseng's modified f-b splitting and korpelevich's methods for generalized variational inequalities with applications to saddle point and convex optimization problems. Manuscript, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA, July 2010. Submitted to *SIAM Journal on Optimization*.

[16] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15:229–251, 2005.

[17] Adam Ouorou. Epsilon-proximal decomposition method. *Math. Program.*, 99(1, Ser. A):89–108, 2004.

[18] J. Povh, F. Rendl, and A. Wiegele. A boundary point method to solve semidefinite programs. *Computing*, 78(3):277–286, 2006.

[19] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33:209–216, 1970.

[20] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[21] M. V. Solodov. A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework. *Optim. Methods Softw.*, 19(5):557–575, 2004.

[22] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[23] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[24] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, 25(2):214–230, 2000.

[25] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[26] B. F. Svaiter. A family of enlargements of maximal monotone operators. *Set-Valued Anal.*, 8(4):311–328, 2000.

[27] Paul Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446 (electronic), 2000.