Instituto Nacional de Matemática Pura e Aplicada

SPLITTING METHODS FOR THE SUM OF TWO MAXIMAL MONOTONE OPERATORS AND THEIR COMPLEXITY ANALYSIS

Majela Pentón Machado
Advisor: Prof. Benar Fux Svaiter

Rio de Janeiro
March, 2016

Abstract

In this thesis we are concerned with the study of pointwise and ergodic iteration-complexities of a family of splitting-projective methods proposed by Eckstein and Svaiter, for finding a zero of the sum of two maximal monotone operators. We also present two inexact variants of specific instances of this family of algorithms, and derive corresponding convergence rate results.

Keywords: monotone operators, splitting algorithms, complexity, projective algorithms, inclusion problem.

Resumo

Nesta tese apresentamos a análise de complexidade de uma família de métodos de decomposição-projetiva proposta por Eckstein e Svaiter para resolver o problema de encontrar um zero da soma de dois operadores monótonos maximais. Também introduzimos variantes inexatas de dois casos específicos desta família de algoritmos e obtemos suas taxas de convergência.

Palavras-chave: operadores monótonos, métodos de decomposição, métodos projetivos, complexidade, problema de inclusão.

Acknowledgements

I would like to express my gratitude to my advisor, Benar Fux Svaiter, for guiding my research and for his patience. Also, I would like to thank the committee members: Alfredo Iusem, Antonio Leitão, Mauricio Sicre and Hector Flores. Thanks for the useful comments which helped to improve this thesis. Special thanks to Alfredo and Antonio for all the support and advices.

I am grateful to the Department of Ensino of IMPA for the support to finish this work. In special Isabel Cherques and Andrea Nascimento.

My deepest thanks to my family, specially to my parents Odalis and Raúl, for their encouragement and love, despite the distance. To my sister Mariela my many thanks for her friendship and constant support. I leave my most sincere gratitude to my husband Edgar for his infinite patience and for always being there for me.

I am greatly indebted to Carlos Antonio Galeano Ríos for a careful reading of this work and the many helpful suggestions which improved the exposition considerably. I would also like to thank him for his friendship.

Many thanks also go to all my friends and colleagues. Specially I thank to Felippe, Susana, Juan and Diego, for their companionship and friendship.

Also, I would like to thank Emanuel Carneiro for his support to start my Ph.D. studies, and to the IMPA for the great environment of work. I acknowledge CAPES and FAPERJ for the financial support.

Contents

Introduction **1**

1 Background material **3**

1.1 Preliminaries . . . 3

1.2 The general splitting-projection framework . . . 6

1.2.1 The generated sequences . . . 9

1.2.2 The ergodic sequences . . . 12

2 Splitting-projective methods **15**

2.1 Complexity analysis . . . 15

2.2 Specialized complexity results . . . 21

2.3 Spingarn's splitting method . . . 23

2.4 A parallel inexact case . . . 27

2.5 A sequential inexact case . . . 35

3 Application **43**

3.1 Convex optimization problems . . . 43

3.2 TV denoising . . . 50

3.2.1 Numerical tests . . . 53

Bibliography **61**

Introduction

Monotone operators were first introduced by Minty in the early sixties [25] for the study of electrical networks. Since then, these operators have been used in several branches of applied mathematics such as optimization, partial differential equations, mathematical economics, and much more.

The classical *monotone inclusion problem* (MIP) consists in locating a root of a maximal monotone operator. This problem provides a powerful framework for the study of a broad class of problems, which include optimization problems, variational inequalities and equilibrium problems.

In this thesis we are concerned with the MIP defined by the sum of two maximal monotone operators. *Splitting methods* for this problem, which are inspired by decomposition techniques from numerical linear algebra, encompasses extensive literature with a variety of applications, see for instances [18, 23, 30, 17, 19, 13, 40, 41, 10, 14, 3], and references therein. This list is surely not exhaustive and is given simply as a statement of the richness and fruitfulness of the field.

A family of splitting algorithms for finding a zero of the sum of two maximal monotone operators was introduced in [14] by Eckstein and Svaiter. Their framework is based on reformulating the MIP as a convex feasibility problem defined by a certain closed convex "extended" solution set. For this latter problem the authors presented successive projection algorithms which use, on each iteration, independent calculations involving each operator. In this thesis we study this family of methods and establish its iteration-complexity.

Our complexity study for the methods described in [14] is motivated by the analysis presented in [27] and the subsequent papers [28, 29]; where the *hybrid proximal extragradient* method [36] was used as a general framework to derive iteration-complexity results for specific algorithms for solving various type of structured MIPs.

Outline of the Thesis. This work is organized in three chapters. Chapter 1 contains two sections. Section 1.1 reviews some definitions and facts on maximal monotone operators and convex functions that will be used along this work. Section 1.2 presents a relaxed projection method which extends the general framework introduced in [14]; and also proves some properties of such method.

Chapter 2 is devoted to the study of specific instances of the general framework proposed in Section 1.2. It contains five sections as follows. Section 2.1 introduces the family of splitting-projective methods proposed in [14] for solving the MIP of the sum of two maximal monotone operators. Global convergence rate results for this family of algorithms are also obtained in this section. Section 2.2 specializes the general complexity bounds that were derived in the previous section, for the case where global convergence for the methods were obtained in [14]. Section 2.3 derives iteration-complexity bounds for Spingarn's method of partial inverses [39]. Sections 2.4 and 2.5 introduce two inexact versions of the method discussed in Section 2.1 and derives iteration-complexity results for them.

Chapter 3 presents an application of the algorithm proposed in Section 2.5 to convex programming problems. Section 3.1 studies a splitting-projective method for solving a class of linearly constrained optimization problems with proper closed convex objective functions. This algorithm is applied, in Section 3.2, to solve the TV denoising model for image restoration [35]. Finally, Subsection 3.2.1, describes some preliminary computational experiments.

Notations

$\mathbb{R}$	real numbers
$\mathbb{R}_+$	nonnegative real numbers
$\mathbb{E}$	the product set $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$
$\langle \cdot, \cdot \rangle$	inner product
$\\|\cdot\\|$	norm associated to the inner product
$T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$	point-to-set operator
$\mathrm{Gr}\,(T)$	graph of T
$f : \mathbb{R}^n \to (-\infty, \infty]$	extend real valued function
$\mathrm{dom}\, f$	domain of f
∂f	subdifferential of f
f^*	Fenchel-Legendre conjugate of f
∇f	gradient of a differentiable function f
$\mathrm{ri}\,(D)$	relative interior of the convex set D
$\\|x\\|_1$	ℓ_1 norm $\sum_{i=1}^n \lvert x_i \rvert$
M^T	transpose of matrix M
$\\|M\\|_F$	Frobenius norm of matrix M, $\sqrt{\mathrm{trace}(M^T M)}$
$sgn(\cdot)$	sign function
C^*	the adjoint operator of the linear map C

Chapter 1

Background material

We present in this chapter the definitions and some basic properties of a point-to-set maximal monotone operator and its ϵ-enlargements. We also describe some basic definitions and facts on convex functions. These results will be used in our subsequent presentation.

In Section 1.2 we introduce the general splitting-projective scheme that we are interested in study in this work and we prove some technical results regarding it. These results will be useful for our complexity study of specific instances of the general framework, which will be developed in the following chapters.

1.1 Preliminaries

A point-to-set operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a relation $T \subseteq \mathbb{R}^n \times \mathbb{R}^n$ and

$$T(x) = \{v \in \mathbb{R}^n : (x, v) \in T\}, \qquad x \in \mathbb{R}^n.$$

Given $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ its graph is the set

$$\mathrm{Gr}\,(T) = \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n : v \in T(x)\}.$$

An operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *monotone* if

$$\left\langle x - x', v - v' \right\rangle \geq 0 \qquad \forall (x, v), (x', v') \in \mathrm{Gr}\,(T)\,,$$

and it is *maximal monotone* if it is monotone and maximal in the family of monotone operators of $\mathbb{R}^n$ into $\mathbb{R}^n$, with respect to the partial order of inclusion. This is, if $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a monotone operator such that $\mathrm{Gr}\,(T) \subseteq \mathrm{Gr}\,(S)$, then $S = T$.

Given an extended real valued function $f : \mathbb{R}^n \to (-\infty, \infty]$, the domain of f is the set

$$\operatorname{dom} f = \{x \in \mathbb{R}^n \,:\, f(x) < \infty\}\,.$$

We will say that f is a *proper* function if $\operatorname{dom} f \neq \emptyset$. We will also say that f is *closed* if it is a lower semicontinuous function.

The *subdifferential* of f is the point-to-set operator $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined as

$$\partial f(x) = \left\{v \,:\, f(x') \geq f(x) + \left\langle v, x' - x\right\rangle \quad \forall x' \in \mathbb{R}^n \right\}, \quad \forall x \in \mathbb{R}^n.$$

A vector $v \in \mathbb{R}^n$ is called a *subgradient* of f at $x \in \mathbb{R}^n$ if $v \in \partial f(x)$. The operator ∂f is trivially monotone if f is proper. In addition, if f is a proper closed convex function, then ∂f is maximal monotone [33].

The *Fenchel-Legendre conjugate* of f is the function $f^* : \mathbb{R}^n \to (-\infty, \infty]$ defined as

$$f^*(v) = \sup_{x \in \mathbb{R}^n} \{\langle v, x\rangle - f(x)\}, \qquad \forall v \in \mathbb{R}^n.$$

It is simple to see that f^* is a convex closed function. Furthermore, if f is proper, closed and convex, then f^* is a proper function [4].

The *ϵ-enlargement* of a maximal monotone operator was introduced in [6] by Burachik, Iusem and Svaiter. In [27], Monteiro and Svaiter extended this notion to a generic point-to-set operator as follows. Given $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $\epsilon \in \mathbb{R}$, define the operator ϵ-enlargement of T, $T^\epsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, by

$$T^\epsilon(x) = \{v \in \mathbb{R}^n : \left\langle x' - x, v' - v\right\rangle \geq -\epsilon, \quad \forall (x', v') \in \operatorname{Gr}(T)\}, \quad \forall x \in \mathbb{R}^n.$$

The following proposition presents some important properties of T^ϵ, its proof can be found in [27].

Proposition 1.1. *Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Then,*

(a) if $\epsilon' \leq \epsilon$ it holds that $T^{\epsilon'}(x) \subseteq T^\epsilon(x)$ for all $x \in \mathbb{R}^n$;

(b) T is monotone if and only if $T \subseteq T^0$;

(c) T is maximal monotone if and only if $T = T^0$.

Observe that items (a) and (c) above imply that if $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone, then $T(x) \subseteq T^\epsilon(x)$ for all $x \in \mathbb{R}^n$ and $\epsilon \geq 0$, hence $T^\epsilon(x)$ is indeed an enlargement of $T(x)$.

The ϵ-enlargement is a generalization of the *ϵ-subdifferential* of a extended real function.

Given $\epsilon \geq 0$, the ϵ-subdifferential of a function f is the operator $\partial_\epsilon f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined as

$$\partial_\epsilon f(x) = \{v \,:\, f(x') \geq f(x) + \langle v, x' - x\rangle - \epsilon, \quad \forall x' \in \mathbb{R}^n\} \quad \forall x \in \mathbb{R}^n.$$

It is trivial to verify that $\partial_0 f(x) = \partial f(x)$ and $\partial f(x) \subseteq \partial_\epsilon f(x)$, for every $x \in \mathbb{R}^n$ and $\epsilon \geq 0$. The proposition below lists some useful properties of the ϵ-subdifferential of a proper closed convex function which will be needed in our presentation.

Proposition 1.2. *If $f : \mathbb{R}^n \to (-\infty, \infty]$ is a proper closed convex function, $g : \mathbb{R}^n \to \mathbb{R}$ is a convex differentiable function in $\mathbb{R}^n$, and $M : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation, then*

(a) $\partial_\epsilon f(x) \subseteq (\partial f)^\epsilon(x)$ *for any* $\epsilon \geq 0$ *and* $x \in \mathbb{R}^n$*;*

(b) $v \in \partial_\epsilon f(x)$ *if and only if* $x \in \partial_\epsilon f^*(v)$ *for all* $\epsilon \geq 0$*;*

(c) $\partial(f + g)(x) = \partial f(x) + \nabla g(x)$ *for all* $x \in \mathbb{R}^n$*;*

(d) $M^*\partial f(Mx) \subseteq \partial(f \circ M)(x)$ *for all* $x \in \mathbb{R}^m$*. If, in addition,* $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{range} M \neq \emptyset$*, then* $\partial(f \circ M)(x) = M^*\partial f(Mx)$ *for every* $x \in \mathbb{R}^m$*.*

Proof. Statement (a) was proved in [6, Proposition 3], and (b)-(d) are classical results which can be found, for example, in [22] and [32]. □

We now state the *weak transportation formula* [7] for computing points in the graph of T^ϵ. This formula will be used in the complexity analysis of some ergodic iterates generated by the algorithms studied in this work, see Subsection 1.2.2.

Theorem 1.1. *Assume that $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a maximal monotone operator. Let $x_i, v_i \in \mathbb{R}^n$ and $\epsilon_i, \alpha_i \in \mathbb{R}_+$, for $i = 1, \ldots, k$, be such that*

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

and define

$$\overline{x} = \sum_{i=1}^k \alpha_i x_i, \qquad \overline{v} = \sum_{i=1}^k \alpha_i v_i, \qquad \overline{\epsilon} = \sum_{i=1}^k \alpha_i(\epsilon_i + \langle x_i - \overline{x}, v_i\rangle).$$

Then, the following statements hold:

(a) $\overline{\epsilon} \geq 0$ *and* $\overline{v} \in T^{\overline{\epsilon}}(\overline{x})$*;*

(b) if, in addition, $T = \partial f$*, for some proper closed convex function* f*, and* $v_i \in \partial_{\epsilon_i} f(x_i)$ *for* $i = 1, \ldots, k$*, then* $\overline{v} \in \partial_{\overline{\epsilon}} f(\overline{x})$*.*

1.2 The general splitting-projection framework

If $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a maximal monotone operator the *monotone inclusion problem* (MIP) consists of

$$\text{finding} \quad z \in \mathbb{R}^n \quad \text{such that} \quad 0 \in T(z).$$

It is well known that MIPs have many applications, such as optimization and min-max problems, complementarity problems and variational inequalities. An important tool for the design and analysis of several implementable methods for solving MIPs is the *proximal point algorithm* (PPA), proposed by Martinet [24] and generalized by Rockafellar [34]. The PPA, in its exact version, computes a sequence $\{z_k\}$ obeying the recursion

$$z_{k+1} = (I + \lambda_k T)^{-1}(z_k)$$

where I is the identity mapping and $\lambda_k > 0$ is known as the *proximal parameter*. Even though the PPA has good global and local convergence properties [34], its major drawback is that it requires the evaluation of the *resolvent mapping* (or *proximal mapping*) $(I + \lambda_k T)^{-1}$. The difficulty lies in the fact that inverting operators $I + \lambda_k T$ can be equally complicated as solving the original problem.

One alternative to surmount this difficulty is to find maximal monotone operators A and B such that $T = A + B$ and the evaluation of the resolvents $(I + \lambda A)^{-1}$ and $(I + \lambda B)^{-1}$ is simple to do. Then, one can devise methods that use independently these proximal mappings.

In this work, we are interested in the problem of finding $z \in \mathbb{R}^n$ such that

$$0 \in A(z) + B(z) \tag{1.1}$$

where $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are maximal monotone operators.

Splitting methods for this problem generate sequences which converge to a solution using, in each iteration, the resolvents $(I + \lambda A)^{-1}$ and $(I + \lambda B)^{-1}$, rather than $(I + \lambda(A + B))^{-1}$. *Peaceman-Rachford* and *Douglas-Rachford* methods are examples of this type of algorithms. These were first introduced in [31] and [12] for the particular case of linear mappings and then generalized in [23] by Lions and Mercier, to address monotone inclusion problems. *Forward-Backward* methods [23, 30, 41], which generalize standard gradient projection methods for variational inequalities and optimization problems, are also examples of splitting algorithms for problem (1.1).

Eckstein and Svaiter introduced in [14] a family of splitting-projection methods to find a solution of problem (1.1). They defined a certain closed convex extended solution set in a product space, and constructed a class of methods for solving the MIP (1.1), which is essentially a standard projection method. For constructing the separating hyperplanes the authors used individually evaluations of the resolvent mappings $(I+\lambda A)^{-1}$ and $(I+\lambda B)^{-1}$, which make their scheme truly a splitting method for (1.1).

The *extended solution set* of (1.1), introduced in [14], is defined as

$$S_e(A,B) := \{(z,w) \in \mathbb{R}^n \times \mathbb{R}^n : \, w \in B(z), \, -w \in A(z)\}. \tag{1.2}$$

The next result, which establishes two important properties of $S_e(A,B)$, was proved in [14, Lemma 1]. For the sake of completeness, we include its proof here.

Lemma 1.1. *Let $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be two maximal monotone operators. The following statements hold.*

(a) A point $z \in \mathbb{R}^n$ is a solution of (1.1) if and only if there is $w \in \mathbb{R}^n$ such that $(z,w) \in S_e(A,B)$.

(b) $S_e(A,B)$ is a closed and convex subset of $\mathbb{R}^n \times \mathbb{R}^n$.

Proof. (a) A point $z \in \mathbb{R}^n$ satisfies $0 \in (A+B)(z)$ if and only if there is a $w \in \mathbb{R}^n$ such that $w \in B(z)$ and $-w \in A(z)$. Hence, z is a solution of (1.1) if and only if there exist w with $(z,w) \in S_e(A,B)$.

(b) Define the linear map $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ as $F(z,w) = (w,-z)$. Define also the maximal monotone operator $T : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n$ as $T(z,w) = A(z) \times B^{-1}(w)$. Observe now that $F+T$ is maximal monotone, and

$$\begin{aligned} 0 \in (F+T)(z,w) &\Leftrightarrow 0 \in w + A(z), \; 0 \in -z + B^{-1}(w) \\ &\Leftrightarrow -w \in A(z), \; w \in B(z) \\ &\Leftrightarrow (z,w) \in S_e(A,B). \end{aligned}$$

Hence, $S_e(A,B) = (F+T)^{-1}(0,0)$. The assertion (b) follows from the fact that the root set of a maximal monotone operator is a convex closed set, see [5]. □

According to the above lemma, problem (1.1) is equivalent to the convex feasibility problem of finding a point in $S_e(A,B)$. In order to solve this feasibility problem by successive orthogonal projection methods we need to construct hyperplanes separating points $(z,w) \notin S_e(A,B)$ from $S_e(A,B)$. For this purpose, in [14] it was used points in the graph of A and B to define affine functions, which were called *decomposable separators*, such that $S_e(A,B)$ was contained in the non-positive half-spaces determined by them. Here, we generalize this concept using points in the ϵ-enlargements of A and B.

Definition 1. *Given two triples (x,b,ϵ_x), $(y,a,\epsilon_y) \in \mathbb{E}$ such that $b \in B^{\epsilon_x}(x)$ and $a \in A^{\epsilon_y}(y)$, the* decomposable separator *associated to (x,b,ϵ_x) and (y,a,ϵ_y) is the affine function*

$\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\phi(z, w) = \langle z - x, b - w\rangle + \langle z - y, a + w\rangle - \epsilon_x - \epsilon_y. \tag{1.3}$$

The non-positive level set *of ϕ is*

$$H_\phi = \{(z, w) \ : \ \phi(z, w) \leq 0\}. \tag{1.4}$$

Lemma 1.2. *If ϕ is the decomposable separator associated to (x, b, ϵ_x) and $(y, a, \epsilon_y) \in \mathbb{E}$, where $b \in B^{\epsilon_x}(x)$ and $a \in A^{\epsilon_y}(y)$, and H_ϕ is its non-positive level set, then*

(a) $S_e(A, B) \subseteq H_\phi$;

(b) either $\nabla\phi \neq 0$ or $\phi \leq 0$ in $\mathbb{R}^n \times \mathbb{R}^n$;

(c) either H_ϕ is a closed half-space or $H_\phi = \mathbb{R}^n \times \mathbb{R}^n$.

Proof. Item (a) is a direct consequence of the definitions of the ϵ-enlargement of an operator and the set $S_e(A, B)$. Rewriting $\phi(z, w)$ as

$$\phi(z, w) = \langle z - y, a + b\rangle + \langle w - b, x - y\rangle - \epsilon_x - \epsilon_y \qquad \forall (z, w) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{1.5}$$

and noting that $\nabla\phi = (a + b, x - y)$ and ϵ_x, $\epsilon_y \geq 0$, then (b) and (c) follow immediately. □

In view of Lemma 1.2, if ϕ is a decomposable separator, then the orthogonal projection of $(z, w) \in \mathbb{R}^n \times \mathbb{R}^n$ onto H_ϕ is

$$P_{H_\phi}(z, w) = (z, w) - \gamma\nabla\phi, \qquad \text{where} \qquad \gamma = \begin{cases} 0 & \text{if } \phi(z, w) \leq 0, \\ \dfrac{\phi(z, w)}{\|\nabla\phi\|^2} & \text{otherwise.} \end{cases} \tag{1.6}$$

We now present the general projection scheme for finding a point in $S_e(A, B)$ that will be studied in this work. Algorithm 1 below generalizes the framework introduced in [14], since we use the notion of decomposable separator introduced in Definition 1.

Algorithm 1. *Choose $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$. For $k = 1, 2, \ldots$*

1. *Choose $(x_k, b_k, \epsilon_{x,k})$ and $(y_k, a_k, \epsilon_{y,k}) \in \mathbb{E}$ such that*

$$b_k \in B^{\epsilon_{x,k}}(x_k) \quad \text{and} \quad a_k \in A^{\epsilon_{y,k}}(y_k).$$

2. *Define $\phi_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as the decomposable separator associated to $(x_k, b_k, \epsilon_{x,k})$ and $(y_k, a_k, \epsilon_{y,k})$, and compute $P_{H_{\phi_k}}(z_{k-1}, w_{k-1})$ the orthogonal projection of (z_{k-1}, w_{k-1})*

onto H_{ϕ_k} given in (1.6), i.e. define

$$\gamma_k = \begin{cases} 0 & \text{if } \phi_k(z_{k-1}, w_{k-1}) \le 0, \\ \dfrac{\phi_k(z_{k-1}, w_{k-1})}{\|\nabla \phi_k\|^2} & \text{otherwise;} \end{cases} \tag{1.7}$$

and set

$$P_{H_\phi}(z_{k-1}, w_{k-1}) = (z_{k-1}, w_{k-1}) - \gamma_k \nabla \phi_k. \tag{1.8}$$

3. *Choose $\rho_k \in (0, 2)$ and set*

$$\begin{aligned}(z_k, w_k) &= (z_{k-1}, w_{k-1}) + \rho_k \left[P_{H_{\phi_k}}(z_{k-1}, w_{k-1}) - (z_{k-1}, w_{k-1})\right] \\ &= (z_{k-1}, w_{k-1}) - \rho_k \gamma_k \nabla \phi_k.\end{aligned}$$

Notice that the general form of Algorithm 1 is not sufficient to guarantee convergence of the sequence $\{(z_k, w_k)\}$ to a point in $S_e(A, B)$; for example, if the separation between $(z_{k-1}, w_{k-1}) \notin S_e(A, B)$ and $S_e(A, B)$ by ϕ_k is not strict, then the next iterate is in fact (z_{k-1}, w_{k-1}) itself, which might lead to a constant sequence. Hence, to ensure convergence it is necessary to impose additional conditions on the decomposable separators, see [14] and Section 2 below. However, since Algorithm 1 is a relaxed projection type method it is possible to establish Fejer monotone convergence to $S_e(A, B)$ and boundedness for its generated sequence, as well as other classical properties for this kind of algorithms, see for example [14], [1].

1.2.1 The generated sequences

In this subsection we will analyze some properties of the sequences $\{(z_k, w_k)\}$, $\{\phi_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ generated by Algorithm 1. To that end, let us define the aggregate affine maps $\Phi_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ recursively as

$$\Phi_0 \equiv 0, \qquad \Phi_k = \sum_1^k \rho_j \gamma_j \phi_j \qquad k = 1, 2, \dots \tag{1.9}$$

and β_k as

$$\beta_k = \min_{(z,w)} \left[\frac{1}{2}\|(z, w) - (z_0, w_0)\|^2 + \Phi_k(z, w)\right] \qquad k = 0, 1, \dots. \tag{1.10}$$

Lemma 1.3. *For all $k \in \mathbb{N}$, the following claims hold:*

(a) $(z_k, w_k) = \arg\min_{(z,w)} \left(\frac{1}{2}\|(z, w) - (z_0, w_0)\|^2 + \Phi_k(z, w)\right)$;

(b) $\beta_{k+1} = \beta_k + \frac{1}{2}(2 - \rho_{k+1})\rho_{k+1}\gamma_{k+1}^2 \|\nabla \phi_{k+1}\|^2$;

(c) $\beta_k \geq 0$.

Proof. To prove (a), first we observe that since $\Phi_0 \equiv 0$, it holds that $(z_0, w_0) = \arg\min \frac{1}{2} \|(z,w) - (z_0,w_0)\|^2 + \Phi_0(z,w)$. Next, we notice that for all integer $k \geq 1$ we have

$$\nabla \Phi_k(z,w) = \sum_{j=1}^{k} \rho_j \gamma_j \nabla \phi_j, \qquad \forall (z,w) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Moreover, the update rule in step 3 of Algorithm 1 implies that $(z_k, w_k) = (z_0, w_0) - \sum_{j=1}^{k} \rho_j \gamma_j \nabla \phi_j$. Combining these two expressions we obtain $(z_k, w_k) - (z_0, w_0) + \nabla \Phi_k(z_k, w_k) = 0$. Hence, (z_k, w_k) satisfies the optimality condition for the minimization problem in (a), and the claim in (a) follows.

To prove (b) we first notice that if $\phi_{k+1}(z_k, w_k) \leq 0$ then by (1.7) we have $\gamma_{k+1} = 0$. Therefore, $\Phi_{k+1} = \Phi_k$ and (b) follows from these two latter equalities.

We assume now that $\phi_{k+1}(z_k, w_k) > 0$. We observe that

$$\begin{aligned} \beta_{k+1} &= \min_{(z,w)} \left[\frac{1}{2} \|(z,w) - (z_0,w_0)\|^2 + \Phi_k(z,w) + \rho_{k+1}\gamma_{k+1}\phi_{k+1}(z,w) \right] \\ &= \min_{(z,w)} \left[\beta_k + \frac{1}{2} \|(z,w) - (z_k,w_k)\|^2 + \rho_{k+1}\gamma_{k+1}\phi_{k+1}(z,w) \right]. \end{aligned} \tag{1.11}$$

By statement (a) we have that the optimal solution of the problem above is

$$(z_{k+1}, w_{k+1}) = (z_k, w_k) - \rho_{k+1}\gamma_{k+1}\nabla\phi_{k+1}. \tag{1.12}$$

Therefore, taking $(z,w) = (z_{k+1}, w_{k+1})$ in (1.11), we have

$$\beta_{k+1} = \beta_k + \frac{1}{2} \|\rho_{k+1}\gamma_{k+1}\nabla\phi_{k+1}\|^2 + \rho_{k+1}\gamma_{k+1}\phi_{k+1}(z_{k+1}, w_{k+1}). \tag{1.13}$$

Next we observe that

$$\begin{aligned} \phi_{k+1}(z_{k+1}, w_{k+1}) &= \langle (z_{k+1}, w_{k+1}) - (y_{k+1}, b_{k+1}), \nabla\phi_{k+1} \rangle - \epsilon_{x,k+1} - \epsilon_{y,k+1} \\ &= \langle (z_k, w_k) - (y_{k+1}, b_{k+1}), \nabla\phi_{k+1} \rangle - \rho_{k+1}\gamma_{k+1} \|\nabla\phi_{k+1}\|^2 - \epsilon_{x,k+1} - \epsilon_{y,k+1} \\ &= \phi_{k+1}(z_k, w_k) - \rho_{k+1}\gamma_{k+1} \|\nabla\phi_{k+1}\|^2 \\ &= \gamma_{k+1} \|\nabla\phi_{k+1}\|^2 - \rho_{k+1}\gamma_{k+1} \|\nabla\phi_{k+1}\|^2, \end{aligned}$$

where the first equality follows rewriting ϕ_{k+1} as in (1.5) and the last one follows from the definition of γ_{k+1} in (1.7). To end the proof of (b) we replace this last identity into (1.13).

Since $\beta_0 = 0$ by equations (1.9) and (1.10), statement (c) follows from item (b) and a simple induction argument. □

Proposition 1.3. *For every* $(z,w) \in \mathbb{R}^n \times \mathbb{R}^n$ *and* $k \geq 1$,

$$\frac{1}{2}\|(z,w)-(z_0,w_0)\|^2 + \Phi_k(z,w) = \frac{1}{2}\|(z,w)-(z_k,w_k)\|^2 + \frac{1}{2}\sum_{j=1}^{k}\rho_j(2-\rho_j)\gamma_j^2\|\nabla\phi_j\|^2. \tag{1.14}$$

Proof. Adding the identity in Lemma 1.3(b) from $j=0$ to k and using that $\beta_0 = 0$, we have

$$\beta_k = \frac{1}{2}\sum_{j=1}^{k}\rho_j(2-\rho_j)\gamma_j^2\|\nabla\phi_j\|^2. \tag{1.15}$$

Moreover, in view of Lemma 1.3(a) and the definition of β_k it holds that

$$\frac{1}{2}\|(z,w)-(z_0,w_0)\|^2 + \Phi_k(z,w) = \beta_k + \frac{1}{2}\|(z,w)-(z_k,w_k)\|^2, \quad \forall (z,w) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Thus, replacing (1.15) into the above equality we conclude (1.14). □

The following theorem, which follows directly from Proposition 1.3, establishes boundedness for the sequence $\{(z_k,w_k)\}$ generated by Algorithm 1 and for the sequence of minima $\{\beta_k\}$ defined in (1.10).

Theorem 1.2. *Take* $(z_0,w_0) \in \mathbb{R}^n \times \mathbb{R}^n$ *and let* $\{(z_k,w_k)\}$, $\{\phi_k\}$, $\{\gamma_k\}$, *and* $\{\rho_k\}$ *be the sequences generated by Algorithm 1. If* $S_e(A,B) \neq \emptyset$, *then for every integer* $k \geq 1$ *we have*

$$2\beta_k = \sum_{j=1}^{k}\rho_j(2-\rho_j)\gamma_j^2\|\nabla\phi_j\|^2 \leq d_0^2 \qquad \text{and} \qquad \|(z_k,w_k)-(z_0,w_0)\| \leq 2d_0, \tag{1.16}$$

where d_0 *is the distance of* (z_0,w_0) *to* $S_e(A,B)$.

Proof. Take (z^*,w^*) the orthogonal projection of (z_0,w_0) onto $S_e(A,B)$. From Lemma 1.2(a) and the definition of Φ_k in (1.9) it follows that $\Phi_k(z^*,w^*) \leq 0$ for all integer $k \geq 1$. Hence, specializing equality (1.14) with (z^*,w^*) we obtain the first bound in (1.16) and the following inequality

$$\|(z_k,w_k)-(z^*,w^*)\| \leq d_0.$$

Since $\|(z_0,w_0)-(z^*,w^*)\| = d_0$, the second estimate in (1.16) follows now from the latter two relations and the triangle inequality for norms. □

It is important to say that Theorem 1.2 can be proved using standard arguments of relaxed projection algorithms. We have chosen the above approach since it will be more convenient for our subsequent analysis.

1.2.2 The ergodic sequences

We now consider sequences obtained by weighted averages of the sequences $\{x_k\}$ and $\{y_k\}$ generated by Algorithm 1 and study its properties.

Let $\{x_k\}$, $\{y_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences computed with Algorithm 1, for every integer $k \geq 1$ assume that $\gamma_k > 0$ and define $\overline{x}_k$ and $\overline{y}_k$ as

$$\overline{x}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j x_j, \qquad \overline{y}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j y_j, \qquad \text{where} \quad \Gamma_k := \sum_{j=1}^{k} \rho_j\gamma_j. \tag{1.17}$$

The following lemma is a direct consequence of the weak transportation formula, Theorem 1.1.

Lemma 1.4. *Let $\{(x_k, b_k, \epsilon_{x,k})\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 1. For every integer $k \geq 1$, suppose that $\gamma_k > 0$ and consider $\overline{x}_k$, $\overline{y}_k$ and Γ_k given as in (1.17), define also*

$$\overline{b}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j b_j, \qquad \overline{\epsilon}_{x,k} := \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j(\epsilon_{x,j} + \langle x_j - \overline{x}_k, b_j\rangle), \tag{1.18}$$

$$\overline{a}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j a_j, \qquad \overline{\epsilon}_{y,k} := \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j(\epsilon_{y,j} + \langle y_j - \overline{y}_k, a_j\rangle). \tag{1.19}$$

Then,

$$\overline{\epsilon}_{x,k} \geq 0, \qquad \overline{b}_k \in B^{\overline{\epsilon}_{x,k}}(\overline{x}_k), \tag{1.20}$$

$$\overline{\epsilon}_{y,k} \geq 0, \qquad \overline{a}_k \in A^{\overline{\epsilon}_{y,k}}(\overline{y}_k). \tag{1.21}$$

Proof. Since for all integer $k \geq 1$, it holds that $b_k \in B^{\epsilon_{x,k}}(x_k)$ and $a_k \in A^{\epsilon_{y,k}}(y_k)$, relations (1.20) and (1.21) follow from Theorem 1.1. □

We will refer to the sequences $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$ and $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{y,k})\}$ defined in (1.17)-(1.18)-(1.19) as the *ergodic sequences* associated to Algorithm 1.

The following lemma presents a more suitable manner of writing $\overline{a}_k + \overline{b}_k$, $\overline{x}_k - \overline{y}_k$ and $\overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k}$ to obtain bounds on its size.

Lemma 1.5. *Let $\{(x_k, b_k, \epsilon_{x,k})\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 1. Assume that $\gamma_k > 0$ for all $k \geq 1$, and define the sequences $\{\overline{x}_k\}$, $\{\overline{y}_k\}$, $\{\Gamma_k\}$, $\{\overline{b}_k\}$, $\{\overline{a}_k\}$ $\{\overline{\epsilon}_{x,k}\}$ and $\{\overline{\epsilon}_{y,k}\}$ as in* (1.17), (1.18) *and* (1.19). *Consider also the sequences of* *aggregate functions* $\{\Phi_k\}$ *defined in* (1.9). *Then, for every integer* $k \geq 1$, *we have*

$$\bar{a}_k + \bar{b}_k = \frac{1}{\Gamma_k}(z_0 - z_k), \tag{1.22}$$

$$\bar{x}_k - \bar{y}_k = \frac{1}{\Gamma_k}(w_0 - w_k), \tag{1.23}$$

$$\bar{\epsilon}_{x,k} + \bar{\epsilon}_{y,k} = -\frac{1}{\Gamma_k}\Phi_k(\bar{y}_k, \bar{b}_k). \tag{1.24}$$

Proof. Direct use of the definitions of $\bar{x}_k$, $\bar{y}_k$, $\bar{b}_k$ and $\bar{a}_k$, yields

$$(\bar{a}_k + \bar{b}_k, \bar{x}_k - \bar{y}_k) = \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j(a_j + b_j, x_j - y_j). \tag{1.25}$$

Since $\nabla\phi_k = (a_k + b_k, x_k - y_k)$ for all integer $k \geq 1$, in view of the update rule in step 3 of Algorithm 1, the definition of Γ_k and (1.25), we have

$$\begin{aligned}(z_k, w_k) &= (z_0, w_0) - \sum_{j=1}^{k} \rho_j\gamma_j(a_j + b_j, x_j - y_j) \\ &= (z_0, w_0) - \Gamma_k(\bar{a}_k + \bar{b}_k, \bar{x}_k - \bar{y}_k).\end{aligned} \tag{1.26}$$

Relation (1.26) clearly implies identities (1.22) and (1.23).

To prove (1.24) first we notice that

$$\begin{aligned}\Phi_k(\bar{y}_k, \bar{b}_k) &= \sum_{j=1}^{k} \rho_j\gamma_j \left(\langle \bar{y}_k - x_j, b_j - \bar{b}_k\rangle + \langle \bar{y}_k - y_j, a_j + \bar{b}_k\rangle - \epsilon_{x,j} - \epsilon_{y,j}\right) \\ &= \sum_{j=1}^{k} \rho_j\gamma_j \left(\langle x_j, \bar{b}_k - b_j\rangle + \langle \bar{y}_k, b_j\rangle + \langle \bar{y}_k - y_j, a_j\rangle - \langle y_j, \bar{b}_k\rangle - \epsilon_{x,j} - \epsilon_{y,j}\right).\end{aligned}$$

Next, we multiply this last equality by $1/\Gamma_k$, and use the definitions of $\bar{y}_k$ and $\bar{b}_k$ to obtain

$$\frac{1}{\Gamma_k}\Phi_k(\bar{y}_k, \bar{b}_k) = \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j \left(\langle x_j, \bar{b}_k - b_j\rangle + \langle \bar{y}_k - y_j, a_j\rangle - \epsilon_{x,j} - \epsilon_{y,j}\right). \tag{1.27}$$

We now observe that $\bar{\epsilon}_{x,k}$ can be rewritten as

$$\bar{\epsilon}_{x,k} = \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j \left(\epsilon_{x,j} + \langle x_j, b_j - \bar{b}_k\rangle\right).$$

Thus, adding $\bar{\epsilon}_{x,k}$ and $\bar{\epsilon}_{y,k}$ and noting (1.27), we deduce equality (1.24). □

Theorem 1.3. *Assume the hypotheses in Lemma 1.5. Let* d_0 *be the distance of* (z_0, w_0) *to* $S_e(A, B)$*. Then, for all integer* $k \geq 1$*, we have*

$$\left\|\bar{a}_k + \bar{b}_k\right\| \leq \frac{2d_0}{\Gamma_k}, \qquad \|\bar{x}_k - \bar{y}_k\| \leq \frac{2d_0}{\Gamma_k}, \tag{1.28}$$

$$\bar{\epsilon}_{x,k} + \bar{\epsilon}_{y,k} \leq \frac{1}{\Gamma_k}\left[\frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j \left\|(y_j, b_j) - (z_{j-1}, w_{j-1})\right\|^2 + 4d_0^2\right]. \tag{1.29}$$

Proof. Combining (1.22), (1.23) and the second inequality in (1.16), we have

$$\left\|(\bar{a}_k + \bar{b}_k, \bar{x}_k - \bar{y}_k)\right\| = \frac{1}{\Gamma_k}\left\|(z_0, w_0) - (z_k, w_k)\right\| \leq \frac{2d_0}{\Gamma_k}. \tag{1.30}$$

Hence, the bounds in (1.28) follow.

We notice that, since Φ_k is an affine function, the definitions of $\bar{y}_k$, $\bar{b}_k$ and Γ_k yield

$$\Phi_k(\bar{y}_k, \bar{b}_k) = \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j\Phi_k(y_j, b_j). \tag{1.31}$$

Now, using the definition of β_k in (1.10), we obtain

$$\beta_k \leq \frac{1}{2}\left\|(y_j, b_j) - (z_0, w_0)\right\|^2 + \Phi_k(y_j, b_j), \qquad \text{for all } j = 1, \ldots k.$$

Therefore, since $\beta_k \geq 0$ we have

$$\begin{aligned} -\Phi_k(y_j, b_j) &\leq \frac{1}{2}\left\|(y_j, b_j) - (z_0, w_0)\right\|^2 \\ &\leq \left\|(y_j, b_j) - (z_{j-1}, w_{j-1})\right\|^2 + \left\|(z_{j-1}, w_{j-1}) - (z_0, w_0)\right\|^2 \\ &\leq \left\|(y_j, b_j) - (z_{j-1}, w_{j-1})\right\|^2 + 4d_0^2, \end{aligned}$$

where the last inequality follows from the second bound in (1.16). Multiplying this latter inequality by $\frac{1}{\Gamma_k}\rho_j\gamma_j$, adding from $j = 1$ to k and noting (1.31), we obtain

$$-\Phi_k(\bar{y}_k, \bar{b}_k) \leq \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j\left\|(y_j, b_j) - (z_{j-1}, w_{j-1})\right\|^2 + 4d_0^2.$$

The bound in (1.29) follows combining the above relation with (1.24). □

Chapter 2

Splitting-projective methods

This chapter is concerned with the complexity analysis of particular instances of the general framework of Algorithm 1.

We start our analysis by defining a termination criterion for the methods in terms of the ϵ-enlargement of the operators A and B. This criterion will enable the obtention of complexity bounds, proportional to the distance of the initial iterate to the extended solution set $S_e(A, B)$, for all the schemes presented in the chapter.

In Section 2.1 we will obtain general complexity bounds for the family of splitting-projective methods introduced in [14]. Such bounds will be expressed in terms of the parameter sequences $\{\lambda_k\}$, $\{\mu_k\}$, $\{\alpha_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$, calculated in each iteration of the method, see Algorithm 2 below. In Section 2.2, we will specialize these results for the case where global convergence was obtained in [14].

Section 2.3 is devoted to derive complexity estimates for Spingarn's splitting algorithm [39] for the two-operator case. Finally, in Sections 2.4 and 2.5, two inexact variants of the scheme presented in Section 2.1 will be proposed.

The ideas of our analysis are very similar to the ones used in [27] for obtaining iteration complexity for the *Hybrid Proximal Extragradient* (HPE) method, which was proposed in [36] by Solodov and Svaiter.

2.1 Complexity analysis

Throughout this chapter, we assume that problem (1.1) has at least one solution, which implies that $S_e(A, B)$ is a non-empty set in view of Lemma 1.1.

We start by stating the splitting-projective method that will be studied in this section.

Algorithm 2. *Choose* $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$. *For* $k = 1, 2, \dots$

1. Choose $\lambda_k, \mu_k > 0$ *and* $\alpha_k \in \mathbb{R}$ *such that*

$$\frac{\mu_k}{\lambda_k} - \left(\frac{\alpha_k}{2}\right)^2 > 0, \tag{2.1}$$

and find $(x_k, b_k) \in \mathrm{Gr}\,(B)$ *and* $(y_k, a_k) \in \mathrm{Gr}\,(A)$ *such that*

$$\lambda_k b_k + x_k = z_{k-1} + \lambda_k w_{k-1}, \tag{2.2}$$

$$\mu_k a_k + y_k = (1-\alpha_k) z_{k-1} + \alpha_k x_k - \mu_k w_{k-1}. \tag{2.3}$$

2. If $\|a_k + b_k\| + \|x_k - y_k\| = 0$ *stop. Otherwise, set*

$$\gamma_k = \frac{\langle z_{k-1} - x_k, b_k - w_{k-1}\rangle + \langle z_{k-1} - y_k, a_k + w_{k-1}\rangle}{\|a_k + b_k\|^2 + \|x_k - y_k\|^2}. \tag{2.4}$$

3. Choose a parameter $\rho_k \in (0, 2)$ *and set*

$$\begin{aligned} z_k &= z_{k-1} - \rho_k \gamma_k (a_k + b_k), \\ w_k &= w_{k-1} - \rho_k \gamma_k (x_k - y_k). \end{aligned} \tag{2.5}$$

Several remarks are in order. Algorithm 2 is the same as [14, Algorithm 2], except for the stopping criterion in step 2 above, and boundedness conditions imposed in the parameters ρ_k, λ_k and μ_k in [14]. Notice that if $\|a_k + b_k\| + \|x_k - y_k\| = 0$ for some k, then $x_k = y_k$, $b_k = -a_k$ and since the points (x_k, b_k) and (y_k, a_k) are chosen in the graph of B and A, respectively, we have $(x_k, b_k) \in S_e\,(A, B)$. Therefore, when Algorithm 2 stops in step 2, it has found a point in the extended solution set.

Observe also that since A and B are maximal monotone, Minty's theorem [26] implies that the mappings $(I + \lambda_k B)^{-1}$ and $(I + \mu_k A)^{-1}$ are everywhere defined and single valued for all integer $k \geq 1$. Hence, by (2.2) and (2.3) the points (x_k, b_k) and (y_k, a_k) exist and are unique.

Moreover, if for all $k = 1, 2, \ldots$, we denote by ϕ_k the decomposable separator associated to the pair of points (x_k, b_k) and (y_k, a_k), calculated in step 1 of Algorithm 2, i.e. $\phi_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\phi_k(z, w) = \langle z - x_k, b_k - w\rangle + \langle z - y_k, a_k + w\rangle\,, \tag{2.6}$$

then, the update rule in step 3 of Algorithm 2 can be restated as

$$(z_k, w_k) = (z_{k-1}, w_{k-1}) - \rho_k \gamma_k \nabla \phi_k. \tag{2.7}$$

Consequently, Algorithm 2 falls within the general framework of Algorithm 1, and the results of Section 1.2 apply.

The next lemma, which will be useful for deriving complexity estimates for Algorithm 2, establishes two lower bounds for $\phi_k(z_{k-1}, w_{k-1})$.

Lemma 2.1. *If* $\{(x_k, b_k)\}$, $\{(y_k, a_k)\}$, $\{(z_k, w_k)\}$, $\{\lambda_k\}$, $\{\mu_k\}$, $\{\alpha_k\}$ *and* $\{\rho_k\}$ *are the sequences generated by Algorithm 2, and* $\{\phi_k\}$ *is the sequence of decomposable separators defined in* (2.6)*, then, for all integer* $k \geq 1$*, the following inequalities hold*

$$\phi_k(z_{k-1}, w_{k-1}) \geq \frac{\theta_k}{\delta_k}\left(\|a_k + b_k\|^2 + \|x_k - y_k\|^2\right), \tag{2.8}$$

$$\phi_k(z_{k-1}, w_{k-1}) \geq \frac{\theta_k}{\mu_k}\left(\|z_{k-1} - y_k\|^2 + \|w_{k-1} - b_k\|^2\right), \tag{2.9}$$

where $\delta_k := \mu_k + (1-\alpha_k)\lambda_k > 0$ *and* $\theta_k > 0$ *is the smallest eigenvalue of the matrix*

$$\begin{pmatrix} 1 & -\frac{\lambda_k|\alpha_k|}{2} \\ -\frac{\lambda_k|\alpha_k|}{2} & \lambda_k\mu_k \end{pmatrix}.$$

Proof. Inequality (2.8) was obtained in [14, Proposition 3] as part of the convergence proof for Algorithm 2 in [14], as were the assertions that θ_k, $\delta_k > 0$ under assumption (2.1). Therefore, we need only to prove here relation (2.9).

If we subtract y_k from both sides of (2.2) and rearrange terms, we obtain

$$x_k - y_k = z_{k-1} - y_k + \lambda_k(w_{k-1} - b_k). \tag{2.10}$$

Now adding $\mu_k b_k$ to both sides of (2.3) and rearranging the terms, yields

$$\begin{aligned} \mu_k(a_k + b_k) &= (1-\alpha_k)z_{k-1} + \alpha_k x_k - y_k - \mu_k(w_{k-1} - b_k) \\ &= \alpha_k(x_k - y_k) + (1-\alpha_k)(z_{k-1} - y_k) - \mu_k(w_{k-1} - b_k). \end{aligned} \tag{2.11}$$

Next, we replace (2.10) into (2.11) and divide by μ_k to obtain

$$\begin{aligned} a_k + b_k &= \frac{\alpha_k}{\mu_k}(z_{k-1} - y_k + \lambda_k(w_{k-1} - b_k)) + \frac{(1-\alpha_k)}{\mu_k}(z_{k-1} - y_k) - (w_{k-1} - b_k) \\ &= \frac{1}{\mu_k}(z_{k-1} - y_k) + \left(\frac{\alpha_k\lambda_k}{\mu_k} - 1\right)(w_{k-1} - b_k). \end{aligned} \tag{2.12}$$

Noting that

$$\phi_k(z_{k-1}, w_{k-1}) = \langle a_k + b_k, z_{k-1} - y_k\rangle + \langle x_k - y_k, w_{k-1} - b_k\rangle, \tag{2.13}$$

and combining the above identity with (2.10) and (2.12) we have

$$\begin{aligned}\phi_k(z_{k-1},w_{k-1}) &= \frac{1}{\mu_k}\left\|z_{k-1}-y_k\right\|^2+\frac{\alpha_k\lambda_k}{\mu_k}\left\langle z_{k-1}-y_k,w_{k-1}-b_k\right\rangle+\lambda_k\left\|w_{k-1}-b_k\right\|^2\\ &\geq \frac{1}{\mu_k}\left\|z_{k-1}-y_k\right\|^2-\frac{|\alpha_k|\lambda_k}{\mu_k}\left\|z_{k-1}-y_k\right\|\left\|w_{k-1}-b_k\right\|+\lambda_k\left\|w_{k-1}-b_k\right\|^2\\ &= \frac{1}{\mu_k}\begin{pmatrix}\|z_{k-1}-y_k\|\\ \|w_{k-1}-b_k\|\end{pmatrix}^T\begin{pmatrix}1 & -\frac{\lambda_k|\alpha_k|}{2}\\ -\frac{\lambda_k|\alpha_k|}{2} & \lambda_k\mu_k\end{pmatrix}\begin{pmatrix}\|z_{k-1}-y_k\|\\ \|w_{k-1}-b_k\|\end{pmatrix},\end{aligned}$$

where the inequality in the above relation follows from Cauchy-Schwartz inequality. Finally, (2.9) follows from the expression above and the definition of θ_k. □

Our goal in the remaining part of the section will be to derive complexity bounds for Algorithm 2. For simplicity, from now on we suppose that the method never stops in step 2, i.e. we are assuming that $\|\nabla\phi_k\| > 0$ for all integer $k \geq 1$. However, we observe that all the results presented hold if (x_k, b_k) is a point in $S_e\,(A, B)$ for some k.

Let us consider the following stopping criterion for Algorithm 1. Given an arbitrary pair of scalars δ, $\epsilon > 0$, Algorithm 1 will stop whenever it finds a pair of points (x, b, ϵ_x), (y, a, ϵ_y) $\in \mathbb{E}$ such that

$$b \in B^{\epsilon_x}(x), \qquad a \in A^{\epsilon_y}(y), \qquad \max\{\|a+b\|, \|x-y\|\} \leq \delta, \qquad \max\{\epsilon_x, \epsilon_y\} \leq \epsilon. \tag{2.14}$$

We observe that, in view of Proposition 1.1, if $\delta = \epsilon = 0$ the above termination criterion is reduced to $b \in B(x)$, $a \in A(y)$, $x = y$ and $b = -a$, in which case $(x, b) \in S_e(A, B)$.

Based on the termination condition (2.14) we can define the following notion of approximate solution for problem (1.1).

Definition 2. *For a given tolerance pair (δ, ϵ) of positive scalars, a pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ is called a (δ, ϵ)-*approximate solution *(or (δ, ϵ)-*solution*) of problem (1.1), if there exist $b, a \in \mathbb{R}^n$ and $\epsilon_x, \epsilon_y \in \mathbb{R}_+$ such that the relations in (2.14) hold.*

Notice that Algorithm 2 generates, on each iteration, a pair (x_k, y_k) and vectors b_k, $a_k \in \mathbb{R}^n$ such that the inclusions in (2.14) hold with $(x, b, \epsilon_x) = (x_k, b_k, 0)$ and $(y, a, \epsilon_y) = (y_k, a_k, 0)$. Hence, we can try to develop bounds for the quantities $\|a_k + b_k\|$ and $\|x_k - y_k\|$ to estimate when an iterate (x_k, y_k) is bound to satisfy the termination condition (2.14).

Theorem 2.1. *Let $\{(z_k, w_k)\}$, $\{(x_k, b_k)\}$, $\{(y_k, a_k)\}$, $\{\lambda_k\}$, $\{\mu_k\}$, $\{\alpha_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 2 and let $\{\phi_k\}$ be the sequence of decomposable separators defined in* (2.6)*. Then, for every integer $k \geq 1$, we have*

$$b_k \in B(x_k), \qquad\qquad a_k \in A(y_k), \tag{2.15}$$

and there exists an index $1 \leq i \leq k$ such that

$$\|a_i + b_i\|^2 + \|x_i - y_i\|^2 \leq \frac{d_0^2}{\sum_{j=1}^{k} \rho_j(2-\rho_j)\left(\frac{\theta_j}{\delta_j}\right)^2}, \tag{2.16}$$

where $d_0 := \operatorname{dist}((z_0, w_0), S_e(A, B))$ and θ_k, δ_k were defined in Lemma 2.1.

Proof. The assertions that $b_k \in B(x_k)$ and $a_k \in A(y_k)$ are direct consequences of step 1 in Algorithm 2. The definitions of γ_j and ϕ_j in step 2 of the method and (2.6), respectively, together with inequality (2.8), yield

$$\gamma_j = \frac{\phi_j(z_{j-1}, w_{j-1})}{\|\nabla \phi_j\|^2} \geq \frac{\theta_j}{\delta_j}, \qquad \text{for } j = 1, 2, \ldots. \tag{2.17}$$

Therefore,

$$\gamma_j^2 \geq \left(\frac{\theta_j}{\delta_j}\right)^2 \qquad \text{for } j = 1, 2, \ldots.$$

Multiplying both sides of the inequality above by $\rho_j(2-\rho_j)\|\nabla \phi_j\|^2$, adding from $j = 1$ to k, and using the first bound in (1.16), we have

$$\sum_{j=1}^{k} \rho_j(2-\rho_j)\left(\frac{\theta_j}{\delta_j}\right)^2 \|\nabla \phi_j\|^2 \leq d_0^2.$$

Taking i such that

$$i \in \arg\min_{j=1,\ldots,k} \left(\|\nabla \phi_j\|^2\right), \tag{2.18}$$

and using the previous inequality we obtain

$$\|\nabla \phi_i\|^2 \sum_{j=1}^{k} \rho_j(2-\rho_j)\left(\frac{\theta_j}{\delta_j}\right)^2 \leq d_0^2.$$

Bound (2.16) now follows from the above relation and noting that $\nabla \phi_i = (a_i + b_i, x_i - y_i)$. □

The theorem above estimates the quality of the best iterate among $(x_1, y_1), \ldots, (x_k, y_k)$ in terms of the stopping criterion (2.14). We refer to this estimate as *pointwise* complexity bounds.

Using the sequences of ergodic iterates associated to Algorithm 2, defined as in Subsection 1.2.2, we will derive different complexity bounds for Algorithm 2. Our study is inspired in the analysis developed in [27] to obtain complexity estimates for the HPE-method using average of the iterates, see [27, Section 4]. Following the notion of this reference, we will also call these kind of estimates as *ergodic* complexity bounds.

Define the sequences of ergodic means $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$ and $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{x,k})\}$, associated to the sequences $\{(x_k, b_k, 0)\}$, $\{(y_k, a_k, 0)\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ generated by Algorithm 2, as in (1.17), (1.18) and (1.19).

According to Lemma 1.4 we can attempt to bound the size of $\|\overline{a}_k + \overline{b}_k\|$, $\|\overline{x}_k - \overline{y}_k\|$, $\overline{\epsilon}_{x,k}$ and $\overline{\epsilon}_{y,k}$ in order to know when the ergodic iterates $\{\overline{x}_k\}$ and $\{\overline{y}_k\}$ will meet the stopping criterion (2.14).

Theorem 2.2. *Assume the hypotheses of Theorem 2.1. In addition, consider the sequence* $\{\Gamma_k\}$ *given in* (1.17)*, and consider also the sequences of ergodic iterates* $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$, $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{y,k})\}$ *associated to Algorithm 2 defined in* (1.17), (1.18) *and* (1.19). *Then, for every integer* $k \geq 1$*, we have*

$$\overline{b}_k \in B^{\overline{\epsilon}_{x,k}}(\overline{x}_k), \qquad \overline{a}_k \in A^{\overline{\epsilon}_{y,k}}(\overline{y}_k), \tag{2.19}$$

and

$$\|\overline{a}_k + \overline{b}_k\| \leq \frac{2d_0}{\Gamma_k}, \qquad \|\overline{x}_k - \overline{y}_k\| \leq \frac{2d_0}{\Gamma_k}, \qquad \overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k} \leq \frac{d_0^2(\varsigma_k + 4)}{\Gamma_k} \tag{2.20}$$

where

$$\varsigma_k = \max_{j=1,\ldots,k} \left\{ \frac{\mu_j}{\theta_j(2-\rho_j)\Gamma_k} \right\}. \tag{2.21}$$

Proof. Inclusions in (2.19) are consequence of Lemma 1.4. Since we are assuming that $\|\nabla\phi_k\| > 0$ for all $k = 1, \ldots$, Lemma 2.1 implies that $\phi_k(z_{k-1}, w_{k-1}) > 0$. Therefore, $\gamma_k > 0$ for all integer $k \geq 1$, and applying Theorem 1.3 we obtain the first two inequalities in (2.20).

Relation (2.9), together with identity $\phi_j(z_{j-1}, w_{j-1}) = \|\nabla\phi_j\|^2 \gamma_j$, yields

$$\frac{\mu_j}{\theta_j} \|\nabla\phi_j\|^2 \gamma_j \geq \|(y_j, b_j) - (z_{j-1}, w_{j-1})\|^2 \qquad \text{for } j = 1, 2, \ldots.$$

We multiply the inequality above by $\frac{1}{\Gamma_k}\rho_j\gamma_j$ and add from $j = 1$ to k, to obtain

$$\begin{aligned}
\frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j \|(y_j, b_j) - (z_{j-1}, w_{j-1})\|^2 &\leq \frac{1}{\Gamma_k}\sum_{j=1}^{k} \frac{\mu_j}{\theta_j}\rho_j\gamma_j^2 \|\nabla\phi_j\|^2 \\
&= \frac{1}{\Gamma_k}\sum_{j=1}^{k} \frac{\mu_j}{\theta_j(2-\rho_j)}\rho_j(2-\rho_j)\gamma_j^2 \|\nabla\phi_j\|^2 \\
&\leq \left(\max_{j=1,\ldots,k}\left\{\frac{\mu_j}{\theta_j(2-\rho_j)\Gamma_k}\right\}\right)\sum_{j=1}^{k}\rho_j(2-\rho_j)\gamma_j^2\|\nabla\phi_j\|^2 \\
&\leq \left(\max_{j=1,\ldots,k}\left\{\frac{\mu_j}{\theta_j(2-\rho_j)\Gamma_k}\right\}\right) d_0^2,
\end{aligned}$$

where the last inequality above follows from the first estimate in (1.16). Combining the relation above with (1.29) and the definition of ς_k in (2.21), we conclude the last bound in (2.20). $\square$

2.2 Specialized complexity results

In [14], global convergence of the sequences $\{(x_k, b_k)\}$, $\{(y_k, -a_k)\}$ and $\{(z_k, w_k)\}$, calculated by Algorithm 2, to a point $(z^*, w^*) \in S_e(A, B)$ was proved under the following assumptions:

A.1 there exist $\underline{\lambda}$ and $\overline{\lambda}$ such that, $\overline{\lambda} \geq \underline{\lambda} > 0$ and $\lambda_k, \mu_k \in [\underline{\lambda}, \overline{\lambda}]$ for all integer $k \geq 1$;

A.2 there exists $\overline{\rho} \in [0, 1)$ such that $\rho_k \in [1 - \overline{\rho}, 1 + \overline{\rho}]$ for all integer $k \geq 1$;

A.3 $\nu := \inf_k \left\{ \frac{\mu_k}{\lambda_k} - \left(\frac{\alpha_k}{2}\right)^2 \right\} > 0.$

Using the results obtained in the previous section we show that Algorithm 2, under hypotheses A.1-A.3, has $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate, while the rate in the ergodic sense is $\mathcal{O}(1/k)$.

Theorem 2.3. *Let $\{(z_k, w_k)\}$, $\{(x_k, b_k)\}$, $\{(y_k, a_k)\}$, $\{\lambda_k\}$, $\{\mu_k\}$, $\{\alpha_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 2 under assumptions A.1-A.3. If d_0 denote the distance of (z_0, w_0) to the extended solution set $S_e(A, B)$. Then, for all integer $k \geq 1$, we have*

$$b_k \in B(x_k), \qquad a_k \in A(y_k), \tag{2.22}$$

and there exists an index $1 \leq i \leq k$ such that

$$\|a_i + b_i\| \leq \frac{d_0 \upsilon}{\sqrt{k}(1 - \overline{\rho})}, \qquad \text{and} \qquad \|x_i - y_i\| \leq \frac{d_0 \upsilon}{\sqrt{k}(1 - \overline{\rho})},$$

where

$$\upsilon = \frac{2\overline{\lambda}\left(1 + \overline{\lambda}^2\right)\left(1 + \sqrt{\overline{\lambda}/\underline{\lambda}}\right)}{\underline{\lambda}^2 \nu}. \tag{2.23}$$

Proof. Inclusions in (2.22) follow from (2.15). Now we notice that condition A.2 implies that

$$\rho_j(2 - \rho_j) \geq (1 - \overline{\rho})^2, \qquad \text{for } \; j = 1, 2, \ldots. \tag{2.24}$$

Next, we observe that relation (2.1) in step 1 of Algorithm 2 yields $|\alpha_j| \leq 2\sqrt{\mu_j/\lambda_j}$. Hence, assumption A.1 implies

$$|\alpha_j| \leq 2\sqrt{\overline{\lambda}/\underline{\lambda}}, \qquad \text{for } \; j = 1, 2, \ldots.$$

The inequality above, together with the definition of δ_j in Lemma 2.1 and assumption A.1, yields

$$\delta_j \le 2\overline{\lambda}\left(1+\sqrt{\overline{\lambda}/\underline{\lambda}}\right). \tag{2.25}$$

Moreover, in [14, Proposition 3] under hypotheses A.1-A.3, it was proved that

$$\begin{aligned}\theta_j &:= \frac{1}{2}\left(1+\lambda_j\mu_j-\sqrt{(1+\lambda_j\mu_j)^2-4(\lambda_j\mu_j-(\lambda_j\alpha_j/2)^2)}\right)\\ &\ge \frac{\lambda_j^2(\mu_j/\lambda_j-(\alpha_j/2)^2)}{1+\lambda_j\mu_j}\\ &\ge \frac{\underline{\lambda}^2\nu}{1+\overline{\lambda}^2}.\end{aligned} \tag{2.26}$$

Thus, combining (2.25) and (2.26) we obtain

$$\frac{\theta_j}{\delta_j} \ge \frac{\underline{\lambda}^2\nu}{\left(1+\overline{\lambda}^2\right)2\overline{\lambda}\left(1+\sqrt{\overline{\lambda}/\underline{\lambda}}\right)} = \frac{1}{\upsilon} \qquad \text{for} \;\; j=1,2,\ldots. \tag{2.27}$$

Now, from (2.27) and (2.24) we deduce that

$$\rho_j(2-\rho_j)\left(\frac{\theta_j}{\delta_j}\right)^2 \ge \frac{(1-\overline{\rho})^2}{\upsilon^2},$$

for all $j=1,\ldots,k$. Hence, adding this last inequality from $j=1$ to k, we have

$$\sum_{j=1}^{k}\rho_j(2-\rho_j)\left(\frac{\theta_j}{\delta_j}\right)^2 \ge k\frac{(1-\overline{\rho})^2}{\upsilon^2}.$$

The theorem now follows from the above expression and inequality (2.16) in Theorem 2.1. □

Theorem 2.4. *Assume the hypotheses of Theorem 2.3 and consider the sequences of ergodic iterates* $\{(\overline{x}_k,\overline{b}_k,\overline{\epsilon}_{x,k})\}$, $\{(\overline{y}_k,\overline{a}_k,\overline{\epsilon}_{y,k})\}$ *associated to Algorithm 2 defined in (1.17), (1.18) and (1.19). Then, for every integer* $k\ge 1$, *we have*

$$\overline{b}_k \in B^{\overline{\epsilon}_{x,k}}(\overline{x}_k), \qquad \overline{a}_k \in A^{\overline{\epsilon}_{y,k}}(\overline{y}_k), \tag{2.28}$$

and

$$\|\overline{a}_k+\overline{b}_k\| \le \frac{2d_0\upsilon}{k(1-\overline{\rho})}, \qquad \|\overline{x}_k-\overline{y}_k\| \le \frac{2d_0\upsilon}{k(1-\overline{\rho})}, \tag{2.29}$$

$$\overline{\epsilon}_{x,k}+\overline{\epsilon}_{y,k} \le \frac{d_0^2\upsilon(\upsilon_k'+4)}{k(1-\overline{\rho})}, \tag{2.30}$$

where υ is given in (2.23) and

$$\upsilon_k' = \frac{\overline{\lambda}\left(1+\overline{\lambda}^2\right)\upsilon}{\underline{\lambda}^2\nu(1-\overline{\rho})^2 k}.$$

Proof. The inclusions in (2.28) follow from Lemma 1.4. The definition of Γ_k, together with assumption A.2 and (2.17), yields

$$\Gamma_k \geq (1-\overline{\rho})\sum_{j=1}^{k}\frac{\theta_j}{\delta_j}.$$

Now, by (2.27) and the inequality above we have

$$\Gamma_k \geq \frac{(1-\overline{\rho})k}{\upsilon}. \tag{2.31}$$

The bounds in (2.29) follow from (2.31) and the first two inequalities in (2.20). To conclude the proof we observe that the definition of ς_k in (2.21), hypothesis A.1, (2.26) and (2.31) imply

$$\varsigma_k \leq \frac{\overline{\lambda}(1+\overline{\lambda}^2)\upsilon}{\underline{\lambda}^2\nu(1-\overline{\rho})^2 k}.$$

Thus, combining the above relation, the last inequality in (2.20), the definition of υ_k' and (2.31), we obtain inequality (2.30). □

2.3 Spingarn's splitting method

In [39] Spingarn introduced a method to find a zero of the sum of m maximal monotone operators based on the concept of *partial inverses*. Although here we restrict our study to the $m=2$ case, we make some remarks on the general case. Spingarn's method computes independent proximal subproblems on each of the m operators involved in the problem at each iteration, and then finds the next iterate essentially averaging the results. This algorithm is actually a special case of Douglas-Rachford splitting method [13], and it was also proved that it is a particular instance of the general projective-splitting methods for sums of m maximal monotone operators, which was introduced in [15].

For the two-operator case, Eckstein and Svaiter proved in [14] that Spingarn's method is a special case of a scaled variant of Algorithm 2. Interpreting Spingarn's algorithm as an instance of Algorithm 2 allows us to use the analysis developed in the previous sections for obtaining its complexity bounds.

For this purpose, let us begin with a brief discussion of the reformulation of problem (1.1) studied in [14], obtained by *rescalation*.

Let $\eta > 0$ be a fix scalar, multiplying both sides of (1.1) by η, gives the reformulated problem

$$0 \in \eta A(z) + \eta B(z). \tag{2.32}$$

Observe that the solution set of (1.1) and (2.32) remains the same, but the extended solution set associated to operators ηA and ηB is transformed into

$$S_e(\eta A, \eta B) = \{(z, \eta w) : (z, w) \in S_e(A, B)\}.$$

If we apply Algorithm 2 to ηA and ηB, and consider ηa_k, ηb_k and ηw_k instead of a_k, b_k and w_k, respectively, after some algebraic manipulations, we obtain a scheme identical to Algorithm 2, except that (2.2)-(2.5) are modified to

$$\lambda_k \eta b_k + x_k = z_{k-1} + \lambda_k \eta w_{k-1}, \qquad b_k \in B(x_k); \tag{2.33}$$

$$\mu_k \eta a_k + y_k = (1-\alpha_k) z_{k-1} + \alpha_k x_k - \mu_k \eta w_{k-1}, \qquad a_k \in A(y_k); \tag{2.34}$$

$$\gamma_k = \frac{\langle z_{k-1} - x_k, b_k - w_{k-1}\rangle + \langle z_{k-1} - y_k, a_k + w_{k-1}\rangle}{\eta \|a_k + b_k\|^2 + \frac{1}{\eta}\|x_k - y_k\|^2}, \tag{2.35}$$

$$z_k = z_{k-1} - \rho_k \gamma_k \eta (a_k + b_k), \tag{2.36}$$

$$w_k = w_{k-1} - \frac{\rho_k \gamma_k}{\eta}(x_k - y_k). \tag{2.37}$$

The general pointwise and ergodic complexity bounds for the method above are obtained as a direct consequence of Theorems 2.1 and 2.2, replacing a_i, b_i, $\overline{a}_k$, $\overline{b}_k$, $\overline{\epsilon}_{x,k}$ and $\overline{\epsilon}_{y,k}$ by ηa_i, ηb_i, $\eta \overline{a}_k$, $\eta \overline{b}_k$, $\eta \overline{\epsilon}_{x,k}$ and $\eta \overline{\epsilon}_{y,k}$, respectively.

If $\eta > 0$, in our notation, Spingarn's splitting method is reduced to the following set of recursions:

$$\eta b_k + x_k = z_{k-1} + \eta w_{k-1}, \qquad b_k \in B(x_k); \tag{2.38}$$

$$\eta a_k + y_k = z_{k-1} - \eta w_{k-1}, \qquad a_k \in A(y_k); \tag{2.39}$$

$$z_k = (1-\rho_k) z_{k-1} + \frac{\rho_k}{2}(x_k + y_k), \tag{2.40}$$

$$w_k = (1-\rho_k) w_{k-1} + \frac{\rho_k}{2}(b_k - a_k). \tag{2.41}$$

Notice that if we take $\lambda_k = \mu_k = 1$ and $\alpha_k = 0$ for all integer $k \geq 1$, then (2.33)-(2.34) and (2.38)-(2.39) are identical. Moreover, the remaining calculations (2.35), (2.36) and (2.37) can be rewritten into the form (2.40)-(2.41).

Theorem 2.5. *Fix $\lambda_k = \mu_k = 1$ and $\alpha_k = 0$ in (2.33)-(2.37) for every integer $k \geq 1$. Then the recursions (2.33)-(2.37) and (2.38)-(2.41) are identical. Hence, Spingarn's method is a special case of Algorithm 2.*

Proof. This result was proved in [14, Subsection 4.2]. □

The following theorem establishes the global convergence rate for Spingarn's method in terms of the termination criterion (2.14).

Theorem 2.6. *Let $\eta > 0$ and let $\{(x_k, b_k)\}$, $\{(y_k, a_k)\}$ and $\{\rho_k\}$ be the sequences generated by Spingarn's splitting method (2.38)-(2.41). For every $k \geq 1$, define*

$$P_k = \sum_{j=1}^{k} \rho_j, \tag{2.42}$$

and

$$\overline{x}_k = \frac{1}{P_k}\sum_{j=1}^{k} \rho_j x_j, \qquad \overline{b}_k = \frac{1}{P_k}\sum_{j=1}^{k} \rho_j b_j, \qquad \overline{\epsilon}_{x,k} = \frac{1}{P_k}\sum_{j=1}^{k} \rho_j \left\langle x_j - \overline{x}_k, b_j \right\rangle, \tag{2.43}$$

$$\overline{y}_k = \frac{1}{P_k}\sum_{j=1}^{k} \rho_j y_j, \qquad \overline{a}_k = \frac{1}{P_k}\sum_{j=1}^{k} \rho_j a_j, \qquad \overline{\epsilon}_{y,k} = \frac{1}{P_k}\sum_{j=1}^{k} \rho_j \left\langle y_j - \overline{y}_k, a_j \right\rangle. \tag{2.44}$$

Assume that hypothesis A.2 holds and set $d_0 := \operatorname{dist}\left((z_0, \eta w_0), S_e\left(\eta A, \eta B\right)\right)$. Then the following statements hold.

(a) For every integer $k \geq 1$ we have

$$b_k \in B(x_k), \qquad a_k \in A(y_k), \tag{2.45}$$

and there exists an index $1 \leq i \leq k$ such that

$$\|a_i + b_i\| \leq \frac{2d_0}{\eta\sqrt{k}(1-\overline{\rho})}, \qquad \|x_i - y_i\| \leq \frac{2d_0}{\sqrt{k}(1-\overline{\rho})}.$$

(b) For every integer $k \geq 1$ we have

$$\overline{b}_k \in B^{\overline{\epsilon}_{x,k}}(\overline{x}_k), \qquad \overline{a}_k \in A^{\overline{\epsilon}_{y,k}}(\overline{y}_k),$$

and

$$\left\|\overline{a}_k + \overline{b}_k\right\| \leq \frac{4d_0}{\eta k(1-\overline{\rho})}, \qquad \|\overline{x}_k - \overline{y}_k\| \leq \frac{4d_0}{k(1-\overline{\rho})}, \tag{2.46}$$

$$\overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k} \leq \frac{2d_0^2}{\eta k(1-\overline{\rho})}\left(\frac{2}{(1-\overline{\rho})^2} + 4\right). \tag{2.47}$$

Proof. (a) If we fix $\lambda_k = \mu_k = 1$ and $\alpha_k = 0$ for all integer $k \geq 1$, in the set of recursions (2.33)-(2.37), then the definitions of δ_k and θ_k in the statement of Lemma 2.1 yield $\delta_k = 2$ and $\theta_k = 1$ for every integer $k \geq 1$.

Since by Theorem 2.5, Spingarn's algorithm is a special case of (2.33)-(2.37) with this choice of the parameters λ_k, μ_k and α_k, the claims in (a) follow from Theorem 2.1 applied to (2.33)-(2.37) with $\lambda_k = \mu_k = 1$, $\alpha_k = 0$, and assumption A.2.

(b) The first assertion in (b) follows from the definitions of Γ_k, $\overline{x}_k$, $\overline{b}_k$, $\overline{\epsilon}_{x,k}$, $\overline{y}_k$, $\overline{a}_k$, $\overline{\epsilon}_{y,k}$ in (2.42), (2.43) and (2.44), the inclusions in (2.38), (2.39) and Theorem 1.1.

Now we set, $\lambda_k = \mu_k = 1$ and $\alpha_k = 0$ in (2.33)-(2.37) for all $k = 1, 2, \dots$.

Claim: $\gamma_k = \frac{1}{2}$ for every integer $k \geq 1$.

Equalities in (2.33) and (2.34) are reduced to

$$\eta b_k + x_k = z_{k-1} + \eta w_{k-1},$$
$$\eta a_k + y_k = z_{k-1} - \eta w_{k-1};$$

hence, $z_{k-1} - x_k = \eta(b_k - w_{k-1})$ and $z_{k-1} - y_k = \eta(a_k + w_{k-1})$, which implies

$$\langle z_{k-1} - x_k, b_k - w_{k-1}\rangle + \langle z_{k-1} - y_k, a_k + w_{k-1}\rangle = \eta \|b_k - w_{k-1}\|^2 + \eta \|a_k + w_{k-1}\|^2 . \quad (2.48)$$

We observe that

$$\begin{aligned}\frac{1}{\eta}\|x_k - y_k\|^2 &= \frac{1}{\eta}\|x_k - z_{k-1}\|^2 + 2\frac{1}{\eta}\langle x_k - z_{k-1}, z_{k-1} - y_k\rangle + \frac{1}{\eta}\|z_{k-1} - y_k\|^2 \\ &= \eta\|b_k - w_{k-1}\|^2 + 2\eta\langle w_{k-1} - b_k, a_k + w_{k-1}\rangle + \eta\|a_k + w_{k-1}\|^2 ,\end{aligned}$$

and

$$\eta\|a_k + b_k\|^2 = \eta\|a_k + w_{k-1}\|^2 + 2\eta\langle a_k + w_{k-1}, b_k - w_{k-1}\rangle + \eta\|b_k - w_{k-1}\|^2 .$$

Hence, adding these two last identities we obtain

$$\eta\|a_k + b_k\|^2 + \frac{1}{\eta}\|x_k - y_k\|^2 = 2\eta\|a_k + w_{k-1}\|^2 + 2\eta\|b_k - w_{k-1}\|^2 . \quad (2.49)$$

Combining (2.35), (2.48) and (2.49) we prove the claim.

Theorem 2.5 implies that the sequences $\{(x_k, \eta b_k)\}$ and $\{(y_k, \eta a_k)\}$ can be viewed as generated by Algorithm 2 applied to the operators ηA and ηB, with $\lambda_k = \mu_k = 1$ and $\alpha_k = 0$ for all $k = 1, 2, \dots$.

Moreover, since $\gamma_k = 1/2$ for all integer $k \geq 1$, the sequences of ergodic iterates associated to $\{(x_k, \eta b_k)\}$, $\{(y_k, \eta a_k)\}$, $\{\rho_k\}$ and $\{\gamma_k\}$, which are obtained by equations (1.17), (1.18) and (1.19) with $\Gamma_k = (1/2)P_k$, are exactly as defined in (2.43) and (2.44), but with $\eta\overline{b}_k$, $\eta\overline{\epsilon}_{x,k}$, $\eta\overline{a}_k$ and $\eta\overline{\epsilon}_{y,k}$ instead of $\overline{b}_k$, $\overline{\epsilon}_{x,k}$, $\overline{a}_k$ and $\overline{\epsilon}_{y,k}$.

Hence, applying Theorem 2.2 we have

$$\left\|\eta(\overline{a}_k+\overline{b}_k)\right\| \le \frac{2d_0}{(1/2)P_k}, \quad \|\overline{x}_k-\overline{y}_k\| \le \frac{2d_0}{(1/2)P_k}, \quad \eta(\overline{\epsilon}_{x,k}+\overline{\epsilon}_{y,k}) \le \frac{d_0^2(\varsigma_k+4)}{(1/2)P_k}, \tag{2.50}$$

where d_0 is the distance of $(z_0, \eta w_0)$ to $S_e(\eta A, \eta B)$ and $\varsigma_k = \max_{j=1,\ldots,k}\left\{\frac{\mu_j}{\theta_j(2-\rho_j)(1/2)P_k}\right\}$.

We notice that condition A.2 yields that $\rho_j \ge 1-\overline{\rho}$ for every j, therefore by the definition of P_k we have

$$P_k \ge k(1-\overline{\rho}). \tag{2.51}$$

Furthermore, since in this case $\mu_j = 1$, $\theta_j = 1$ and $2-\rho_j \ge 1-\overline{\rho}$ for all integer $j \ge 1$, these latter relations together with the definition of ς_k and (2.51) imply

$$\varsigma_k \le \frac{2}{(1-\overline{\rho})^2 k} \le \frac{2}{(1-\overline{\rho})^2}, \qquad \text{for all} \quad k = 1, 2, \ldots. \tag{2.52}$$

Thus, (b) follows combining the inequality above, (2.51), condition A.2 and the bounds in (2.50). □

2.4 A parallel inexact case

Algorithm 2 has to solve two proximal subproblems on each iteration to construct decomposable separators. Since finding the exact solution of subproblems (2.2) and (2.3) could be a challenging task, one might wish to allow approximate evaluations of the resolvents mapping, while keeping convergence of the method.

It is customary to appeal to the theory of approximation criteria for the PPA and related methods, when attempting to approximate solutions of proximal subproblems. The first inexact versions of the PPA were introduced in [34] by Rockafellar and are based on summable absolute error criteria. For instance, one of the approximation criterion proposed in [34] for the PPA is

$$\left\|z_{k+1}-(I+\lambda_k T)^{-1}(z_k)\right\| \le s_k, \qquad \sum_{k=1}^{\infty} s_k < \infty. \tag{2.53}$$

This kind of approximation criteria, which involve a theoretical sequence $\{s_k\} \subset [0,\infty)$ such that $\sum_{k=1}^{\infty} s_k < \infty$, has as a practical disadvantage the fact that there is no constructive way of choosing it. Therefore, it is of great importance to develop error conditions for approximating proximal subproblems that could be computable during the progress of the algorithm. Relative error criteria of this kind were proposed in [37], [36] and [38].

In this section we will include a relative error criterion to evaluate approximately the resolvent mappings in Algorithm 2, in the special case of taking $\alpha_k = 0$ for all k, which possibly allows the subproblems to be performed in parallel. To solve inexactly subproblems

(2.2) and (2.3) we will use the notion of approximate solution for a proximal subproblem introduced in [38] by Solodov and Svaiter.

The general projective-splitting framework for the sum of $m \geq 2$ maximal monotone operators [15] admits a relative error condition to solve approximately the proximal subproblems. The criterion used in [15] is a generalization of the relative error criterion for the HPE-method in the case of m maximal monotone operators

We have preferred the framework developed in [38], since it yields a larger relative error condition and evaluation of the ϵ-enlargements of the operators.

We now present the notion of inexact solution of a proximal subproblem introduced in [38]. Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone operator, $\lambda > 0$ and $z \in \mathbb{R}^n$. Consider the *proximal system*

$$\begin{cases} w \in T(z'), \\ \lambda w + z' - z = 0. \end{cases} \tag{2.54}$$

Definition 3. *Given $\sigma \in [0, 1)$, a triplet $(z', w, \epsilon) \in \mathbb{E}$ is called a σ-*approximate solution* of (2.54) at (λ, z) if*

$$\begin{aligned} & w \in T^{\epsilon}(z'), \\ & \left\|\lambda w + z' - z\right\|^2 + 2\lambda\epsilon \leq \sigma \left(\|\lambda w\|^2 + \left\|z' - z\right\|^2\right). \end{aligned} \tag{2.55}$$

We observe that if (z', w) is the exact solution of (2.54), then taking $\epsilon = 0$ the triplet (z', w, ϵ) satisfies the approximation criterion (2.55) for all $\sigma \in [0, 1)$. Conversely, if $\sigma = 0$ only the exact solution of (2.54), taking $\epsilon = 0$, will satisfy (2.55).

The method that will be studied in this section is as follows.

Algorithm 3. *Start with $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$, $\sigma \in [0, 1)$, $\overline{\rho} \in [0, 1)$. Then for $k = 1, 2, \ldots$*

1. *Choose $\lambda_k, \mu_k > 0$ and calculate $(x_k, b_k, \epsilon_{x,k})$ and $(y_k, a_k, \epsilon_{y,k}) \in \mathbb{E}$ such that*

$$b_k \in B^{\epsilon_{x,k}}(x_k), \qquad a_k \in A^{\epsilon_{y,k}}(y_k),$$

and

$$\|\lambda_k(b_k - w_{k-1}) + x_k - z_{k-1}\|^2 + 2\lambda_k\epsilon_{x,k} \leq \sigma \left(\|x_k - z_{k-1}\|^2 + \|\lambda_k(b_k - w_{k-1})\|^2\right), \tag{2.56}$$

$$\|\mu_k(a_k + w_{k-1}) + y_k - z_{k-1}\|^2 + 2\mu_k\epsilon_{y,k} \leq \sigma \left(\|y_k - z_{k-1}\|^2 + \|\mu_k(a_k + w_{k-1})\|^2\right). \tag{2.57}$$

2. If $\|a_k + b_k\| + \|x_k - y_k\| = 0$ stop. Otherwise, set

$$\gamma_k = \frac{\langle z_{k-1} - x_k, b_k - w_{k-1}\rangle + \langle z_{k-1} - y_k, a_k + w_{k-1}\rangle - \epsilon_{x,k} - \epsilon_{y,k}}{\|a_k + b_k\|^2 + \|x_k - y_k\|^2}.$$

3. Choose a parameter $\rho_k \in [1 - \overline{\rho}, 1 + \overline{\rho}]$ and set

$$z_k = z_{k-1} - \rho_k \gamma_k (a_k + b_k),$$
$$w_k = w_{k-1} - \rho_k \gamma_k (x_k - y_k).$$

Notice that for all iteration $k = 1, 2, \ldots$, the triplet $(x_k, b_k, \epsilon_{x,k})$, calculated in step 1 of Algorithm 3, is a σ-approximate solution of (2.54) at (λ_k, z_k), where $T = B - w_{k-1}$. Similarly, $(y_k, a_k, \epsilon_{y,k})$ is a σ-approximate solution of (2.54) at point (μ_k, z_k), with $T = A + w_{k-1}$. Observe also that taking $\sigma = 0$ in Algorithm 3 yields exactly Algorithm 2 with $\alpha_k = 0$ for all integer $k \geq 1$, since condition (2.1) is met.

Let us define, for every integer $k \geq 1$, the decomposable separator ϕ_k associated to the pair $(x_k, b_k, \epsilon_{x,k})$ and $(y_k, a_k, \epsilon_{y,k})$, $\phi_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\phi_k(z, w) = \langle z - x_k, b_k - w\rangle + \langle z - y_k, a_k + w\rangle - \epsilon_{x,k} - \epsilon_{y,k}. \tag{2.58}$$

It will be shown, in the following lemma, that if Algorithm 3 stops at iteration k in step 2 then it has found a point in $S_e(A, B)$. Otherwise we have that $\|\nabla \phi_k\| > 0$, which gives $\phi_k(z_{k-1}, w_{k-1}) > 0$. This clearly implies that Algorithm 3 falls within the general framework of Algorithm 1.

Lemma 2.2. *If $\{(x_k, b_k, \epsilon_{x,k})\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{(z_k, w_k)\}$, $\{\lambda_k\}$, $\{\mu_k\}$, and $\{\rho_k\}$ are the sequences generated by Algorithm 3, and $\{\phi_k\}$ is the sequence of decomposable separators defined in* (2.58)*. Then, for every integer $k \geq 1$, we have*

$$\phi_k(z_{k-1}, w_{k-1}) \geq \frac{1 - \sigma}{4} \xi_k \left(\|a_k + b_k\|^2 + \|x_k - y_k\|^2\right) \geq 0, \tag{2.59}$$

where

$$\xi_k = \min\left\{\lambda_k, \frac{1}{\lambda_k}, \mu_k, \frac{1}{\mu_k}\right\}. \tag{2.60}$$

Furthermore, if $\|\nabla \phi_k\| > 0$ then $\phi_k(z_{k-1}, w_{k-1}) > 0$, and $\|\nabla \phi_k\| = 0$ if and only if $(x_k, b_k) = (y_k, -a_k) \in S_e(A, B)$.

Proof. We define the residual vectors

$$r_{x,k} = \lambda_k (b_k - w_{k-1}) + x_k - z_{k-1} \quad \text{and} \quad r_{y,k} = \mu_k (a_k + w_{k-1}) + y_k - z_{k-1}.$$

By the definition of ϕ_k in (2.58) and direct calculations, we have

$$\begin{aligned}\phi_k(z_{k-1}, w_{k-1}) = &\ \langle z_{k-1} - x_k, b_k - w_{k-1}\rangle + \langle z_{k-1} - y_k, a_k + w_{k-1}\rangle - \epsilon_{x,k} - \epsilon_{y,k} \\ = &\ \frac{1}{2\lambda_k}\left(\|z_{k-1} - x_k\|^2 + \|\lambda_k(b_k - w_{k-1})\|^2 - \|r_{x,k}\|^2 - 2\lambda_k\epsilon_{x,k}\right) \\ &+ \frac{1}{2\mu_k}\left(\|z_{k-1} - y_k\|^2 + \|\mu_k(a_k + w_{k-1})\|^2 - \|r_{y,k}\|^2 - 2\mu_k\epsilon_{y,k}\right). \end{aligned} \tag{2.61}$$

Identity (2.61), together with the error criteria (2.56) and (2.57), implies that

$$\begin{aligned}\phi_k(z_{k-1}, w_{k-1}) \geq &\ \frac{1-\sigma}{2\lambda_k}\left(\|z_{k-1} - x_k\|^2 + \|\lambda_k(b_k - w_{k-1})\|^2\right) \\ &+ \frac{1-\sigma}{2\mu_k}\left(\|z_{k-1} - y_k\|^2 + \|\mu_k(a_k + w_{k-1})\|^2\right). \end{aligned} \tag{2.62}$$

If we interpret the last expression as a quadratic form applied to the $\mathbb{R}^4$ vector $(\|z_{k-1} - x_k\|, \|b_k - w_{k-1}\|, \|z_{k-1} - y_k\|, \|a_k + w_{k-1}\|)^T$, we obtain

$$\begin{aligned}\phi_k(z_{k-1}, w_{k-1}) \geq &\ \frac{1-\sigma}{2}\begin{pmatrix}\|z_{k-1} - x_k\| \\ \|w_{k-1} - b_k\| \\ \|z_{k-1} - y_k\| \\ \|w_{k-1} + a_k\|\end{pmatrix}^T \begin{pmatrix}\frac{1}{\lambda_k} & 0 & 0 & 0 \\ 0 & \lambda_k & 0 & 0 \\ 0 & 0 & \frac{1}{\mu_k} & 0 \\ 0 & 0 & 0 & \mu_k\end{pmatrix}\begin{pmatrix}\|z_{k-1} - x_k\| \\ \|w_{k-1} - b_k\| \\ \|z_{k-1} - y_k\| \\ \|w_{k-1} + a_k\|\end{pmatrix} \\ \geq &\ \frac{1-\sigma}{2}\xi_k\left(\|z_{k-1} - x_k\|^2 + \|b_k - w_{k-1}\|^2 + \|z_{k-1} - y_k\|^2 + \|a_k + w_{k-1}\|^2\right) \end{aligned} \tag{2.63}$$

where ξ_k, defined in (2.60), is the smallest eigenvalue of the matrix in (2.63).

Combining the second inequality in (2.63) with relations

$$\|z_{k-1} - x_k\|^2 + \|z_{k-1} - y_k\|^2 \geq \frac{1}{2}\|x_k - y_k\|^2,$$

$$\|b_k - w_{k-1}\|^2 + \|a_k + w_{k-1}\|^2 \geq \frac{1}{2}\|a_k + b_k\|^2;$$

we obtain (2.59).

Since $\xi_k > 0$ and $\sigma \in [0, 1)$, inequality (2.59) clearly implies that $\phi_k(z_{k-1}, w_{k-1}) > 0$ whenever $\|\nabla\phi_k\| > 0$. To prove the last assertion of the lemma we rewrite $\phi_k(z_{k-1}, w_{k-1})$ as

$$\phi_k(z_{k-1}, w_{k-1}) = \langle a_k + b_k, z_{k-1} - y_k\rangle + \langle x_k - y_k, w_{k-1} - b_k\rangle - \epsilon_{x,k} - \epsilon_{y,k},$$

then, if $\|\nabla\phi_k\| = 0$ it follows that $x_k = y_k$, $b_k = -a_k$ and

$$\phi_k(z_{k-1}, w_{k-1}) = -\epsilon_{x,k} - \epsilon_{y,k}.$$

From (2.59), the equation above and the fact that $\epsilon_{x,k}, \epsilon_{y,k} \geq 0$; we obtain $\epsilon_{x,k} = \epsilon_{y,k} = 0$. Hence, $b_k \in B(x_k)$, $a_k \in A(y_k)$ and we conclude that $(x_k, b_k) = (y_k, -a_k) \in S_e(A, B)$. □

For deriving complexity bounds for Algorithm 3 we will assume, as was done in the preceding section, that the stopping condition in step 2 of Algorithm 3 does not hold for all iteration k. Thus, from now on we suppose that $\|\nabla\phi_k\| > 0$ for every integer $k \geq 1$.

Theorem 2.7. *Take $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$ and let $\{(x_k, b_k, \epsilon_{x,k})\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\lambda_k\}$, $\{\mu_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 3. Let d_0 be the distance of (z_0, w_0) to the set $S_e(A, B)$ and for all integer $k \geq 1$ define ξ_k by (2.60). Then, for every integer $k \geq 1$ we have*

$$b_k \in B^{\epsilon_{x,k}}(x_k), \qquad a_k \in A^{\epsilon_{y,k}}(y_k), \tag{2.64}$$

and there exists an index $1 \leq i \leq k$ such that

$$\|a_i + b_i\|^2 + \|x_i - y_i\|^2 \leq \frac{16 d_0^2}{(1-\sigma)^2(1-\overline{\rho})^2 \xi_i \sum\limits_{j=1}^{k} \xi_j}, \tag{2.65}$$

$$\epsilon_{x,i} + \epsilon_{y,i} \leq \frac{4\sigma d_0^2}{(1-\sigma)^2(1-\overline{\rho})^2 \sum\limits_{j=1}^{k} \xi_j}. \tag{2.66}$$

Proof. The inclusions in (2.64) are due to step 1 of Algorithm 3. Since $\gamma_k = \dfrac{\phi_k(z_{k-1}, w_{k-1})}{\|\nabla\phi_k\|^2}$, using (2.59) we have

$$\gamma_k \geq \frac{1-\sigma}{4}\xi_k, \qquad \text{for } k = 1, 2, \ldots. \tag{2.67}$$

Thus, squaring both sides of (2.67) and multiplying by $\|\nabla\phi_k\|^2$ we obtain

$$\gamma_k^2 \|\nabla\phi_k\|^2 \geq \left(\frac{1-\sigma}{4}\right)^2 \xi_k^2 \|\nabla\phi_k\|^2. \tag{2.68}$$

We observe that the error criteria (2.56) and (2.57) imply

$$\epsilon_{x,k} \leq \frac{\sigma}{2\lambda_k}\left(\|z_{k-1} - x_k\|^2 + \|\lambda_k(b_k - w_{k-1})\|^2\right)$$

and

$$\epsilon_{y,k} \leq \frac{\sigma}{2\mu_k}\left(\|z_{k-1} - y_k\|^2 + \|\mu_k(a_k + w_{k-1})\|^2\right).$$

Adding these two inequalities and considering relation (2.62) we obtain

$$\epsilon_{x,k} + \epsilon_{y,k} \leq \frac{\sigma}{1-\sigma}\phi_k(z_{k-1}, w_{k-1}) = \frac{\sigma}{1-\sigma}\gamma_k \|\nabla\phi_k\|^2.$$

Multiplying the latter inequality by γ_k, using (2.67) and multiplying both sides of the resulting expression by $\dfrac{1-\sigma}{\sigma}$ we have

$$\frac{(1-\sigma)^2}{4\sigma}\xi_k(\epsilon_{x,k}+\epsilon_{y,k}) \leq \gamma_k^2 \|\nabla\phi_k\|^2 \qquad \text{for } k=1,2,\dots. \tag{2.69}$$

Now we define

$$\psi_k := \max\left\{\left(\frac{1-\sigma}{4}\right)^2 \xi_k \|\nabla\phi_k\|^2, \frac{(1-\sigma)^2}{4\sigma}(\epsilon_{x,k}+\epsilon_{y,k})\right\},$$

and we combine (2.68) and (2.69) to obtain

$$\xi_k\psi_k \leq \gamma_k^2\|\nabla\phi_k\|^2 \qquad \text{for } k=1,2,\dots.$$

Next, adding the inequality above from $j=1$ to k, using the first inequality in (1.16), and the fact that $\rho_k \in [1-\overline{\rho}, 1+\overline{\rho}]$ for all integer $k \geq 1$, we obtain

$$\sum_{j=1}^{k}\xi_j\psi_j \leq \frac{d_0^2}{(1-\overline{\rho})^2}, \tag{2.70}$$

and consequently

$$\left(\min_{j=1,\dots,k}\{\psi_j\}\right)\sum_{j=1}^{k}\xi_j \leq \frac{d_0^2}{(1-\overline{\rho})^2}.$$

The theorem now follows from this last inequality and the definition of ψ_k. □

If $\{(x_k, b_k, \epsilon_{x,k})\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ are the sequences generated by Algorithm 3, we define the sequences of ergodic iterates $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$, $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{y,k})\}$ as in (1.17), (1.18) and (1.19). Since Algorithm 3 is a special instance of Algorithm 1 the results of Subsection 1.2.2 hold for the ergodic sequences associated to Algorithm 3. Thus, combining Theorem 1.3 and Lemma 2.2 we can state ergodic complexity estimates for the method.

Theorem 2.8. *Let $\{(x_k, b_k, \epsilon_{x,k})\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 3 and let $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$ and $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{y,k})\}$ be the sequences of ergodic iterates associated to Algorithm 3 defined in (1.17), (1.18) and (1.19). Consider ξ_k given by* (2.60). *Then, for all integer $k \geq 1$, we have*

$$\overline{b}_k \in B^{\overline{\epsilon}_{x,k}}(\overline{x}_k), \qquad \overline{a}_k \in A^{\overline{\epsilon}_{y,k}}(\overline{y}_k) \tag{2.71}$$

and

$$\|\bar{a}_k + \bar{b}_k\| \leq \frac{2d_0}{\Gamma_k}, \qquad \|\bar{x}_k - \bar{y}_k\| \leq \frac{2d_0}{\Gamma_k}, \qquad \bar{\epsilon}_{x,k} + \bar{\epsilon}_{y,k} \leq \frac{d_0^2(\varphi_k + 4)}{\Gamma_k}, \tag{2.72}$$

where d_0 *is the distance of* (z_0, w_0) *to* $S_e(A, B)$ *and*

$$\varphi_k = \left(\frac{2}{1-\sigma}\right) \max_{j=1,\ldots,k} \left\{\frac{1}{\xi_j(2-\rho_j)\Gamma_k}\right\}.$$

Proof. The inclusions in (2.71) follow from Lemma 1.4. The first two bounds in (2.72) are due to (1.28) in Theorem 1.3.

We notice now that the second inequality in (2.63) implies that

$$\phi_j(z_{j-1}, w_{j-1}) \geq \frac{1-\sigma}{2}\xi_j \left(\|z_{j-1} - y_j\|^2 + \|b_j - w_{j-1}\|^2\right), \qquad \text{for } j = 1, \ldots.$$

The relation above, together with the definition of γ_j, yields

$$\frac{2}{(1-\sigma)\xi_j}\gamma_j \|\nabla\phi_j\|^2 \geq \|z_{j-1} - y_j\|^2 + \|b_j - w_{j-1}\|^2, \qquad \text{for } j = 1, \ldots.$$

Multiplying the above inequality by $\dfrac{1}{\Gamma_k}\rho_j\gamma_j$ and adding from $j = 1$ to k, we obtain

$$\begin{aligned} \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j\gamma_j \|(y_j, b_j) - (z_{j-1}, w_{j-1})\|^2 &\leq \frac{1}{\Gamma_k}\sum_{j=1}^{k} \frac{2}{(1-\sigma)\xi_j}\rho_j\gamma_j^2 \|\nabla\phi_j\|^2 \\ &= \frac{1}{\Gamma_k}\sum_{j=1}^{k} \frac{2}{(1-\sigma)\xi_j(2-\rho_j)}\rho_j(2-\rho_j)\gamma_j^2 \|\nabla\phi_j\|^2 \\ &\leq \varphi_k \sum_{j=1}^{k} \rho_j(2-\rho_j)\gamma_j^2 \|\nabla\phi_j\|^2 \\ &\leq \varphi_k d_0^2; \end{aligned} \tag{2.73}$$

where the second and the third inequalities are due to the definition of φ_k and the first bound in (1.16), respectively. Replacing (2.73) into (1.29), in Theorem 1.3, we obtain the last bound in (2.72). □

Theorems 2.7 and 2.8 provide general complexity results for Algorithm 3. Observe that the derived bounds are expressed in terms of ξ_k and Γ_k. The next result presents iteration-complexity bounds for Algorithm 3 to obtain (δ, ϵ)-approximate solutions of problem (1.1).

Theorem 2.9. *Assume the hypotheses of Theorem 2.8. Suppose also that there exist* $\overline{\lambda}$ *and* $\underline{\lambda}$ *such that,* $\overline{\lambda} \geq \underline{\lambda} > 0$ *and for all integer* $k \geq 1$ *the proximal parameters* λ_k *and* μ_k *in step 1* *of Algorithm 3 are chosen within the interval* $[\underline{\lambda}, \overline{\lambda}]$*. Defining* $\xi = \min\left\{\underline{\lambda}, \frac{1}{\overline{\lambda}}\right\}$*, then, for all* δ*,* $\epsilon > 0$*, the following statements hold.*

(a) There exists an index

$$i = \mathcal{O}\left(\max\left\{\frac{d_0^2}{\xi^2\delta^2}, \frac{d_0^2}{\xi\epsilon}\right\}\right)$$

such that, the iterate (x_i, y_i) *is a* (δ, ϵ)*-solution of problem* (1.1).

(b) There exists an index

$$k_0 = \mathcal{O}\left(\max\left\{\frac{d_0}{\xi\delta}, \frac{d_0^2}{\xi\epsilon}\right\}\right)$$

such that, for all integer $k \geq k_0$ *the ergodic iterate* $(\overline{x}_k, \overline{y}_k)$ *is a* (δ, ϵ)*-solution of problem* (1.1).

Proof. We first notice that

$$\xi_j \geq \xi \qquad \text{for all } \ j = 1, 2, \ldots. \tag{2.74}$$

Thus, statement (a) is a direct consequence of (2.74) and Theorem 2.7.

Now, we combine the definition of Γ_k in (1.17) with (2.67) and (2.74), to obtain

$$\Gamma_k \geq \frac{(1-\overline{\rho})(1-\sigma)}{4}\sum_{j=1}^{k}\xi_j \geq \frac{(1-\overline{\rho})(1-\sigma)}{4}\xi k. \tag{2.75}$$

Therefore, inequalities (2.74) and (2.75) imply that

$$\frac{1}{\xi_j(2-\rho_j)\Gamma_k} \leq \frac{4}{\xi^2(1-\overline{\rho})^2(1-\sigma)k}, \qquad \text{for } \ j = 1, \ldots, k.$$

The inequality above, together with the definition of φ_k, yields

$$\varphi_k \leq \frac{8}{(1-\overline{\rho})^2(1-\sigma)^2\xi^2 k} \leq \frac{8}{(1-\overline{\rho})^2(1-\sigma)^2\xi^2}. \tag{2.76}$$

Defining the last term in (2.76) as $\overline{\varphi}$, and using Theorem 2.8, inequalities (2.75) and (2.76), we conclude that

$$\begin{aligned}
\|\overline{a}_k + \overline{b}_k\| &\leq \frac{8d_0}{\xi(1-\overline{\rho})(1-\sigma)k}, \\
\|\overline{x}_k - \overline{y}_k\| &\leq \frac{8d_0}{\xi(1-\overline{\rho})(1-\sigma)k}, \\
\overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k} &\leq \frac{4d_0^2(\overline{\varphi}+4)}{\xi(1-\overline{\rho})(1-\sigma)k},
\end{aligned}$$

for all integer $k \geq 1$. The inclusions in (2.71) together with the inequalities above imply statement (b). $\square$

2.5 A sequential inexact case

In this section we will study a variant of a sequential case of Algorithm 2. We observe that, unless $\alpha_k = 0$, subproblems (2.2) and (2.3) cannot be solved in parallel. For example, taking $\alpha_k = 1$ for all k, in step 2 of Algorithm 2, we have to perform on each iteration the following steps

$$\lambda_k b_k + x_k = z_{k-1} + \lambda_k w_{k-1}, \qquad b_k \in B(x_k); \tag{2.77}$$

$$\mu_k a_k + y_k = x_k - \mu_k w_{k-1}, \qquad a_k \in A(y_k). \tag{2.78}$$

Therefore (2.78) must to be solved after (2.77); these steps cannot be performed simultaneously like the proximal subproblems in step 1 of Algorithm 3. However, this choice of α_k could be an advantage since subproblem (2.78) uses more recent information, that is x_k instead of z_{k-1}.

In this section we will set $\alpha_k = 1$ for all integer $k \geq 1$ in Algorithm 2, and we will also allow the solution of the second proximal subproblem to be approximate, provided that the approximate solution satisfies the relative error condition of Definition 3.

Algorithm 4. *Start with* $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$, $\sigma \in [0, 1/2)$, $\overline{\rho} \in [0, 1)$. *For* $k = 1, 2, \ldots.$

1. *Choose* $\lambda_k > 0$ *and calculate* $(x_k, b_k) \in \mathbb{R}^n \times \mathbb{R}^n$ *and* $(y_k, a_k, \epsilon_{y,k}) \in \mathbb{E}$ *such that*

$$\lambda_k b_k + x_k = z_{k-1} + \lambda_k w_{k-1}, \qquad b_k \in B(x_k); \tag{2.79}$$

and

$$\lambda_k a_k + y_k = x_k - \lambda_k w_{k-1} + r_k, \qquad a_k \in A^{\epsilon_{y,k}}(y_k); \tag{2.80}$$

$$\|r_k\|^2 + 2\lambda_k \epsilon_{y,k} \ \leq \ \sigma \left(\|y_k - x_k\|^2 + \|\lambda_k (a_k + w_{k-1})\|^2 \right). \tag{2.81}$$

2. *If* $\|a_k + b_k\| + \|x_k - y_k\| = 0$ *stop. Otherwise, set*

$$\gamma_k = \frac{\langle z_{k-1} - x_k, b_k - w_{k-1} \rangle + \langle z_{k-1} - y_k, a_k + w_{k-1} \rangle - \epsilon_{y,k}}{\|a_k + b_k\|^2 + \|x_k - y_k\|^2}.$$

3. Choose a parameter $\rho_k \in [1-\overline{\rho}, 1+\overline{\rho}]$ *and set*

$$z_k = z_{k-1} - \rho_k \gamma_k (a_k + b_k),$$
$$w_k = w_{k-1} - \rho_k \gamma_k (x_k - y_k).$$

We notice that the maximum tolerance for the relative error in the resolution of (2.80)-(2.81) is 1/2, instead of 1 as in Algorithm 3. We also notice that the proximal parameter in step 1 of Algorithm 4 is not allowed to change from one subproblem to another within an iteration.

For all integer $k \geq 1$, the decomposable separator $\phi_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, associated to $(x_k, b_k, 0)$ and $(y_k, a_k, \epsilon_{y,k})$, is given by

$$\phi_k(z, w) = \langle z - x_k, b_k - w\rangle + \langle z - y_k, a_k + w\rangle - \epsilon_{y,k}. \tag{2.82}$$

Hence, if $\phi_k(z_{k-1}, w_{k-1}) > 0$ we have that Algorithm 4 is an instance of the general scheme presented in Section 1.2.

The following lemma implies that Algorithm 4 stops in step 2 whenever it has found a point in the extended solution set $S_e(A, B)$.

Lemma 2.3. *If* $\{(x_k, b_k)\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{(z_k, w_k)\}$, $\{\lambda_k\}$, *and* $\{\rho_k\}$ *are the sequences generated by Algorithm 4, and* $\{\phi_k\}$ *is the sequence of decomposable separators defined in* (2.82)*. Then, for all integer* $k \geq 1$*, we have*

$$\phi_k(z_{k-1}, w_{k-1}) \geq \frac{1-2\sigma}{2}\tau_k \left(\|a_k + b_k\|^2 + \|x_k - y_k\|^2\right) \geq 0, \tag{2.83}$$

where

$$\tau_k = \min\left\{\lambda_k, \frac{1}{\lambda_k}\right\}. \tag{2.84}$$

Furthermore, if $\|\nabla\phi_k\| > 0$ *then* $\phi_k(z_{k-1}, w_{k-1}) > 0$*, and* $\|\nabla\phi_k\| = 0$ *if and only if* $(x_k, b_k) = (y_k, -a_k) \in S_e(A, B)$*.*

Proof. Specializing identity (2.82) with $(z, w) = (z_{k-1}, w_{k-1})$, adding and subtracting $\langle x_k, a_k + w_{k-1}\rangle$ on the right-hand side and regrouping the terms, we obtain

$$\begin{aligned}\phi_k(z_{k-1}, w_{k-1}) &= \langle z_{k-1} - x_k, b_k + a_k\rangle + \langle x_k - y_k, a_k + w_{k-1}\rangle - \epsilon_{y,k}\\ &= \lambda_k \langle b_k - w_{k-1}, b_k + a_k\rangle + \frac{1}{2\lambda_k}\left[\|x_k - y_k\|^2 + \|\lambda_k(a_k + w_{k-1})\|^2\right]\\ &\quad - \frac{1}{2\lambda_k}\left[\|r_k\|^2 + 2\lambda_k\epsilon_{y,k}\right],\end{aligned} \tag{2.85}$$

where we have used in the last equality the identity in (2.79) and r_k is given by (2.80). We observe that

$$\lambda_k \langle b_k - w_{k-1}, b_k + a_k \rangle = \frac{\lambda_k}{2} \left[\|b_k - w_{k-1}\|^2 + \|b_k + a_k\|^2 - \|a_k + w_{k-1}\|^2 \right]. \tag{2.86}$$

Hence, combining equalities (2.85) and (2.86), and the error criterion (2.81) we have

$$\phi_k(z_{k-1}, w_{k-1}) \geq \frac{\lambda_k}{2} \|b_k - w_{k-1}\|^2 + \frac{\lambda_k}{2} \|a_k + b_k\|^2 + \frac{1-\sigma}{2\lambda_k} \|x_k - y_k\|^2 - \frac{\sigma \lambda_k}{2} \|a_k + w_{k-1}\|^2.$$

Since $\|a_k + w_{k-1}\|^2 \leq 2\|a_k + b_k\|^2 + 2\|b_k - w_{k-1}\|^2$, we deduce that

$$\phi_k(z_{k-1}, w_{k-1}) \geq \frac{\lambda_k(1-2\sigma)}{2} \|b_k - w_{k-1}\|^2 + \frac{\lambda_k(1-2\sigma)}{2} \|a_k + b_k\|^2 + \frac{1-\sigma}{2\lambda_k} \|x_k - y_k\|^2. \tag{2.87}$$

The inequalities in (2.83) now follow from the above relation, the definition of τ_k and noting that $1 - \sigma \geq 1 - 2\sigma > 0$.

The claim that $\|\nabla \phi_k\| > 0$ implies $\phi_k(z_{k-1}, w_{k-1}) > 0$ is obtained as a direct consequence of (2.83). To prove the last assertion of the lemma we notice that if $\|\nabla \phi_k\| = 0$ then $x_k = y_k$, $b_k = -a_k$, and it follows from (2.83), the first equality in (2.85) and the fact that $\epsilon_{y,k} \in \mathbb{R}_+$, that $\epsilon_{y,k} = 0$. Thus, we have $(x_k, b_k) \in S_e(A, B)$. □

We assume, from now on, that $\|\nabla \phi_k\| > 0$ for every integer $k \geq 1$.

The next result establishes pointwise iteration-complexity bounds for Algorithm 4. It will be proved in much the same way as Theorem 2.7, using Lemma 2.3 instead of Lemma 2.2.

Theorem 2.10. *Take $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$ and let $\{(x_k, b_k)\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\lambda_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 4. Let d_0 be the distance of (z_0, w_0) to $S_e(A, B)$, and for all integer $k \geq 1$ let τ_k be given by (2.84). Then, for every integer $k \geq 1$, we have*

$$b_k \in B(x_k), \qquad a_k \in A^{\epsilon_{y,k}}(y_k), \tag{2.88}$$

and there exists an index $1 \leq i \leq k$ such that

$$\|a_i + b_i\|^2 + \|x_i - y_i\|^2 \leq \frac{4d_0^2}{(1-2\sigma)^2(1-\overline{\rho})^2 \tau_i \sum_{j=1}^{k} \tau_j}, \tag{2.89}$$

$$\epsilon_{y,i} \leq \frac{4\sigma d_0^2}{(1-2\sigma)^2(1-\overline{\rho})^2 \sum_{j=1}^{k} \tau_j}. \tag{2.90}$$

Proof. The inclusions in (2.88) are due to step 1 of Algorithm 4. It follows from the definition of γ_k and inequality (2.83) that

$$\gamma_k \geq \left(\frac{1-2\sigma}{2}\right)\tau_k, \qquad \text{for } \; k = 1, 2, \ldots. \tag{2.91}$$

Squaring both sides of the above inequality and multiplying by $\|\nabla\phi_k\|^2$ we obtain

$$\gamma_k^2 \|\nabla\phi_k\|^2 \geq \left(\frac{1-2\sigma}{2}\right)^2 \tau_k^2 \|\nabla\phi_k\|^2, \qquad \text{for } \; k = 1, 2, \ldots. \tag{2.92}$$

Now, we notice that the error criterion (2.81) implies

$$\epsilon_{y,k} \leq \frac{\sigma}{2\lambda_k}\left[\|x_k - y_k\|^2 + \|\lambda_k(a_k + b_k)\|^2\right].$$

Hence,

$$\epsilon_{y,k} \leq \frac{\sigma}{2\lambda_k}\|x_k - y_k\|^2 + \sigma\lambda_k \|a_k + w_{k-1}\|^2 + \sigma\lambda_k \|b_k - w_{k-1}\|^2.$$

The above inequality, together with (2.87), yields

$$\epsilon_{y,k} \leq \frac{2\sigma}{1-2\sigma}\phi_k(z_{k-1}, w_{k-1}).$$

If we multiply the above relation by γ_k and combine with (2.91), after some manipulations, we obtain

$$\frac{(1-2\sigma)^2}{4\sigma}\tau_k\epsilon_{y,k} \leq \gamma_k^2 \|\nabla\phi_k\|^2. \tag{2.93}$$

Next, we define

$$\psi_k = \max\left\{\frac{(1-2\sigma)^2}{4}\tau_k \|\nabla\phi_k\|^2, \frac{(1-2\sigma)^2}{4\sigma}\epsilon_{y,k}\right\},$$

and using (2.92) and (2.93) we can conclude the proof proceeding analogously to the proof of Theorem 2.7. □

As in the previous sections, we will also derive ergodic complexity bounds for Algorithm 4.

Theorem 2.11. *Let $\{(x_k, b_k)\}$, $\{(y_k, a_k, \epsilon_{y,k})\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by Algorithm 4 and let $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$ and $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{y,k})\}$ be the associated sequences of ergodic iterates defined in (1.17), (1.18) and (1.19). Consider τ_k given by (2.84). Then, for all integer $k \geq 1$, we have*

$$\overline{b}_k \in B^{\overline{\epsilon}_{x,k}}(\overline{x}_k), \qquad \overline{a}_k \in A^{\overline{\epsilon}_{y,k}}(\overline{y}_k), \tag{2.94}$$

and

$$\left\|\bar{a}_k + \bar{b}_k\right\| \leq \frac{2d_0}{\Gamma_k}, \qquad \|\bar{x}_k - \bar{y}_k\| \leq \frac{2d_0}{\Gamma_k}, \qquad \bar{\epsilon}_{x,k} + \bar{\epsilon}_{y,k} \leq \frac{d_0^2(\vartheta_k + 4)}{\Gamma_k}; \tag{2.95}$$

where d_0 *is the distance of* (z_0, w_0) *to* $S_e(A, B)$ *and*

$$\vartheta_k = \max_{j=1,\dots,k} \left\{ \frac{8}{\tau_j(1-2\sigma)(2-\rho_j)\Gamma_k} \right\}.$$

Proof. Since Algorithm 4 is an instance of Algorithm 1, Lemma 1.4 and Theorem 1.3 apply, therefore the inclusions in (2.94) and the first two inequalities in (2.95) follow.

We derive now an estimate for the sum on the right-hand side of (1.29). We notice that (2.87) implies

$$\phi_j(z_{j-1}, w_{j-1}) \geq \lambda_j \left(\frac{1-2\sigma}{2}\right) \|b_j - w_{j-1}\|^2, \tag{2.96}$$

for all integer $j \geq 1$. We also notice that

$$z_{j-1} - y_j = z_{j-1} - x_j + x_j - y_j = \lambda_j(b_j - w_{j-1}) + x_j - y_j,$$

where the last identity is due to the equality in (2.79). This last expression and the triangle inequality for norms yield

$$\|z_{j-1} - y_j\| \leq \lambda_j \|b_j - w_{j-1}\| + \|x_j - y_j\|.$$

Moreover, squaring both sides of the inequality above and making some manipulations, we have

$$\begin{aligned} \frac{1}{2\lambda_j} \|z_{j-1} - y_j\|^2 &\leq \lambda_j \|b_j - w_{j-1}\|^2 + \frac{1}{\lambda_j} \|x_j - y_j\|^2 \\ &\leq \frac{2}{1-2\sigma} \phi_j(z_{j-1}, w_{j-1}), \end{aligned} \tag{2.97}$$

where the last inequality follows from (2.87). Adding now (2.96) and (2.97) we obtain

$$\frac{1-2\sigma}{4} \frac{1}{\lambda_j} \|z_{j-1} - y_j\|^2 + \frac{1-2\sigma}{2} \lambda_j \|b_j - w_{j-1}\|^2 \leq 2\phi_j(z_{j-1}, w_{j-1}). \tag{2.98}$$

The above relation, together with the definitions of γ_j and τ_j, implies

$$\|b_j - w_{j-1}\|^2 + \|z_{j-1} - y_j\|^2 \leq \frac{8}{(1-2\sigma)\tau_j} \gamma_j \|\nabla\phi_j\|^2.$$

Multiplying both sides of the above relation by $\frac{1}{\Gamma_k}\rho_j\gamma_j$, and adding from $j = 1$ to k, we obtain the desired bound, i.e.

$$\begin{aligned}\frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j\left[\|b_j-w_{j-1}\|^2+\|z_{j-1}-y_j\|^2\right] &\le \frac{1}{\Gamma_k}\sum_{j=1}^{k}\frac{8}{(1-2\sigma)\tau_j}\rho_j\gamma_j^2\left\|\nabla\phi_j\right\|^2\\ &= \frac{1}{\Gamma_k}\sum_{j=1}^{k}\frac{8}{(1-2\sigma)\tau_j(2-\rho_j)}\rho_j(2-\rho_j)\gamma_j^2\left\|\nabla\phi_j\right\|^2\\ &\le \vartheta_k\sum_{j=1}^{k}\rho_j(2-\rho_j)\gamma_j^2\left\|\nabla\phi_j\right\|^2\\ &\le \vartheta_k d_0^2,\end{aligned}$$

where the second and the third inequalities above follow from the definition of ϑ_k and (1.16), respectively. The proof of the last bound in (2.95) now follows combining the above relation with (1.29). □

We finish this section with the following theorem, which provides complexity bounds for Algorithm 4 to find a (δ,ϵ)-approximate solution of problem (1.1) in terms of the stopping criterion (2.14). It will be proved in much the same manner as Theorem 2.9.

Theorem 2.12. *Assume the hypotheses of Theorem 2.11. Suppose also that there exist $\overline{\lambda}$ and $\underline{\lambda}$ such that, $\overline{\lambda}\ge\underline{\lambda}>0$ and $\lambda_k\in[\underline{\lambda},\overline{\lambda}]$ for all integer $k\ge 1$. Defining $\tau=\min\left\{\underline{\lambda},\frac{1}{\overline{\lambda}}\right\}$, then, for every δ, $\epsilon>0$, the following claims hold.*

(a) There exists an index

$$i=\mathcal{O}\left(\max\left\{\frac{d_0^2}{\tau^2\delta^2},\frac{d_0^2}{\tau\epsilon}\right\}\right)$$

such that, the point (x_i,y_i) calculated by Algorithm 4 is a (δ,ϵ)-approximate solution of problem (1.1)*.*

(b) There exists an index

$$k_0=\mathcal{O}\left(\max\left\{\frac{d_0}{\tau\delta},\frac{d_0^2}{\tau\epsilon}\right\}\right)$$

such that, for all integer $k\ge k_0$ the ergodic iterate $(\overline{x}_k,\overline{y}_k)$ is a (δ,ϵ)-approximate solution of problem (1.1)*.*

Proof. We observe that the definitions of τ_k and τ in (2.84) and the statement of the theorem, respectively, imply that

$$\tau_k\ge\tau,\qquad \text{for } k=1,2,\ldots. \tag{2.99}$$

Therefore, applying Theorem 2.10 and combining with (2.99), we have that there exists and index $1 \le i \le k$ such that

$$\|a_i + b_i\| \le \frac{2d_0}{(1-2\sigma)(1-\overline{\rho})\tau\sqrt{k}}, \qquad \|x_i - y_i\| \le \frac{2d_0}{(1-2\sigma)(1-\overline{\rho})\tau\sqrt{k}},$$
$$\epsilon_{y,i} \le \frac{4d_0^2}{(1-2\sigma)^2(1-\overline{\rho})^2 k\tau}.$$

Item *(a)* now follows from the inclusions in (2.88) and the inequalities above.

To prove statement *(b)* we first notice that the definition of Γ_k in (1.17), together with inequality (2.91), yields

$$\Gamma_k \ge (1-\overline{\rho})\left(\frac{1-2\sigma}{2}\right)\sum_{j=1}^{k}\tau_j.$$

Combining the inequality above with (2.99) we obtain

$$\Gamma_k \ge (1-\overline{\rho})\left(\frac{1-2\sigma}{2}\right)\tau k, \qquad \text{for } k = 1, 2, \ldots. \tag{2.100}$$

We next observe that the definition of ϑ_k in the statement of Theorem 2.11 and relations (2.99) and (2.100) imply

$$\vartheta_k \le \frac{16}{\tau^2(1-2\sigma)^2(1-\overline{\rho})^2 k} \le \frac{16}{\tau^2(1-2\sigma)^2(1-\overline{\rho})^2}, \qquad \text{for } k = 1, 2, \ldots.$$

Defining the last term in the above relation as $\overline{\vartheta}$, and using Theorem 2.11 and inequality (2.100), we obtain

$$\left\|\overline{a}_k + \overline{b}_k\right\| \le \frac{4d_0}{(1-\overline{\rho})(1-2\sigma)\tau k},$$
$$\|\overline{x}_k - \overline{y}_k\| \le \frac{4d_0}{(1-\overline{\rho})(1-2\sigma)\tau k},$$
$$\overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k} \le \frac{2d_0(\overline{\vartheta}+4)}{(1-\overline{\rho})(1-2\sigma)\tau k}.$$

Statement *(b)* follows from the above inequalities and inclusions (2.94). □

Chapter 3

Application

In this chapter we will develop an specific instance of the splitting-projective framework of Section 2.5. Namely, we will use Algorithm 4 to solve a class of linearly constrained optimization problems.

In Section 3.1 we briefly discuss Lagrangian duality theory for convex optimization, for more details we refer the reader to [32]. We also introduce in this section the scheme that we study and we obtain complexity bounds for the method using the results of Section 2.5.

In Section 3.2 we present the ROF [35] model for image restoration and apply the algorithm developed in the previous section to solve this problem. Some numerical tests are included in Subsection 3.2.1.

3.1 Convex optimization problems

Consider the following optimization problem:

$$\min_{(u,v)} \{f(u) + g(v) \,:\, Mu + Cv = d\}, \tag{3.1}$$

where $d \in \mathbb{R}^n$, $M : \mathbb{R}^{m_1} \to \mathbb{R}^n$ and $C : \mathbb{R}^{m_2} \to \mathbb{R}^n$ are linear operators, and $f : \mathbb{R}^{m_1} \to (-\infty, \infty]$ and $g : \mathbb{R}^{m_2} \to (-\infty, \infty]$ are proper closed convex functions.

The *Lagrangian function* $L : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \mathbb{R}^n \to (-\infty, \infty]$ for problem (3.1) is defined as

$$L(u, v, z) = f(u) + g(v) + \langle Mu + Cv - d, z\rangle\,. \tag{3.2}$$

The *dual function* is the concave function $q : \mathbb{R}^n \to [-\infty, \infty)$ defined by

$$q(z) = \inf_{(u,v)\in\mathbb{R}^{m_1}\times\mathbb{R}^{m_2}} L(u, v, z).$$

The *dual problem* to (3.1) is

$$\max_{z\in\mathbb{R}^n} q(z). \tag{3.3}$$

Problem (3.1) will be called the *primal problem*. A straightforward application of the Fenchel-Young inequality shows that weak duality holds, i.e. $q^* \leq p^*$, where p^* and q^* are the optimal values of (3.1) and (3.3), respectively. Observe that the dual function q can be written in terms of the Fenchel-Legendre conjugates of the functions f and g. Specifically,

$$\begin{aligned} q(z) &= \inf_{(u,v)} \left[f(u) + g(v) + \langle Mu + Cv - d, z\rangle\right] \\ &= \inf_u \left[f(u) + \langle Mu, z\rangle\right] + \inf_v \left[g(v) + \langle Cv, z\rangle\right] - \langle d, z\rangle \\ &= - f^*(-M^*z) - g^*(-C^*z) - \langle d, z\rangle . \end{aligned}$$

Moreover, if we define the functions $h_1(z) = \left(f^* \circ -M^*\right)(z)$ and $h_2(z) = \left(g^* \circ -C^*\right)(z) + \langle d, z\rangle$, the dual problem (3.3) is equivalent to minimizing $h_1 + h_2$ over $\mathbb{R}^n$. Thus, z^* is a solution of (3.3) if and only if

$$0 \in \partial(h_1 + h_2)(z^*). \tag{3.4}$$

A vector (u^*, v^*, z^*) is said to be a *saddle point* of the Lagrangian function L, if $L(u^*, v^*, z^*)$ is finite and

$$\min_{(u,v)\in\mathbb{R}^{m_1}\times\mathbb{R}^{m_2}} L(u, v, z^*) = L(u^*, v^*, z^*) = \max_{z\in\mathbb{R}^n} L(u^*, v^*, z). \tag{3.5}$$

Finding optimal solutions of problems (3.1) and (3.3) is equivalent to finding saddle points of L. That is, (u^*, v^*) is an optimal primal solution and z^* is an optimal dual solution if and only if (u^*, v^*, z^*) satisfy (3.5), see [32]. Furthermore, if a saddle point of L exists then $p^* = q^*$, i.e. there is no duality gap [32].

Note that, if (u^*, v^*, z^*) is a saddle point of L, from (3.2) and (3.5) it follows that

$$f(u) + g(v) + \langle Mu + Cv - d, z^*\rangle \geq L(u^*, v^*, z^*) \geq f(u^*) + g(v^*) + \langle Mu^* + Cv^* - d, z\rangle$$

for all $u \in \mathbb{R}^{m_1}$, $v \in \mathbb{R}^{m_2}$, $z \in \mathbb{R}^n$. From these relations we can directly derive the Kuhn-Tucker optimality conditions for problem (3.1)

$$\begin{aligned} 0 &= Mu^* + Cv^* - d, \\ 0 &\in \partial f(u^*) + M^*z^*, \\ 0 &\in \partial g(v^*) + C^*z^*. \end{aligned} \tag{3.6}$$

An iterative method used for finding saddle points of the Lagrangian function L is the *Alternating Direction Method of Multipliers* (ADMM), which goes back to Glowinski and Marroco [20], and Gabay and Mercier [18]. Gabay [17] showed that ADMM can be interpreted as Douglas-Rachford splitting algorithm applied to the operators ∂h_1 and ∂h_2. This approach motivates us to solve problems (3.1)-(3.3) by applying the projective-splitting framework developed in the previous chapter.

We make the following assumptions throughout this section:

B.1 There exists (u^*, v^*, z^*) a saddle point of L;

B.2 $\mathrm{ri}\,(\mathrm{dom}\, f^*) \cap \mathrm{range}\, M^* \neq \emptyset$;

B.3 $\mathrm{ri}\,(\mathrm{dom}\, g^*) \cap \mathrm{range}\, C^* \neq \emptyset$.

We notice that conditions B.2 and B.3 imply, respectively, that h_1 and h_2 are proper functions. Therefore ∂h_1 and ∂h_2 are maximal monotone operators.

Lemma 3.1. *If (u^*, v^*, z^*) is a saddle point of L, then*

$$(z^*, d - Cv^*) \in S_e\left(\partial h_1, \partial h_2\right).$$

Proof. If (u^*, v^*, z^*) is a saddle point of the Lagrangian function, then the Kuhn-Tucker optimality conditions hold and by the inclusions in (3.6) and Proposition 1.2(b) we have

$$u^* \in \partial f^*(-M^*z^*), \qquad v^* \in \partial g^*(-C^*z^*).$$

Thus,

$$-Mu^* \in -M\partial f^*(-M^*z^*) \subseteq \partial(f^* \circ -M^*)(z^*), \tag{3.7}$$
$$-Cv^* \in -C\partial g^*(-C^*z^*) \subseteq \partial(g^* \circ -C^*)(z^*); \tag{3.8}$$

where the last inclusions in (3.7) and (3.8) follow from Proposition 1.2(d). Adding d to both sides of (3.8) and using the definitions of h_2 and Proposition 1.2(c) we have $d - Cv^* \in \partial h_2(z^*)$. Now, adding this last inclusion with (3.7) and noting the definition of h_1 we conclude that

$$-Mu^* + d - Cv^* \in \partial h_1(z^*) + h_2(z^*).$$

The lemma follows combining the relation above with the equality in (3.6) and the definition of $S_e\left(\partial h_1, \partial h_2\right)$. □

The next result shows how we can invert operators $I + \lambda\partial h_1$ and $I + \lambda\partial h_2$ for all $\lambda > 0$.

Lemma 3.2. *Let $z \in \mathbb{R}^n$, $\lambda > 0$ and assume conditions B.1-B.3, then the following claims hold.*

(a) If $\tilde{u} \in \mathbb{R}^{m_1}$ is a solution of the following problem

$$\min_{u \in \mathbb{R}^{m_1}} \left\{ f(u) + \langle z, Mu \rangle + \frac{\lambda}{2} \|Mu\|^2 \right\}, \tag{3.9}$$

then, $-M\tilde{u} \in \partial h_1(\tilde{z})$ where $\tilde{z} = z + \lambda M\tilde{u}$. Hence, $\tilde{z} = (I + \lambda\partial h_1)^{-1}(z)$. Furthermore, the set of optimal solutions of (3.9) is nonempty.

(b) If $\tilde{v} \in \mathbb{R}^{m_2}$ is a solution of problem

$$\min_{v \in \mathbb{R}^{m_2}} \left\{ g(v) + \langle z, Cv - d \rangle + \frac{\lambda}{2} \|Cv - d\|^2 \right\}, \tag{3.10}$$

then, $d - C\tilde{v} \in \partial h_2(\hat{z})$ where $\hat{z} = z + \lambda(C\tilde{v} - d)$. Hence, $\hat{z} = (I + \lambda\partial h_2)^{-1}(z)$. Furthermore, the set of optimal solutions of (3.10) is nonempty.

Proof. (a) If $\tilde{u} \in \mathbb{R}^{m_1}$ is a solution of (3.9), deriving the optimality condition of this minimization problem, we have

$$0 \in \partial f(\tilde{u}) + M^* z + \lambda M^* M\tilde{u} = \partial f(\tilde{u}) + M^*\tilde{z},$$

where the last identity follows from the definition of $\tilde{z}$. The inclusion above and Proposition 1.2(b),(d) yield

$$-M\tilde{u} \in \partial(f^* \circ -M^*)(\tilde{z}).$$

Using the above relation and the definition of h_1, we deduce that $-M\tilde{u} \in \partial h_1(\tilde{z})$. The assertion that $\tilde{z} = (I + \lambda\partial h_1)^{-1}(z)$ is a direct consequence of this last inclusion and the definition of $\tilde{z}$.

Since ∂h_1 is maximal monotone, Minty's theorem [26] asserts that for all $z \in \mathbb{R}^n$ there exist $\tilde{z}$, $w \in \mathbb{R}^n$ such that

$$\begin{cases} w \in \partial h_1(\tilde{z}), \\ \lambda w + \tilde{z} = z. \end{cases} \tag{3.11}$$

Moreover, assumption B.2 and Proposition 1.2(d) imply that

$$\partial h_1(z) = -M\partial f^*(-M^* z) \qquad \forall z \in \mathbb{R}^n. \tag{3.12}$$

Therefore, by (3.12) there exists $\tilde{u} \in \partial f^*(-M^*\tilde{z})$ such that $w = -M\tilde{u}$, where w is given in (3.11). This last inclusion implies that $-M^*\tilde{z} \in \partial f(\tilde{u})$, from which we deduce that

$$0 \in \partial f(\tilde{u}) + M^*\tilde{z} = \partial f(\tilde{u}) + M^*(z + \lambda M\tilde{u}),$$

where the last identity is obtained replacing w by $-M\tilde{u}$ in the equality in (3.11). Thus, it follows from the relation above that $\tilde{u}$ is an optimal solution of problem (3.9).

(b) The same proof of statement (a) remains valid for item (b), noting that condition B.3 and Proposition 1.2(c),(d) yield

$$\partial h_2(z) = \partial(g^* \circ -C^*)(z) + d = -C\partial g^*(-C^* z) + d, \qquad \forall z \in \mathbb{R}^n. \tag{3.13}$$

□

We notice that equation (3.12) implies that for all z, $w \in \mathbb{R}^n$ with $w \in \partial h_1(z)$ there exists $u \in \mathbb{R}^{m_1}$ such that $u \in \partial f^*(-M^* z)$ and $w = -Mu$. Similarly, if w is a subgradient of h_2 at z, by (3.13) we have that there is $v \in \mathbb{R}^{m_2}$ satisfying $v \in \partial g^*(-C^* z)$ and $w = d - Cv$. Hence, by the definition of the extended solution set, if $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$, there exist u^* and v^* with $w^* = -Mu^*$ and $-w^* = d - Cv^*$, from which follows that (u^*, v^*, z^*) is a saddle point of L.

Thus, according to Lemma 3.1, we can attempt to find a saddle point of the Lagrangian function (3.2), by seeking a point in the extended solution set $S_e(\partial h_1, \partial h_2)$.

We now state the method that we wish to study in this section.

Algorithm 5. *Let $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$, $\lambda > 0$ and $\overline{\rho} \in [0, 1)$ be given. For $k = 1, 2, \ldots$.*

1. *Compute $v_k \in \mathbb{R}^{m_2}$ as*

$$v_k = \arg\min_v \{g(v) + \langle z_{k-1} + \lambda w_{k-1}, Cv - d\rangle + \frac{\lambda}{2}\|Cv - d\|^2\}, \tag{3.14}$$

and $u_k \in \mathbb{R}^{m_1}$ as

$$u_k = \arg\min_u \{f(u) + \langle z_{k-1} + \lambda(Cv_k - d), Mu\rangle + \frac{\lambda}{2}\|Mu\|^2\}. \tag{3.15}$$

2. *If $\|Mu_k + Cv_k - d\| + \|Mu_k - w_{k-1}\| = 0$ stop. Otherwise, set*

$$\gamma_k = \frac{\lambda\|Cv_k - d + w_{k-1}\|^2 + \lambda\langle d - Cv_k - Mu_k, w_{k-1} - Mu_k\rangle}{\|Mu_k + Cv_k - d\|^2 + \lambda^2\|Mu_k - w_{k-1}\|^2}.$$

3. *Choose $\rho_k \in [1 - \overline{\rho}, 1 + \overline{\rho}]$ and set*

$$z_k = z_{k-1} + \rho_k\gamma_k(Mu_k + Cv_k - d),$$
$$w_k = w_{k-1} - \rho_k\gamma_k\lambda(w_{k-1} - Mu_k).$$

Proposition 3.1. *Algorithm 5 is a special instance of Algorithm 4 applied to operators*

$A = \partial h_1$ and $B = \partial h_2$, with $\sigma = 0$,

$$\lambda_k = \lambda, \qquad \epsilon_{y,k} = 0, \tag{3.16}$$

and

$$\begin{aligned} x_k &= z_{k-1} + \lambda w_{k-1} + \lambda(Cv_k - d), & b_k &= d - Cv_k, \\ y_k &= x_k - \lambda(w_{k-1} - Mu_k), & a_k &= -Mu_k, \end{aligned} \tag{3.17}$$

for every integer $k \geq 1$.

Proof. Applying Lemma 3.2(b) with $z = z_{k-1} + \lambda w_{k-1}$ and $\tilde{v} = v_k$ we have that x_k and b_k, defined in (3.17), satisfy $b_k \in \partial h_2(x_k)$ and $x_k = (I + \lambda \partial h_2)^{-1}(z_{k-1} + \lambda w_{k-1})$. Therefore, we deduce that (x_k, b_k) is the solution of subproblem (2.79) with $\lambda_k = \lambda$ and $B = \partial h_2$. Furthermore, applying Lemma 3.2(a) with $z = x_k - \lambda w_{k-1}$ and $\tilde{u} = u_k$ we have that the points y_k and a_k, given in (3.17), satisfy (2.80) and (2.81) with $A = \partial h_1$, $\lambda_k = \lambda$, $\epsilon_{y,k} = 0$ and $\sigma = 0$.

Moreover, identities in (3.17) yield

$$b_k + a_k = d - Cv_k - Mu_k, \qquad x_k - y_k = \lambda(w_{k-1} - Mu_k), \tag{3.18}$$

and

$$z_{k-1} - y_k = \lambda(d - Mu_k - Cv_k).$$

Using these relations above and the definitions of x_k, b_k, y_k and a_k in (3.17), we can rewrite γ_k in step 2 of Algorithm 5 as

$$\gamma_k = \frac{\langle z_{k-1} - x_k, b_k - w_{k-1} \rangle + \langle z_{k-1} - y_k, a_k + w_{k-1} \rangle}{\|a_k + b_k\|^2 + \|x_k - y_k\|^2}.$$

Finally, we observe that (3.18) and the update rule of step 3 of Algorithm 5 imply that

$$(z_k, w_k) = (z_{k-1}, w_{k-1}) - \rho_k \gamma_k (a_k + b_k, x_k - y_k).$$

Thus, the proposition is proven. □

We are now able to establish the convergence rate result for Algorithm 5.

Theorem 3.1. *Consider the sequences $\{(u_k, v_k)\}$, $\{(z_k, w_k)\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ generated by Algorithm 5, and the sequences $\{x_k\}$, $\{b_k\}$, $\{y_k\}$ and $\{a_k\}$ defined in (3.17). Moreover, consider the associated sequences of ergodic iterates $\{(\overline{x}_k, \overline{b}_k, \overline{\epsilon}_{x,k})\}$ and $\{(\overline{y}_k, \overline{a}_k, \overline{\epsilon}_{y,k})\}$ given by (1.17), (1.18) and (1.19). Let d_0 be the distance of (z_0, w_0) to $S_e(\partial h_1, \partial h_2)$ and define $\tau = \min\left\{\lambda, \dfrac{1}{\lambda}\right\}$. Then for all integer $k \geq 1$, the following statements hold.*

(a) $b_k \in \partial h_2(x_k)$, $a_k \in \partial h_1(y_k)$, *and there exists* $i \le k$ *such that*

$$\|a_i + b_i\| \le \frac{2d_0}{(1-\overline{\rho})\tau\sqrt{k}}, \qquad \|x_i - y_i\| \le \frac{2d_0}{(1-\overline{\rho})\tau\sqrt{k}}.$$

(b) *We have*

$$\overline{b}_k \in \partial_{\overline{\epsilon}_{x,k}} h_2(\overline{x}_k), \qquad \overline{a}_k \in \partial_{\overline{\epsilon}_{y,k}} h_1(\overline{y}_k), \tag{3.19}$$

and

$$\left\|\overline{a}_k + \overline{b}_k\right\| \le \frac{4d_0}{k(1-\overline{\rho})\tau}, \qquad \|x_k - y_k\| \le \frac{4d_0}{k(1-\overline{\rho})\tau},$$
$$\overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k} \le \frac{2d_0^2}{k(1-\overline{\rho})\tau}(\vartheta + 4),$$

where

$$\vartheta := \frac{16}{\tau^2(1-\overline{\rho})^2}. \tag{3.20}$$

Proof. This result follows immediately from Proposition 3.1 and Theorems 2.10 and 2.11 by specializing the last two results to the case where $\tau_k = \tau$, $\sigma = 0$, and noting that Γ_k and ϑ_k in Theorem 2.11 satisfy

$$\Gamma_k \ge k(1-\overline{\rho})\frac{\tau}{2} \qquad \text{and} \qquad \vartheta_k \le \frac{16}{\tau^2(1-\overline{\rho})^2 k} \le \vartheta, \qquad \text{for } k = 1, 2, \ldots.$$

Observe also that (3.19) follows from Theorem 1.1(b) and the fact that $b_k \in \partial h_2(x_k)$ and $a_k \in \partial h_1(y_k)$, according to statement (a). □

The above result can be translated from the context of the monotone inclusion problem (1.1) with ∂h_1, ∂h_2, to the context of the minimization problem (3.1) and the Kuhn-Tucker optimality conditions (3.6).

Theorem 3.2. *Assume the hypotheses of Theorem 3.1 and define for every integer* $k \ge 1$

$$\overline{u}_k = \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j \gamma_j u_j, \qquad \overline{v}_k = \frac{1}{\Gamma_k}\sum_{j=1}^{k} \rho_j \gamma_j v_j, \qquad \textit{where} \quad \Gamma_k = \sum_{j=1}^{k} \rho_j \gamma_j. \tag{3.21}$$

Then,

(a) *for all* $k = 1, 2, \ldots,$

$$0 \in \partial g(v_k) + C^* x_k, \qquad 0 \in \partial f(u_k) + M^* y_k, \tag{3.22}$$

and there exist and index $i \leq k$ such that

$$\|Mu_i + Cv_i - d\| \leq \frac{2d_0}{(1-\overline{\rho})\tau\sqrt{k}}, \qquad \|x_i - y_i\| \leq \frac{2d_0}{(1-\overline{\rho})\tau\sqrt{k}}; \tag{3.23}$$

(b) for all $k = 1, 2, \ldots,$

$$0 \in \partial_{\overline{\epsilon}_{x,k}} g(\overline{v}_k) + C^*\overline{x}_k, \qquad 0 \in \partial_{\overline{\epsilon}_{y,k}} f(\overline{u}_k) + M^*\overline{y}_k, \tag{3.24}$$

and

$$\begin{aligned} &\|M\overline{u}_k + C\overline{v}_k - d\| \leq \frac{4d_0}{k(1-\overline{\rho})\tau}, \qquad \|\overline{x}_k - \overline{y}_k\| \leq \frac{4d_0}{k(1-\overline{\rho})\tau}, \\ &\overline{\epsilon}_{x,k} + \overline{\epsilon}_{y,k} \leq \frac{2d_0^2}{k(1-\overline{\rho})\tau}(\vartheta + 4); \end{aligned} \tag{3.25}$$

where ϑ is given in (3.20)*.*

Proof. First we observe that the optimality condition of problems (3.14) and (3.15), together with the definitions of x_k and y_k in (3.17), imply inclusions in (3.22). The estimates in (3.23) are due to Theorem 3.1(a) and the fact that $a_k + b_k = d - Mu_k - Cv_k$, for all integer $k \geq 1$.

Moreover, the inclusions in (3.24) follow from (3.21), (3.22), Theorem 1.1(b) and linearity of the M^* and C^* operators. The definitions of a_k, b_k in (3.17), together with the definitions of $\overline{u}_k$, $\overline{v}_k$, $\overline{b}_k$ and $\overline{a}_k$ in (3.21), (1.18) and (1.19), respectively, yield

$$\overline{a}_k + \overline{b}_k = d - M\overline{u}_k - C\overline{v}_k.$$

Therefore, the bounds in (3.25) follow from the identity above and Theorem 3.1(b). □

Clearly, it is also possible to apply Algorithms 2 and 3 to operators ∂h_1 and ∂h_2 to find a saddle point of the Lagrangian function of problem (3.1), and use the results obtained in Sections 2.1 and 2.4 to establish its convergence rate. For instance, using Algorithm 3 for locating a point in $S_e(\partial h_1, \partial h_2)$, with $\sigma = 0$ and $\lambda_k = \mu_k = \lambda$ for all integer $k \geq 1$, gives a parallel method with pointwise and ergodic convergence rates of $\mathcal{O}(1/\sqrt{k})$ and $\mathcal{O}(1/k)$, respectively.

3.2 TV denoising

In this section we discuss the specialization of Algorithm 5 to the total variation model for image denoising (TV denoising). Total variation or Rudin-Osher-Fatemi (ROF) model is a common image restoration model first introduced in [35], which has been the object of intense research, see for instance [8, 11, 42, 2, 9], and references therein.

Let $b \in \mathbb{R}^{m\times n}$ be an observed noisy image and $u \in \mathbb{R}^{m\times n}$ the original image to be recovered, the anisotropic TV problem for image restoration is the minimization problem

$$\min_{u\in\mathbb{R}^{m\times n}} \zeta TV(u) + \frac{1}{2}\|u-b\|_F^2, \tag{3.26}$$

where $\zeta > 0$, $\|\cdot\|_F$ is the Frobenius norm for matrix and TV is the total variation norm defined as

$$\begin{aligned} TV(u) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1} |u_{i,j}-u_{i+1,j}| + |u_{i,j}-u_{i,j+1}| \\ &+ \sum_{i=1}^{m-1}|u_{i,n}-u_{i+1,n}| + \sum_{j=1}^{n-1}|u_{m,j}-u_{m,j+1}|, \end{aligned} \tag{3.27}$$

where we assumed in (3.27) standard reflexive boundary conditions

$$u_{m+1,j} - u_{m,j} = 0, \quad \forall j \qquad \text{and} \qquad u_{i,n+1} - u_{i,n} = 0, \quad \forall i.$$

If we consider the discrete forward gradients $\nabla_1 : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ and $\nabla_2 : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ in the first and second direction, respectively, given by

$$\begin{aligned} (\nabla_1 u)_{ij} &= u_{i+1,j} - u_{i,j}, \\ (\nabla_2 u)_{ij} &= u_{i,j+1} - u_{i,j}, \end{aligned} \qquad \forall i = 1,\dots,m,\ j=1,\dots,n, \quad \forall u \in \mathbb{R}^{m\times n}; \tag{3.28}$$

then the TV norm can be stated as

$$TV(u) = \|\nabla_1 u\|_1 + \|\nabla_2 u\|_1.$$

Thus, problem (3.26) can be rewritten as

$$\min_{u\in\mathbb{R}^{m\times n}} \zeta\|\nabla_1 u\|_1 + \zeta\|\nabla_2 u\|_1 + \frac{1}{2}\|u-b\|_F^2. \tag{3.29}$$

Let us state the problem above in the form of a linearly constrained minimization problem (3.1). Define $\Omega := \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$, and the linear map $\nabla : \mathbb{R}^{m\times n} \to \Omega$ by

$$\nabla u = (\nabla_1 u, \nabla_2 u).$$

If we set $v = \nabla u \in \Omega$, then problem (3.29) is equivalent to

$$\min_{(u,v)} \left\{ \zeta\|v\|_1 + \frac{1}{2}\|b-u\|_F^2 \,:\, \nabla u - v = 0 \right\}. \tag{3.30}$$

To solve the minimization problem above we will apply Algorithm 5 with $f(u) = \frac{1}{2} \|u - b\|_F^2$, $g(v) = \zeta \|v\|_1$, $M = \nabla$, $C = -I$ and $d = 0$. For simplicity of notation and to avoid confusion, from now on we write the iteration counter k, at the top of the iterated vectors. Given z^{k-1}, $w^{k-1} \in \Omega$, Algorithm 5 requires the solution of subproblems,

$$v^k = \arg \min_{v \in \Omega} \left\{ \zeta \|v\|_1 - \left\langle z^{k-1} + \lambda w^{k-1}, v \right\rangle + \frac{\lambda}{2} \|v\|_F^2 \right\}, \tag{3.31}$$

and

$$u^k = \arg \min_{u \in \mathbb{R}^{m \times n}} \left\{ \frac{1}{2} \|u - b\|_F^2 + \left\langle z^{k-1} - \lambda v^k, \nabla u \right\rangle + \frac{\lambda}{2} \|\nabla u\|_F^2 \right\}. \tag{3.32}$$

The optimality condition of problem (3.31), yields

$$0 \in \zeta \partial \|\cdot\|_1 (v^k) - (z^{k-1} + \lambda w^{k-1}) + \lambda v^k,$$

hence,

$$v^k = \left(I + \frac{\zeta}{\lambda} \partial \|\cdot\|_1 \right)^{-1} \left(\frac{1}{\lambda} z^{k-1} + w^{k-1} \right).$$

Furthermore, the resolvent above can be computed explicitly as

$$\begin{aligned} v1_{i,j}^k &= \max \left[0, \left| \frac{1}{\lambda} z1_{i,j}^{k-1} + w1_{i,j}^{k-1} \right| - \frac{\zeta}{\lambda} \right] sgn \left(\frac{1}{\lambda} z1_{i,j}^{k-1} + w1_{i,j}^{k-1} \right), \\ v2_{i,j}^k &= \max \left[0, \left| \frac{1}{\lambda} z2_{i,j}^{k-1} + w2_{i,j}^{k-1} \right| - \frac{\zeta}{\lambda} \right] sgn \left(\frac{1}{\lambda} z2_{i,j}^{k-1} + w2_{i,j}^{k-1} \right). \end{aligned} \tag{3.33}$$

Deriving now the optimality condition for subproblem (3.32) we have that

$$0 = u^k - b + \nabla^*(z^{k-1} - \lambda v^k) + \lambda \nabla^* \nabla u^k,$$

from which follows that u^k has to be the solution of the system of linear equations

$$(I + \lambda \nabla^* \nabla) u^k = b - \nabla^*(z^{k-1} - \lambda v^k).$$

Thus, the iterations of Algorithm 5 for solving problem (3.30) are given by

$$v^k = \max\left[0, \left|\frac{1}{\lambda}z^{k-1} + w^{k-1}\right| - \frac{\zeta}{\lambda}\right] sgn\left(\frac{1}{\lambda}z^{k-1} + w^{k-1}\right), \tag{3.34}$$

$$(I + \lambda\nabla^*\nabla)u^k = b - \nabla^*(z^{k-1} - \lambda v^k), \tag{3.35}$$

$$\gamma_k = \frac{\lambda\left\|w^{k-1} - v^k\right\|^2 + \lambda\left\langle v^k - \nabla u^k, w^{k-1} - \nabla u^k\right\rangle}{\left\|\nabla u^k - v^k\right\|^2 + \lambda^2\left\|\nabla u^k - w^{k-1}\right\|^2}, \tag{3.36}$$

$$z^k = z^{k-1} + \rho_k\gamma_k(\nabla u^k - v^k), \tag{3.37}$$

$$w^k = w^{k-1} - \rho_k\gamma_k\lambda(w^{k-1} - \nabla u^k); \tag{3.38}$$

where identity (3.34) stands for the formulas (3.33).

3.2.1 Numerical tests

We will now exhibit some preliminary numerical experiments to illustrate the performance of Algorithm 5 when solving the anisotropic TV denoising model. We tested our method with two images: the first was an image of size 512 × 512 which we referred to as *Man*, and the second was the famous image *Lena* of size 512 × 512. Both images were noised with Gaussian noise using the Matlab function 'imnoise' with different variance values. Algorithm 5 (Projective) was implemented in Matlab code and it was chosen $\lambda = 1$ and $\rho_k = 1$ for all integer $k \geq 1$, in all tests. We solved problem (3.35) using the Conjugate Gradient (CG) method with tolerance 10^{-5}.

For comparison, we also report the results obtained with the split Bregman (SB) method [21], which is actually equivalent to ADMM [16]. As in [21] iterations were terminated when condition $\left\|u^k - u^{k-1}\right\| / \left\|u^k\right\| \leq 10^{-3}$ was met; since, this stopping criterion is satisfied faster while yielding good denoised images. However, we will report the residuals for the Kuhn-Tucker optimality conditions for problem (3.30) for both methods, i.e. we will plot the curves $\left\|\nabla u^k - v^k\right\|$ and $\left\|x^k - y^k\right\|$, where x^k and y^k are defined in (3.17) with $M = \nabla$ and $C = -I$. Moreover, we will refer to the vectors $\nabla u^k - v^k$ and $x^k - y^k$ as the primal and dual residuals, respectively.

Figures 3.1 and 3.3 show the performance of Projective and SB methods for denoising the images Man and Lena, respectively. Man was contaminated with Gaussian noise with variance $v = 0.01$ and Lena with variance $v = 0.03$. To solve the TV problems of the experiments which are presented in Figures 3.1 and 3.3, it was chosen $\zeta = 20$ and $\zeta = 40$, respectively. The number of iterations and total computation time (sec), required for the Projective and SB methods to reach the stopping criterion, are reported below each figure. We observe that in the experiments, the Projective method executed fewer iterations than SB method, and the Projective was faster.

In Figures 3.2 and 3.4 it is plotted the residual errors respective to the tests presented in Figures 3.1 and 3.3. We notice that the primal errors converges fast in the first iterations, but then slow down as the exact solution is reached. We also notice that after the first iterations primal and dual errors of Projective method are smaller than the respective errors of SB.

Figure 3.1: *Denoising of Man. (top left) Original image. (top right) Noise contaminated with variance 0.01. (bottom left) Denoised with Projective, the stopping criterion was satisfied at iteration 14 (2.313). (bottom right) Denoised with SB, the stopping criterion was satisfied at iteration 16 (2.540).*

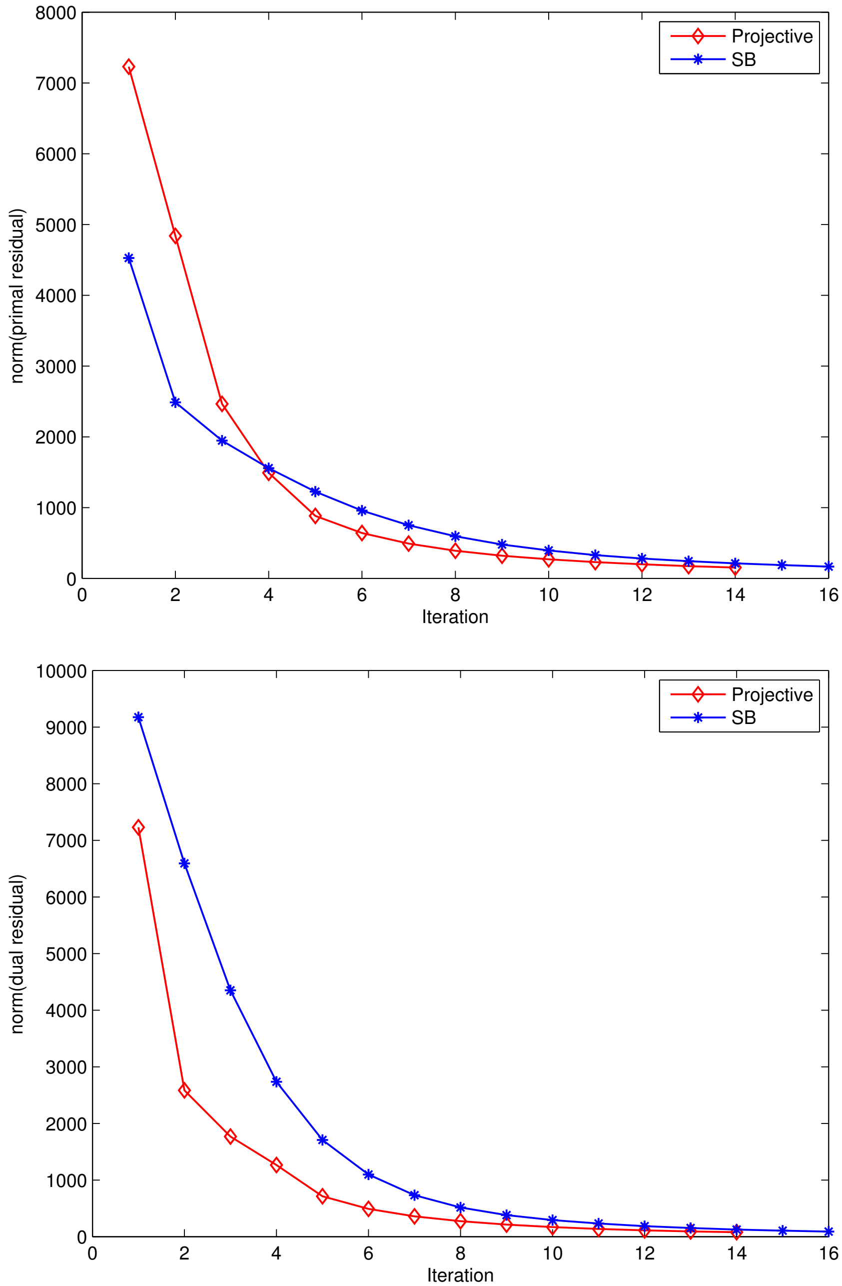

Figure 3.2: *Residual curves for Projective algorithm and SB method. (top) Primal error* $\left\|\nabla u^k - v^k\right\|$ *vs iteration number* k*. (bottom) Dual error* $\left\|x^k - y^k\right\|$ *vs iteration number* k*. Convergence results are for the tested image Man with Gaussian noise (variance* 0.01*) and with* $\zeta = 20$*.*

Figure 3.3: *Denoising of Lena. (top left) Original image. (top right) Noise contaminated with variance 0.03. (bottom left) Denoised with Projective, the stopping criterion was satisfied at iteration 17 (2.738). (bottom right) Denoised with SB, the stopping criterion was satisfied at iteration 19 (2.940).*

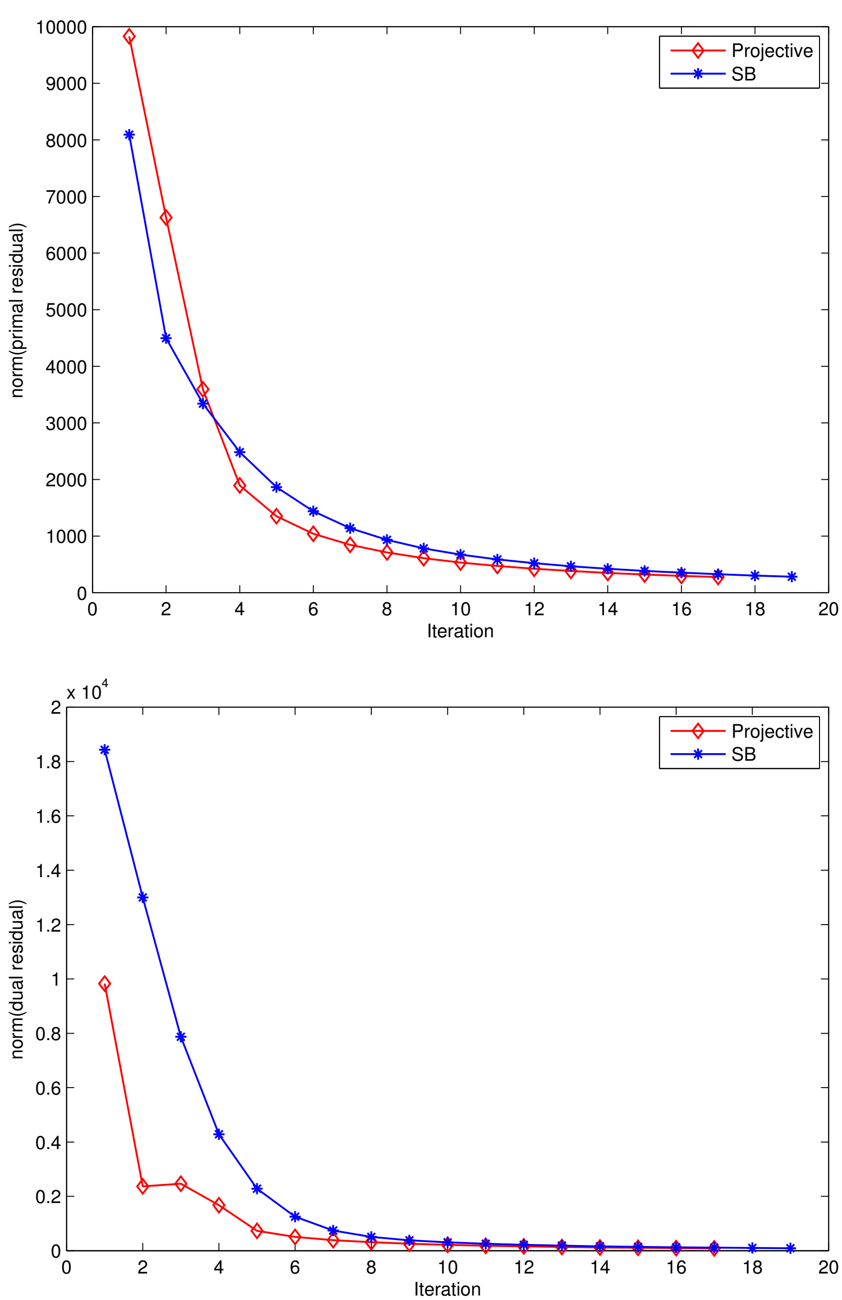

Figure 3.4: *Residual curves for Projective algorithm and SB method. (top) Primal error* $\|\nabla u^k - v^k\|$ *vs iteration number* k*. (bottom) Dual error* $\|x^k - y^k\|$ *vs iteration number* k*. Convergence results are for the tested image Lena with Gaussian noise (variance* 0.03*) and with* $\zeta = 40$*.*

We now show the total number of iteration executed by the CG method in each algorithm for some specific experiments. In the tests presented in Table 3.1, 3.2 and 3.3 both methods were stopped at iteration 20. We notice that the CG method in Projective algorithm performed fewer iterations than in SB method.

Image	Projective	SB
Man	113	119
Lena	109	115

Table 3.1: Total number of iteration of CG method. Images noised with Gaussian noise ($v = 0.01$). Resulst for $\zeta = 20$.

Image	Projective	SB
Man	120	125
Lena	118	121

Table 3.2: Total number of iteration of CG method. Images noised with Gaussian noise ($v = 0.03$). Resulst for $\zeta = 40$.

Image	Projective	SB
Man	125	127
Lena	120	125

Table 3.3: Total number of iteration of CG method. Images noised with Gaussian noise ($v = 0.05$). Resulst for $\zeta = 50$

However, the authors of [21] observed that the SB method attained optimal efficiency executing, at each iteration, just a single iteration of an iterative method to solve subproblem (3.35). Figures 3.3 and 3.6 below show that the Projective method also yields good denoised images performing one iteration of CG method.

These results suggest that the Projective framework is competitive when solving TV denoising models. To confirm this hypothesis more experiments are needed with a larger sample of images and different levels of noise. We notice that Algorithm 5 can also be applied to solve the isotropic TV model, i.e. the minimization problem

$$\min_{u\in\mathbb{R}^{m\times n}} \zeta \sum_i \sqrt{|\nabla_1 u|_i^2 + |\nabla_2 u|_i^2} + \frac{1}{2}\left\|u-b\right\|_F^2 .$$

Therefore, it might also be of interest to test the Projective algorithm with the above problem and compare its performance with algorithms available in the literature, such as SB method [21].

Clearly, Algorithm 3, 4 and 5 should be applied to a various kind of problems such as variational inequality, large-scale optimization problems arising in statistics, machine learning

Figure 3.5: *Denoising of Lena with one iteration of CG method per iteration. (left) Noise contaminated with variance 0.01. (center) Denoised with Projective, the stopping criterion was satisfied at iteration 14 (1.167). (right) Denoised with SB, the stopping criterion was satisfied at iteration 16 (1.190). Results for* $\zeta = 20$.

Figure 3.6: *Denoising of Man with one iteration of CG method per iteration. (left) Noise contaminated with variance 0.05. (center) Denoised with Projective, the stopping criterion was satisfied at iteration 24 (1.848). (right) Denoised with SB, the stopping criterion was satisfied at iteration 25 (1.760). Results for* $\zeta = 50$.

an related areas; and computational tests should be done to deduce the practical performance of these methods. These are interesting topics for a future work.

Bibliography

[1] Bauschke, H. H., and Borwein, J. M. On projection algorithms for solving convex feasibility problems. *SIAM Rev. 38*, 3 (1996), 367–426.

[2] Beck, A., and Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process. 18*, 11 (2009), 2419–2434.

[3] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning 3*, 1 (2011), 1–122.

[4] Brezis, H. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[5] Browder, F. E. Multi-valued monotone nonlinear mappings and duality mappings in Banach spaces. *Trans. Amer. Math. Soc. 118* (1965), 338–351.

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

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

[8] Chambolle, A. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision 20*, 1-2 (2004), 89–97. Special issue on mathematics and image analysis.

[9] Chambolle, A., Levine, S. E., and Lucier, B. J. An upwind finite-difference method for total variation-based image smoothing. *SIAM Journal on Imaging Sciences 4*, 1 (2011), 277–299.

[10] Combettes, P. L., and Pesquet, J.-C. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *Selected Topics in Signal Processing, IEEE Journal of 1*, 4 (2007), 564–574.

[11] Darbon, J., and Sigelle, M. A fast and exact algorithm for total variation minimization. In *Pattern recognition and image analysis*. Springer, 2005, pp. 351–359.

[12] Douglas, Jr., J., and Rachford, Jr., H. H. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc. 82* (1956), 421–439.

[13] Eckstein, J., and Bertsekas, D. P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming 55*, 3, Ser. A (1992), 293–318.

[14] Eckstein, J., and Svaiter, B. F. A family of projective splitting methods for the sum of two maximal monotone operators. *Math. Program. 111*, 1-2, Ser. B (2008), 173–199.

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

[16] Esser, E. Applications of lagrangian-based alternating direction methods and connections to split bregman. *CAM report 9* (2009), 31.

[17] Gabay, D. Chapter ix applications of the method of multipliers to variational inequalities. *Studies in mathematics and its applications 15* (1983), 299–331.

[18] Gabay, D., and Mercier, B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications 2*, 1 (1976), 17–40.

[19] Glowinski, R., and Le Tallec, P. *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, vol. 9 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.

[20] Glowinski, R., and Marrocco, A. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér. 9*, R-2 (1975), 41–76.

[21] Goldstein, T., and Osher, S. The split Bregman method for $L1$-regularized problems. *SIAM J. Imaging Sci. 2*, 2 (2009), 323–343.

[22] Hiriart-Urruty, J.-B., and Lemaréchal, C. *Convex analysis and minimization algorithms. II*, vol. 306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Advanced theory and bundle methods.

[23] Lions, P.-L., and Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal. 16*, 6 (1979), 964–979.

[24] Martinet, B. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle 4*, Ser. R-3 (1970), 154–158.

[25] Minty, G. J. Monotone networks. *Proc. Roy. Soc. London. Ser. A 257* (1960), 194–212.

[26] Minty, G. J. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J. 29* (1962), 341–346.

[27] Monteiro, R. D. C., and Svaiter, B. F. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim. 20*, 6 (2010), 2755–2787.

[28] Monteiro, R. D. C., and Svaiter, B. F. Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM J. Optim. 21*, 4 (2011), 1688–1720.

[29] Monteiro, R. D. C., and Svaiter, B. F. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim. 23*, 1 (2013), 475–507.

[30] Passty, G. B. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl. 72*, 2 (1979), 383–390.

[31] Peaceman, D. W., and Rachford, Jr., H. H. The numerical solution of parabolic and elliptic differential equations. *J. Soc. Indust. Appl. Math. 3* (1955), 28–41.

[32] Rockafellar, R. T. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

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

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

[35] Rudin, L. I., Osher, S., and Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D 60*, 1-4 (1992), 259–268. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

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

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

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

[39] Spingarn, J. E. Partial inverse of a monotone operator. *Appl. Math. Optim. 10*, 3 (1983), 247–265.

[40] Tseng, P. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim. 29*, 1 (1991), 119–138.

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

[42] Wang, Y., Yin, W., and Zhang, Y. Caam technical report tr07-10. *Department of Computational and Applied Mathematics* (2007).