Samara Costa Lima

On the Spingarn's partial inverse method: inexact versions, convergence rates and applications to operator splitting and optimization

Tese submetida ao Programa de Pós-Graduação em Matemática Pura e Aplicada da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Matemática Pura e Aplicada, com área de concentração em Matemática Aplicada.
Orientador: Prof. Dr. Maicon Marques Alves

Florianópolis, 2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Lima, Samara Costa
On the Spingarn's partial inverse method: inexact versions, convergence rates and applications to operator splitting and optimization / Samara Costa Lima ; orientador, Maicon Marques Alves, 2018.
79 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Programa de Pós-Graduação em Matemática Pura e Aplicada, Florianópolis, 2018.

Inclui referências.

1. Matemática Pura e Aplicada. 2. Otimizacão convexa, Matemática Aplicada. I. Marques Alves, Maicon. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Matemática Pura e Aplicada. III. Título.

Samara Costa Lima

On the Spingarn's partial inverse method: inexact versions, convergence rates and applications to operator splitting and optimization

Esta Tese foi julgada para a obtenção do Título de Doutor em Matemática Pura e Aplicada, área de Concentração em Matemática Aplicada, e aprovada em sua forma final pelo Curso de Pós-Graduação em Matemática Pura e Aplicada

---

Prof. Dr. **Ruy Coimbra Charão**
Coordenador do Curso de Pós-Graduação

Comissão Examinadora:

Prof. Dr. **Maicon Marques Alves**
Orientador: Universidade Federal de Santa Catarina - UFSC

Prof. Dr. **Orizon Pereira Ferreira**
Universidade Federal de Goiás - UFG

Prof. Dr. **Adriano De Cezaro**
Universidade Federal do Rio Grande - FURG

Prof. Dr. **Douglas Soares Gonçalves**
Universidade Federal de Santa Catarina - UFSC

Prof. Dr. **Licio Hernanes Bezerra**
Universidade Federal de Santa Catarina - UFSC

Florianópolis, 15 de fevereiro de 2018.

Este trabalho teve suposte da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES

Dedico à minha mãe e minhas irmãs.

Agradecimentos

Agradeço à minha família e meu namorado Valdir, pela compreensão e paciência, que permitiram a estabilidade necessária ao longo deste período.

Agradeço ao professor Maicon Marques Alves, pela orientação, paciência e dedicação que foram indispensáveis para concretização deste trabalho.

Agradeço aos amigos que me ajudaram de forma direta ou indireta para a conclusão deste trabalho.

Agradeço aos professores do departamento de matemática da UFSC , pelo conhecimento transmitido. Agradeço ainda, a Elisa pelas informaçoes prestadas.

Agradeço aos Professores Orizon Pereira Ferreira, Adriano De Cezaro, Douglas Soares Gonçalves, Licio Hernanes Bezerra por terem aceitado participar da banca avaliadora. Muito obrigada pelos comentários e sugestões.

Agradeço a CAPES pelo apoio financeiro.

"Ninguém baterá tão forte quanto a vida. Porém, não se trata de quão forte pode bater, se trata de quão forte pode ser atingido e continuar seguindo em frente. É assim que a vitória é conquistada."

Rocky Balboa.

Resumo

Neste trabalho, propomos e estudamos a complexidade computacional (em número de iterações) de uma versão inexata do método das inversas parciais de Spingarn. Os principais resultados de complexidade são obtidos através de uma análise do método proposto no contexto do *hybrid proximal extragradient* (HPE) *method* de Solodov e Svaiter, para o qual resultados de complexidade pontual e ergódica foram obtidos recentemente por Monteiro e Svaiter. Como aplicações, propomos e analisamos a complexidade computacional de um algoritmo inexato de decomposição – que generaliza o algoritmo de decomposição de Spingarn – e de um algoritmo paralelo do tipo *forward-backward* para otimização convexa com múltiplos termos na função objetivo. Além disso, mostramos que o algoritmo *scaled proximal decomposition on the graph of a maximal monotone operator* (SPDG), originalmente introduzido e estudado por Mahey, Oualibouch e Tao (1995), pode ser analisado através do formalismo das inversas parciais de Spingarn. Mais precisamente, mostramos que sob as hipóteses consideradas por Mahey, Oualibouch and Tao, a inversa parcial de Spingarn (do operador monótono maximal que define o problema em consideração) é um operador fortemente monótono, o que permite empregar resultados recentes sobre convergência e complexidade computational de métodos proximais para operadores fortemente monótonos. Ao fazer isso, obtemos adicionalmente uma convergência potencialmente mais rápida para o algorítmo SPDG e um limite superior mais preciso sobre o número de iterações necessárias para alcançar tolerâncias prescritas, especialmente para problemas mal-condicionados.

Palavras-Chave: Métodos de ponto proximal inexatos. Método das inversas parciais de Spingarn. Algoritmos de decomposição. Otimização convexa. Algorítmos do tipo *forward-backward*. Otimização paralela. Complexidade computacional. Algoritmo SPDG. Operadores fortemente monótonos. Taxas de convergência.

Resumo expandido

Introdução
Seja $\mathcal{H}$ um espaço de Hilbert com produto interno $\langle \cdot, \cdot \rangle$ e induzido pela norma $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

Um problema de inclusão monótona (MIP) consiste em encontrar $z \in \mathcal{H}$ tal que

$$0 \in T(z) \tag{1}$$

onde $T : \mathcal{H} \rightrightarrows \mathcal{H}$ é um operador ponto-conjunto monótono maximal. O problema (1) inclui diversos problemas em matemática aplicada e otimização. Um método clássico para resolução do problema (1) é o método de ponto proximal desenvolvido inicialmente por Martinet, no contexto de otimização, e posteriormente por Rockafellar neste contexto mais geral. Novas variantes inexatas deste método foram desenvolvidas por Solodov e Svaiter, e a complexidade computacional de uma destas variantes, os métodos híbridos proximais extragradiente (HPE), foi estabelecida em 2010 por Monteiro e Svaiter.

Em 1983, J. E. Spingarn propôs o problema de encontrar $x, u \in \mathcal{H}$ tal que

$$x \in V, \quad u \in V^{\perp} \text{ and } u \in T(x) \tag{2}$$

onde V é um subespaço vetorial fechado de $\mathcal{H}$, denotamos por $V^{\perp}$ o *complemento ortogonal* de V, e $T : \mathcal{H} \rightrightarrows \mathcal{H}$ é monótono maximal. Note que (2) se reduz a (1) quando $V = \mathcal{H}$. Além disso, (2) engloba problemas importantes na matemática aplicada, incluindo minimização de funções convexas sobre subespaço vetorial fechado e problemas de inclusão dado por soma finita de vários operadores monónotos maximais (ver, por exemplo, [1, 8, 9, 15]). Em [44], Spingarn também propôs um algoritmo para resolver (2), chamado método das inversas parciais, Spingarn provou que a análise de convergência desse método consiste em ver esse método como um caso especial do método de ponto proximal para resolver o problema de inclusão monótono (1), com T trocado por T_V, i.e.,

$$0 \in T_V(z) \tag{3}$$

onde T_V é a inversa parcial de T com respeito a V definida por Spingarn em [44].

O método das inversas parciais de Spingarn tem sido usado por diversos autores para análise de diferentes algoritmos práticos em otimização e aplicações relacionadas (veja, por exemplo, [1, 8, 9, 10, 25, 35]).

Objetivos

Os objetivos deste trabalho são propor e estudar a complexidade computacional (em número de iterações) de uma versão inexata do método das inversas parciais de Spingarn. Como aplicações, queremos propor e analisar a complexidade computacional de um algoritmo inexato de decomposição e de um algoritmo paralelo do tipo *forward-backward* para otimização convexa com múltiplos termos na função objetivo. Além disso, queremos mostrar que o algoritmo *scaled proximal decomposition on the graph of a maximal monotone operator* (SPDG), originalmente introduzido e estudado por Mahey, Oualibouch e Tao (1995), pode ser analisado através do formalismo das inversas parciais de Spingarn.

Metodologia

Os principais resultados de complexidade da versão inexata do método das inversas parciais de Spingarn são obtidos através de uma análise do método proposto no contexto do *hybrid proximal extragradient* (HPE) *method* de Solodov e Svaiter, para o qual resultados de complexidade pontual e ergódica foram obtidos recentemente por Monteiro e Svaiter. Mostramos ainda que sob as hipóteses consideradas por Mahey, Oualibouch and Tao, a inversa parcial de Spingarn (do operador monótono maximal que define o problema em consideração) é um operador fortemente monótono, o que permite empregar resultados recentes sobre convergência e complexidade computational de métodos proximais para operadores fortemente monótonos.

Resultados, discussão e Considerações finais

Propomos e estudamos a complexidade de uma versão inexata do método das inversas parciais de Spingarn. Como aplicações, propomos e analisamos a complexidade de um algoritmo inexato de decomposição e de um algoritmo paralelo do tipo *forward-backward* para otimização

convexa com múltiplos termos na função objetivo. Mostramos que o algoritmo SPDG pode ser analisado dentro do metodo das inversas parciais de Spingarn. Adicionalmente, obtemos uma convergência potencialmente mais rapida para o algoritmo SPDG e um limite superior mais preciso sobre o número de iterações necessárias para alcançar tolerâncias prescritas, especialmente para problemas mal-condicionados.

Palavras-Chave: Métodos de ponto proximal inexatos. Método das inversas parciais de Spingarn. Algoritmos de decomposição. Otimização convexa. Algorítmos do tipo *forward-backward*. Otimização paralela. Complexidade computacional. Algoritmo SPDG. Operadores fortemente monótonos. Taxas de convergência.

Abstract

In this work, we propose and study the iteration-complexity of an inexact version of the Spingarn's partial inverse method. Its complexity analysis is performed by viewing it in the framework of the hybrid proximal extragradient (HPE) method, for which pointwise and ergodic iteration-complexity has been established recently by Monteiro and Svaiter. As applications, we propose and analyze the iteration-complexity of an inexact operator splitting algorithm – which generalizes the original Spingarn's splitting method – and of a parallel forward-backward algorithm for multi-term composite convex optimization. We also show that the scaled proximal decomposition on the graph of a maximal monotone operator (SPDG) algorithm introduced and analyzed by Mahey, Oualibouch and Tao (1955) can be analyzed within the original Spingarn's partial inverse framework. We show that under the assumptions considered by Mahey, Oualibouch and Tao, the Spingarn's partial inverse of the underlying maximal monotone operator is strongly monotone, which allows one to employ recent results on the convergence and iteration-complexity of proximal point type methods for strongly monotone operators. By doing this, we additionally obtain a potentially faster convergence for the SPDG algorithm and a more accurate upper bound on the number of iterations needed to achieve prescribed tolerances, specially for ill-conditioned problems.

Keywords: Inexact proximal point methods. Spingarn's partial inverse method. Splitting algorithms. Convex optimization. Forward-backward type algorithms. Parallel methods. Iteration-complexity. SPDG algorithm. Strongly monotone operators. Rates of convergence.

Contents

Introduction

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. A set-valued map $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is a *monotone operator* if

$$\langle z - z', v - v' \rangle \geq 0 \qquad \forall v \in T(z), v' \in T(z'). \tag{4}$$

It is *maximal monotone* if it is monotone and its graph $G(T) := \{(z, v) \in \mathcal{H} \times \mathcal{H} \, : \, v \in T(z)\}$ is not properly contained in the graph of any other monotone operator on $\mathcal{H}$. Maximal monotone operators are (set-valued) nonlinear versions of (not necessarily symmetric) semipositive linear operators and they appear in different situations in nonlinear analysis and optimization [6, 39].

A *monotone inclusion problem* (MIP) consists of finding $z \in \mathcal{H}$ such that

$$0 \in T(z) \tag{5}$$

where $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone. Problem (5) includes many problems in applied mathematics and optimization as its special cases. For instance, if T is the subdifferential of a convex function f on $\mathcal{H}$, then (5) reduces to the problem of minimizing f over $\mathcal{H}$. Other important instances of (5) include saddle-point problems and variational inequalities for monotone operators (see, e.g., [14, 19, 21, 37, 40] and references therein).

One of the most popular algorithms for finding approximate solutions of (5) is the *proximal point* (PP) *method*. It consists of an iterative scheme, which can be described by the following iteration (which is well defined due to the Minty theorem [27])

$$z_k = (\lambda_k T + I)^{-1} z_{k-1} \qquad \forall k \geq 1, \tag{6}$$

where z_{k-1} and z_k are the current and new iterate, respectively, and $\lambda_k > 0$ is a stepsize parameter. Tracing back to the work of Martinet

[26], the PP method was popularized by the work of Rockafellar [36] and is widely used in nowadays research as a framework for the design and analysis of many practical algorithms (see, e.g., [4, 16, 20, 21, 24, 33, 40]).

In this thesis, we propose and analyze the iteration-complexity of an inexact version of a proximal point type method, proposed by Spingarn in [44], for solving a more general version of (5), which we discuss next.

In [44], J. E. Spingarn posed the problem of finding $x, u \in \mathcal{H}$ such that

$$x \in V, \quad u \in V^{\perp} \text{ and } u \in T(x) \tag{7}$$

where V is a closed vector subspace of $\mathcal{H}$, we denote by $V^{\perp}$ the *orthogonal complement* of V, and $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone. Note that (7) reduces to (5) when $V = \mathcal{H}$. Moreover, (7) encompasses important problems in applied mathematics, including minimization of convex functions over closed subspaces and inclusion problems given by the sum of finitely many maximal monotone operators (see, e.g., [1, 8, 9, 15] and references therein, and Chapter 2). In [44], Spingarn also proposed an algorithm for solving (7), called the *partial inverse method* (PIM), which can be described as follows, for all $k \geq 1$:

$$\begin{cases} \tilde{x}_k = (T + I)^{-1}(x_{k-1}), \\ u_k = x_{k-1} - \tilde{x}_k, \\ x_k = P_V(\tilde{x}_k) + P_{V^{\perp}}(u_k), \end{cases} \tag{8}$$

where P_V and $P_{V^{\perp}}$ stand for the orthogonal projection onto V and $V^{\perp}$, respectively. Note that when $V = \mathcal{H}$, we obtain $P_V = I$ and $P_{V^{\perp}} = 0$, which, in particular, gives that the Spingarn's partial inverse method (8) reduces to the PP method (6) (with $\lambda_k \equiv 1$). The original Spingarn's approach for analyzing the convergence of (8) consists of viewing it as a special instance of (6) (with $\lambda_k \equiv 1$) for solving the MIP induced by the *partial inverse* of T. The partial inverse of T with respect to V is the (maximal monotone) operator $T_V : \mathcal{H} \rightrightarrows \mathcal{H}$ whose graph is

$$G(T_V) := \{(z, v) \in \mathcal{H} \times \mathcal{H} \; : \; P_V(v) + P_{V^{\perp}}(z) \in T(P_V(z) + P_{V^{\perp}}(v))\}. \tag{9}$$

It follows from (9) that $0 \in T_V(z)$, if and only if $P_{V^{\perp}}(z) \in T(P_V(z))$, in which case $(x, u) := (P_V(z), P_{V^{\perp}}(z))$ is a solution of (7). Moreover, it follows directly from the above definition that $T_{\{0\}} = T^{-1}$ and $T_{\mathcal{H}} = T$, where T^{-1} is the usual (set-valued) inverse of T, i.e., $v \in T^{-1}(z)$ if and only if $z \in T(v)$.

As we mentioned earlier, Spingarn proved in [44] that (8) is a particular instance of (6) (with $\lambda_k \equiv 1$) for solving (5) with T replaced by T_V, i.e.,

$$0 \in T_V(z) \tag{10}$$

which, in turn, as we pointed out before, is equivalent to (7). This gives, in particular, that the partial inverse method (8) converges either under the assumption of exact computation of the resolvent $(T+I)^{-1}$ or under summable error criterion [36]. Since then, the Spingarn's partial inverse method has been used by many authors as a framework for the design and analysis of different practical algorithms in optimization and related applications (see, e.g., [1, 8, 9, 10, 25, 35]).

Since, in practical situations, the computation of the (exact) resolvent $(\lambda T + I)^{-1}$ may be numerically expensive, it follows that any attempt to implement either (6) or (8) depends on strategies to compute $(\lambda T + I)^{-1}$ only inexactly. This motivated Rockafellar [36] to propose and analyze an inexact version of the PP method (6) based on a *summable error criterion*. More precisely, at every iteration $k \geq 1$, if z_k is computed such that

$$\sum_{k=1}^{+\infty} \|z_k - (\lambda_k T + I)^{-1}(z_{k-1})\| < +\infty \tag{11}$$

and the sequence $\{\lambda_k\}$ is bounded away from zero, then $\{z_k\}$ converges (weakly) to a solution of (5) (if there exists at least one of them).

In the last two decades, as an alternative to (11), inexact versions of the Rockafellar's PP method (6) based on *relative error criterion* have deserved the attention of several researchers (see, e.g., [4, 7, 14, 20, 21, 23, 31, 40, 41, 42, 43]). This type of inexact PP method, which started with the pioneering work of Solodov and Svaiter [40], is widely used both in the theory and practice of numerical convex analysis and optimization (see, e.g., [7, 14, 20, 21, 23, 31, 33, 42, 43]). Among the above mentioned inexact versions of the PP method, which are based on relative error tolerance criterion, the hybrid proximal extragradient (HPE) method of Solodov and Svaiter [40] became very popular in recent years (see, e.g., [4, 7, 14, 20, 21, 31]), specially due to its robustness as a framework for the design and analysis of several algorithms for monotone inclusion, variational inequalities, saddle-point and convex optimization problems. The HPE method can be described according to the following iteration, for all $k \geq 1$,

$$\begin{cases} v_k \in T^{\varepsilon_k}(\tilde{z}_k), \\ \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2, \\ z_k = z_{k-1} - \lambda_k v_k, \end{cases} \tag{12}$$

where $\varepsilon_k \geq 0$, T^{ε_k} denotes the ε_k-enlargement of T (see (1.7) and Proposition 1.1.1), $\lambda_k > 0$ is a stepsize parameter and $\sigma \in [0, 1)$. The HPE method (12), which will be formally discussed in Section 1.2, is an inexact version of the Rockafellar's PP method (6) in which, at every iteration $k \geq 1$, errors – in the solution of the corresponding subproblems – are allowed within a relative error tolerance $\sigma \in [0, 1)$ (see the remarks after Algorithm 1).

In this thesis, we propose and analyze the iteration-complexity of an inexact version of the Spingarn's partial inverse method (8) in the light of recent developments in the iteration-complexity of the HPE method (12), which have been recently established by Monteiro and Svaiter [31]. We obtain $\mathcal{O}(1/\sqrt{k})$ pointwise and $\mathcal{O}(1/k)$ ergodic (global) convergence rates for finding approximate solutions of (7) by viewing the proposed method as a special instance of the HPE method for solving (10).

As we mentioned earlier, one of the most important special cases of (7) is the problem of solving MIPs given by the sum of finitely many maximal monotone operators:

$$0 \in \sum_{i=1}^{m} T_i(x) \tag{13}$$

where T_i is maximal monotone on $\mathcal{H}$, for each $i = 1, \ldots, m$. The most efficient algorithms for solving (13) belong to the family of so called *splitting methods* (see, e.g., [5, 10, 12, 17, 18, 44]). The attractive feature of this family of methods is that, at every iteration, each resolvent $(\lambda T_i + I)^{-1}$ can be computed individually, instead of the resolvent of the (sum) operator given in (13), which can be numerically expensive to evaluate. Motivated by this, Spingarn proposed in [44] a fully splitting algorithm for solving (13) for which, at each iteration, the solution of corresponding subproblems can be performed in *parallel*. This contrasts to other well-known splitting methods for solving (13), like, e.g., the Douglas-Rachford splitting method (in which case, $m = 2$), which demand a serial computation of the resolvents.

In this work, we apply the results obtained for the proposed inexact version of the Spingarn's partial inverse method to study and analyze the iteration-complexity of an inexact version of the Spingarn's operator splitting algorithm for solving (13). Moreover, applying these results for $T_i = \nabla f_i + \partial\varphi_i$, for all $i = 1, \ldots, m$, we obtain the iteration-complexity of a *parallel forward-backward splitting method* for solving the problem:

$$\min_{x\in\mathcal{H}} \sum_{i=1}^{m} (f_i + \varphi_i)(x), \tag{14}$$

where $m \geq 2$, for each $i = 1, \ldots, m$, f_i differentiable with Lipschitz continuous gradient, φ_i proper, convex and closed with an easily computable resolvent, and the solution set of (14) is nonempty.

As mentioned in [22], the performance of (8) is very sensitive to scaling factor variations. Motivated by this, Mahey, Oualibouch and Tao proposed in [25] the *scaled proximal decomposition on the graph of a maximal monotone operator* (SPDG) *algorithm* for solving (7), which can be described as follows, for all $k \geq 1$:

$$\begin{cases} \tilde{x}_k = (\gamma T + I)^{-1}(x'_{k-1}), \\ u_k = \gamma^{-1}(x'_{k-1} - \tilde{x}_k), \\ x'_k = P_V(\tilde{x}_k) + \gamma P_{V^\perp}(u_k). \end{cases} \tag{15}$$

Note that, if $\gamma = 1$, then it follows that the sequence $\{x'_k\}$ coincides with the sequence $\{x_k\}$ generated in (8), i.e., (15) is a scaled version of the Spingarn's partial inverse method. The convergence rate of the SPDG algorithm was previously analyzed in [25, Theorem 4.2] under the assumptions that the operator T in (7) is η-strongly monotone and L-Lipschitz continuous. Note that these assumptions imply, in particular, that the operator T is at most single-valued and $L \geq \eta$. The number $L/\eta \geq 1$ is known in the literature as the condition number of the problem (7). The influence of L/η as well as of scaling factors in the convergence speed of the SPDG algorithm, specially for solving ill-conditioned problems, was discussed in [25, Section 4] for the special case of quadratic programming.

Analogously to the latter reference, in this thesis, we analyze the convergence rate of the SPDG algorithm under the assumptions that the operator T in (7) is strongly monotone and Lipschitz continuous. We show that the algorithm falls in the framework of the PP method (6) for the (scaled) partial inverse of T, which, under the assumptions that the operator T in (7) is strongly monotone and Lipschitz continuous, is shown to be strongly monotone. This contrasts to the approach adopted in [25], which relies on fixed point techniques. By showing that the (scaled) partial inverse of T – with respect to V– is strongly monotone, we obtain a potentially faster convergence to the

SPDG algorithm when compared to the one proved in [25] by means of fixed point techniques. Moreover, the convergence rates obtained in this thesis allows one to measure the convergence speed of the SPDG algorithm on three different measures of approximate solution to the problem (7) (see Theorem 3.0.2 and the remarks right below it).

Summarizing, the contributions of this thesis are as follows:

- We propose and study the iteration-complexity of an inexact version of the Spingarn's partial inverse method. As applications, we propose and analyze the iteration-complexity of an inexact Spingarn's operator splitting algorithm and of a parallel forward-backward algorithm for multi-term composite convex optimization.

- We show that the SPDG algorithm can be analyzed within the Spingarn's partial inverse framework. In doing this, we obtain a potentially faster convergence for the SPDG algorithm and a more accurate upper bound on the number of iterations needed to achieve prescribed tolerances, specially for ill-conditioned problems.

This thesis resulted in the paper [2], which covers Chapter 2 and some results of the Chapter 1 , and in the preprint [3], which covers Chapter 3. The organization of this work is as follows:

Chapter 1 contains three sections. Section 1.1 reviews some important facts and definitions on convex analysis, maximal monotone operators, ε–enlargements and partial inverses of monotone operators, which are used throughout this thesis. Section 1.2 presents the HPE method and its iteration-complexity, and we also discuss a variant of HPE method from [12]. Section 1.3 presents the proximal point method for strongly monotone operators which will be useful in Chapter 3.

Chapter 2 contains three sections as follows. In Section 2.1, we propose an inexact Spingarn's partial inverse method for solving (7) and we analyze its iteration-complexity. In Section 2.2, we propose and analyze the iteration-complexity of an inexact Spingarn's operator splitting algorithm for solving (13). In Section 2.3, we show how the inexact operator splitting algorithm and its iteration-complexity can be used to derive a *parallel forward-backward splitting method* for multi-term composite convex optimization and to study its iteration-complexity.

In **Chapter 3**, we show that the SPDG algorithm can be analyzed by means of the original Spingarn's partial inverse framework.

Chapter 1

Preliminaries and basic results

In this chapter, we first review some important facts and definitions on convex analysis, maximal monotone operators, ε–enlargements and partial inverses of monotone operators, which will be used throughout this thesis. After, we present some results on the iteration-complexity and on a variant of the hybrid proximal extragradient (HPE) method. Lastly, we present the convergence rate of the proximal point (PP) algorithm for finding zeroes of strongly monotone inclusions. Appropriate literature on these topics can be found, e.g., in [2, 4, 6, 11, 12, 13, 15, 26, 27, 31, 34, 36, 38, 39, 40, 44] and the references therein.

1.1 Notation and basic results

This section collects the most important notations, definitions and some basic results that will be useful in this thesis.

In the following, we denote by $\mathcal{H}$ a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$. For $m \geq 2$, the Hilbert space $\mathcal{H}^m := \mathcal{H} \times \mathcal{H} \times \cdots \times \mathcal{H}$ will be endowed with the inner product

$$\langle (z_1, \ldots, z_m), (z'_1, \ldots, z'_m) \rangle := \sum_{i=1}^{m} \langle z_i, z'_i \rangle$$

and norm $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$.

Let $S : \mathcal{H} \rightrightarrows \mathcal{H}$ be a *set-valued* map. Its *graph* and *domain* are taken, respectively, as

$$\mathrm{G}(S) := \{(z, v) \in \mathcal{H} \times \mathcal{H} \,:\, v \in S(z)\} \tag{1.1}$$

and

$$\mathrm{D}(S) := \{z \in \mathcal{H} \,:\, S(z) \neq \emptyset\}. \tag{1.2}$$

The *inverse* of S is $S^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ such that $v \in S(z)$ if and only if $z \in S^{-1}(v)$. Given $S, S' : \mathcal{H} \rightrightarrows \mathcal{H}$ and $\lambda > 0$ we define $S + S' : \mathcal{H} \rightrightarrows \mathcal{H}$ and $\lambda S : \mathcal{H} \rightrightarrows \mathcal{H}$ by $(S + S')(z) = S(z) + S'(z)$ and $(\lambda S)(z) = \lambda S(z)$ for all $z \in \mathcal{H}$, respectively. Given set-valued maps $S_i : \mathcal{H} \rightrightarrows \mathcal{H}$, for $i = 1, \ldots, m$, we define its product by $S_1 \times S_2 \times \cdots \times S_m : \mathcal{H}^m \rightrightarrows \mathcal{H}^m$,

$$(z_1, z_2, \ldots, z_m) \mapsto S_1(z_1) \times S_2(z_2) \times \cdots \times S_m(z_m). \tag{1.3}$$

We denote by I the *identity operator* $z \mapsto z$.

Given an extended-real valued function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, we define its *domain* (or *effective domain*) and its *epigraph*, respectively, by

$$\mathrm{dom} f := \{z \in \mathcal{H} \,:\, f(z) < +\infty\} \tag{1.4}$$

and

$$\mathrm{epi} f := \{(z, \mu) \in \mathcal{H} \times \mathbb{R} \,:\, \mu \geq f(z)\}. \tag{1.5}$$

We say that f is:

(a) *proper*, if $\mathrm{dom} f \neq \emptyset$;

(b) *convex*, if for all $x, y \in \mathcal{H}$ and $\lambda \in [0, 1]$, one has

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y);$$

(c) *closed*, if its epigraph is a closed subset of $\mathcal{H} \times \mathbb{R}$.

The *subdifferential* of f is the set-valued map $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ defined at every $z \in \mathcal{H}$ as

$$\partial f(z) := \{v \in \mathcal{H} \,:\, f(z') \geq f(z) + \langle v, z' - z \rangle \;\; \forall z' \in \mathcal{H}\}.$$

Given $\varepsilon \geq 0$, the ε-*subdifferential* of $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is defined at every $z \in \mathcal{H}$ by

$$\partial_\varepsilon f(z) := \{v \in \mathcal{H} \,:\, f(z') \geq f(z) + \langle v, z' - z \rangle - \varepsilon \;\; \forall z' \in \mathcal{H}\}.$$

Note that $\partial_0 f = \partial f$.

The next two results, about ε-subdifferentials, will be needed in this thesis.

Lemma 1.1.1 ([33, Lemma 3.2]) *If $z, \tilde{z}, v \in \mathcal{H}$ are such that $v \in \partial f(z)$ and $f(\tilde{z}) < +\infty$, then $v \in \partial_\varepsilon f(\tilde{z})$ for every $\varepsilon \geq f(\tilde{z}) - f(z) - \langle v, \tilde{z} - z\rangle \geq 0$.*

Lemma 1.1.2 *The following hold for every $\lambda, \varepsilon > 0$:*

(a) $\partial(\lambda f) = \lambda \partial f$.

(b) $\partial_\varepsilon(\lambda f) = \lambda \partial_{\varepsilon/\lambda} f$.

The following result is well-known in the literature, its proof can be found in [34].

Lemma 1.1.3 *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable such that there exists a nonnegative constant L satisfying*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathcal{H}.$$

Then,

$$0 \leq f(x) - f(y) - \langle \nabla f(y), x - y\rangle \leq \frac{L}{2}\|x - y\|^2 \tag{1.6}$$

for all $x, y \in \mathcal{H}$.

As we mentioned in the Introduction, the theory of (set-valued) maximal monotone operators plays a central role in many areas of nonlinear analysis, optimization and applications.

Definition 1.1.1 *A set-valued map $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is* monotone *if*

$$\langle z - z', v - v'\rangle \geq 0 \quad \text{whenever} \quad (z, v), (z', v') \in G(T).$$

It is maximal monotone *if it is monotone and maximal in the following sense: if $S : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone and $G(T) \subset G(S)$, then $T = S$.*

For $\lambda > 0$, T is (maximal) monotone if and only if λT is (maximal) monotone. It is easy to prove that the subdifferential of any convex function is a monotone operator. In the case when f is proper, closed and convex, then ∂f is maximal monotone [38].

The *resolvent* of a monotone operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is $(T + I)^{-1}$. When T is maximal monotone, its resolvent is single-valued operator being defined everywhere on $\mathcal{H}$, which is a result due to Minty (see, e.g., [27]).

The concept of ε-enlargement of maximal monotone operators was introduced and studied in [13] (see also [11, 31]).

Definition 1.1.2 *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone and $\varepsilon \geq 0$, the ε-enlargement of T is the operator $T^{\varepsilon} : \mathcal{H} \rightrightarrows \mathcal{H}$ defined at any $z \in \mathcal{H}$ by*

$$T^{\varepsilon}(z) := \{v \in \mathcal{H} \;:\; \langle z - z', v - v' \rangle \geq -\varepsilon \;\; \forall (z', v') \in G(T)\}. \qquad (1.7)$$

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

The following summarizes some properties of T^{ε} which will be useful in this thesis.

Proposition 1.1.1 ([31, Proposition 2.1]) *Let $T, S : \mathcal{H} \rightrightarrows \mathcal{H}$ be set-valued maps. Then,*

(a) *if $\varepsilon \leq \varepsilon'$, then $T^{\varepsilon}(z) \subseteq T^{\varepsilon'}(z)$ for every $z \in \mathcal{H}$;*

(b) *$T^{\varepsilon}(z) + S^{\varepsilon'}(z) \subseteq (T+S)^{\varepsilon+\varepsilon'}(z)$ for every $z \in \mathcal{H}$ and $\varepsilon, \varepsilon' \geq 0$;*

(c) *T is monotone if, and only if, $T \subseteq T^0$;*

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

(e) *if $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and closed, then $\partial_{\varepsilon} f(z) \subseteq (\partial f)^{\varepsilon}(z)$ for any $\varepsilon \geq 0$ and $z \in \mathcal{H}$.*

Note that items (a) and (d) above imply that if $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone, then $T(z) \subseteq T^{\varepsilon}(z)$ for all $z \in \mathcal{H}$ and $\varepsilon \geq 0$, i.e., $T^{\varepsilon}(z)$ is indeed an enlargement of $T(z)$.

Next, we present the transportation formula for ε-enlargements. This formula will be used in the complexity analysis of ergodic iterates generated by the algorithms studied in this thesis.

Theorem 1.1.1 ([13, Theorem 2.3]) *Suppose $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone and let $x_\ell, u_\ell \in \mathcal{H}$, $\varepsilon_\ell, \alpha_\ell \in \mathbb{R}_+$, for $\ell = 1, \ldots, k$, be such that*

$$u_\ell \in T^{\varepsilon_\ell}(x_\ell), \quad \ell = 1, \ldots, k, \quad \sum_{\ell=1}^{k} \alpha_\ell = 1,$$

and define

$$x^a := \sum_{\ell=1}^{k} \alpha_\ell \, x_\ell \,, \quad u^a := \sum_{\ell=1}^{k} \alpha_\ell \, u_\ell \,,$$

$$\varepsilon^a := \sum_{\ell=1}^{k} \alpha_\ell \left[\varepsilon_\ell + \langle x_\ell - x^a, u_\ell - u^a \rangle\right].$$

Then, the following statements hold:

(a) $\varepsilon^a \geq 0$ *and* $u^a \in T^{\varepsilon^a}(x^a)$.

(b) *If, in addition,* $T = \partial f$ *for some proper, convex and closed function* f *and* $u_\ell \in \partial_{\varepsilon_\ell} f(x_\ell)$ *for* $\ell = 1, \dots, k$, *then* $u^a \in \partial_{\varepsilon^a} f(x^a)$.

The *Spingarn's partial inverse* [44] of a set-valued map $S : \mathcal{H} \rightrightarrows \mathcal{H}$ with respect to a closed subspace V of $\mathcal{H}$ is the set-valued operator $S_V : \mathcal{H} \rightrightarrows \mathcal{H}$ whose graph is

$$\mathrm{G}(S_V) := \{(z, v) \in \mathcal{H} \times \mathcal{H} \;:\; P_V(v) + P_{V^\perp}(z) \in S(P_V(z) + P_{V^\perp}(v))\}. \tag{1.8}$$

We denote by $V^\perp$ the *orthogonal complement* of V and by P_V and $P_{V^\perp}$ the *orthogonal projectors* onto V and $V^\perp$, respectively. Note that, as observed in the Introduction, $S_{\{0\}} = S^{-1}$ and $S_{\mathcal{H}} = S$. As was pointed out by Spingarn [44], if V is a proper subspace of $\mathcal{H}$, then S_V is "something in between" S^{-1} and S. This motivated Spingarn to choose the term "partial inverse". Moreover, the operator S_V is maximal monotone if and only if S is maximal monotone [44, Proposition 2.1].

The following characterization of the partial inverse of the ε–enlargement of a maximal monotone operator will be important for us.

Lemma 1.1.4 ([12, Lemma 3.1]) *Let* $T : \mathcal{H} \rightrightarrows \mathcal{H}$ *be a maximal monotone operator,* $V \subset \mathcal{H}$ *a closed subspace and* $\varepsilon > 0$. *Then,*

$$(T_V)^\varepsilon = (T^\varepsilon)_V.$$

1.2 On the hybrid proximal extragradient method

Consider the monotone inclusion problem (5), i.e.,

$$0 \in T(z) \tag{1.9}$$

where $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximal monotone operator for which the solution set $T^{-1}(0)$ of (1.9) is nonempty.

In this section, we formally state the *hybrid proximal extragradient* (HPE) method of Solodov and Svaiter [40] for solving (1.9), briefly discussed in (12), present its iteration-complexity, and also discuss a variant of it from [12] (see (1.15), Proposition 1.2.1 and Remarks 1.2.3 and 1.2.4).

In the last few years, starting with the paper [31], a lot of research has been done to study and analyze the *iteration-complexity* of the HPE method and its special instances, including Tseng's forward-backward splitting method, Korpelevich extragradient method, ADMM (see, e.g., [30, 31, 32]). To this end, the following notion of approximate solution to (1.9) was introduced in [31]: for given tolerances $\rho, \epsilon > 0$, find $\bar{z}, \bar{v} \in \mathcal{H}$ and $\bar{\varepsilon} > 0$ such that $(z, v) := (\bar{z}, \bar{v})$ and $\varepsilon := \bar{\varepsilon}$ satisfy

$$v \in T^{\varepsilon}(z), \quad \|v\| \leq \rho, \quad \varepsilon \leq \epsilon. \tag{1.10}$$

Using Proposition 1.1.1(d), we find that if $\rho = \epsilon = 0$ in (1.10) then $0 \in T(\bar{z})$, i.e., $\bar{z}$ is a solution of (1.9).

Next, we formally state the HPE method.

Algorithm 1 Hybrid proximal extragradient (HPE) method for (1.9)

(0) Let $z_0 \in \mathcal{H}$ and $\sigma \in [0, 1)$ be given and set $k = 1$.

(1) Compute $(\tilde{z}_k, v_k, \varepsilon_k) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ and $\lambda_k > 0$ such that

$$v_k \in T^{\varepsilon_k}(\tilde{z}_k), \quad \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2. \tag{1.11}$$

(2) Define

$$z_k = z_{k-1} - \lambda_k v_k, \tag{1.12}$$

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

Remarks.

(i) First, note that condition (1.11) relaxes both the inclusion and the equation in

$$v_k \in T(z_k), \quad \lambda_k v_k + z_k - z_{k-1} = 0, \tag{1.13}$$

which is clearly equivalent to the exact PP method (6), i.e., $z_k = (\lambda_k T + I)^{-1} z_{k-1}$. Here, $T^\varepsilon(\cdot)$ is the ε-enlargement of T; it has the properties that $T^0 = T$ and $T^\varepsilon(z) \supset T(z)$ (see Section 1.1 for details).

(ii) Instead of $\tilde{z}_k$, the next iterate z_k is defined in (1.12) as an extragradient step from z_{k-1}.

(iii) Letting $\sigma = 0$ and using Proposition 1.1.1(d), we conclude from (1.11) and (1.12) that (z_k, v_k) and $\lambda_k > 0$ satisfy (1.13), i.e., Algorithm 1 is an inexact version of the exact Rockafellar's PPM.

(iv) Algorithm 1 serves also as a framework for the analysis and development of several numerical schemes for solving concrete instances of (1.9) (see, e.g., [16, 28, 29, 30, 31, 32]); specific strategies for computing $(\tilde{z}_k, v_k, \varepsilon_k)$ and $\lambda_k > 0$ satisfying (1.11) depend on the particular instance of (1.9) under consideration.

Next, we summarize the main results from [31] about *pointwise* and *ergodic* iteration-complexity of the HPE method that we will need in this thesis. The *aggregate stepsize sequence* $\{\Lambda_k\}$ and the *ergodic sequences* $\{\tilde{z}_k^a\}$, $\{\tilde{v}_k^a\}$, $\{\varepsilon_k^a\}$ associated to $\{\lambda_k\}$ and $\{\tilde{z}_k\}$, $\{v_k\}$, and $\{\varepsilon_k\}$ are, respectively, for $k \geq 1$,

$$\begin{aligned}
\Lambda_k &:= \sum_{\ell=1}^{k} \lambda_\ell\,, \\
\tilde{z}_k^{\,a} &:= \frac{1}{\Lambda_k} \sum_{\ell=1}^{k} \lambda_\ell\, \tilde{z}_\ell, \quad v_k^{\,a} := \frac{1}{\Lambda_k} \sum_{\ell=1}^{k} \lambda_\ell\, v_\ell, \\
\varepsilon_k^{\,a} &:= \frac{1}{\Lambda_k} \sum_{\ell=1}^{k} \lambda_\ell (\varepsilon_\ell + \langle \tilde{z}_\ell - \tilde{z}_k^{\,a}, v_\ell - v_k^{\,a} \rangle).
\end{aligned} \tag{1.14}$$

Theorem 1.2.1 ([31, Theorem 4.4(a) and 4.7]) *Let* $\{\tilde{z}_k\}$, $\{v_k\}$, $\{\varepsilon_k\}$ *and* $\{\lambda_k\}$ *be generated by* Algorithm 1 *and let* $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ *and* $\{\varepsilon_k^{\,a}\}$ *be given in* (1.14). *Let also* d_0 *denote the distance of* z_0 *to* $T^{-1}(0) \neq \emptyset$ *and assume that* $\underline{\lambda} := \inf \lambda_k > 0$. *The following statements hold.*

(a) *For any* $k \geq 1$, *there exists* $i \in \{1, \ldots, k\}$ *such that*

$$v_i \in T^{\varepsilon_i}(\tilde{z}_i), \quad \|v_i\| \leq \frac{d_0}{\underline{\lambda}\sqrt{k}} \sqrt{\frac{1+\sigma}{1-\sigma}}, \quad \varepsilon_i \leq \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\underline{\lambda} k}\,;$$

(b) *for any* $k \geq 1$,

$$v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \quad \|v_k^a\| \leq \frac{2d_0}{\underline{\lambda} k}, \quad \varepsilon_k^a \leq \frac{2(1+\sigma/\sqrt{1-\sigma^2})d_0^2}{\underline{\lambda} k}.$$

Remark 1.2.1 The bounds given in (a) and (b) of Theorem 1.2.1 are called pointwise and ergodic bounds, respectively. Items (a) and (b) can be used, respectively, to prove that, for given tolerances $\rho, \epsilon > 0$, the termination criterion (1.10) is satisfied in at most

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{d_0^2}{\underline{\lambda}^2\rho^2}\right\rceil, \left\lceil \frac{d_0^2}{\underline{\lambda}\epsilon}\right\rceil\right\}\right) \text{ and } \mathcal{O}\left(\max\left\{\left\lceil \frac{d_0}{\underline{\lambda}\rho}\right\rceil, \left\lceil \frac{d_0^2}{\underline{\lambda}\epsilon}\right\rceil\right\}\right)$$

iterations.

The following variant of Algorithm 1 was studied in [12]: Let $z_0 \in \mathcal{H}$ and $\hat{\sigma} \in [0, 1)$ be given and iterate for $k \geq 1$,

$$\begin{cases} v_k \in T^{\varepsilon_k}(\tilde{z}_k), \\ \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \hat{\sigma}^2 \left(\|\tilde{z}_k - z_{k-1}\|^2 + \|\lambda_k v_k\|^2\right), \\ z_k = z_{k-1} - \lambda_k v_k. \end{cases} \tag{1.15}$$

Remark 1.2.2 The inequality in (1.15) is a relative error tolerance proposed in [43] (for a different method); the identity in (1.15) is the same extragradient step of Algorithm 1. Hence, the method described in (1.15) can be interpreted as a HPE variant in which a different relative error tolerance is considered in the solution of each subproblem.

The following result is similar to that in (23) in [43, Lemma 2].

Lemma 1.2.1 ([2, Lemma 2.5]) *Suppose* $\{z_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$ *and* $\{\lambda_k\}$ *satisfy the inequality in* (1.15). *Then, for every* $k \geq 1$,

$$\frac{1-\theta}{1-\hat{\sigma}^2}\|\tilde{z}_k - z_{k-1}\| \leq \|\lambda_k v_k\| \leq \frac{1+\theta}{1-\hat{\sigma}^2}\|\tilde{z}_k - z_{k-1}\|$$

where

$$\theta := \sqrt{1-(1-\hat{\sigma}^2)^2}. \tag{1.16}$$

Proof: From the inequality in (1.15) and the Cauchy-Schwarz inequality we obtain, for every $k \geq 1$,

$$\left(1-\hat{\sigma}^2\right)\|\lambda_k v_k\|^2 - 2\|\tilde{z}_k - z_{k-1}\|\|\lambda_k v_k\| + \left(1-\hat{\sigma}^2\right)\|\tilde{z}_k - z_{k-1}\|^2 \leq 0.$$

Defining

$$a := 1-\hat{\sigma}^2, \quad b := \|\tilde{z}_k - z_{k-1}\|, \quad c := ab^2, \quad t := \|\lambda_k v_k\|,$$

and using the above inequality we find $at^2 - 2bt + c \leq 0$ and, consequently,

$$\frac{2b - \sqrt{4(b^2 - ac)}}{2a} \leq t \leq \frac{2b + \sqrt{4(b^2 - ac)}}{2a},$$

which, in turn, combined with the latter definitions and (1.16) gives the desired result. ■

In what follows, we show that (1.15) is a special instance of Algorithm 1 whenever $\hat{\sigma} \in [0, 1/\sqrt{5}\,)$ and that it may fail to converge if we take $\hat{\sigma} > 1/\sqrt{5}$.

Proposition 1.2.1 ([2, Proposition 2.2]) *Let* $\{z_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$, $\{\varepsilon_k\}$ *and* $\{\lambda_k\}$ *be given in* (1.15) *and assume that* $\hat{\sigma} \in [0, 1/\sqrt{5}\,)$. *Define, for all* $k \geq 1$,

$$\sigma := \hat{\sigma}\sqrt{1 + \left(\frac{1+\theta}{1-\hat{\sigma}^2}\right)^2}, \tag{1.17}$$

where $0 \leq \theta < 1$ *is given in* (1.16). *Then,* $\sigma \geq 0$ *belongs to* $[0,1)$ *and* $z_k, \tilde{z}_k, v_k, \varepsilon_k$ *and* $\lambda_k > 0$ *satisfy* (1.11) *and* (1.12) *for all* $k \geq 1$. *As a consequence, the method of* [12] *defined in* (1.15) *is a special instance of* Algorithm 1 *whenever* $\hat{\sigma} \in [0, 1/\sqrt{5}\,)$.

Proof: The assumption $\hat{\sigma} \in [0, 1/\sqrt{5}\,)$, definition (1.17) and some simple calculations show that $\sigma \in [0,1)$. It follows from (1.15), (1.11) and (1.12) that to finish the proof of the proposition it suffices to prove the inequality in (1.11). To this end, note that from the second inequality in Lemma 1.2.1 and (1.17) we have

$$\begin{aligned}\hat{\sigma}^2\left(\|\tilde{z}_k - z_{k-1}\|^2 + \|\lambda_k v_k\|^2\right) &\leq \hat{\sigma}^2\left(1 + \left(\frac{1+\theta}{1-\hat{\sigma}^2}\right)^2\right)\|\tilde{z}_k - z_{k-1}\|^2 \\ &= \sigma^2\|\tilde{z}_k - z_{k-1}\|^2 \quad \forall k \geq 1,\end{aligned}$$

which in turn gives that the inequality in (1.11) follows from the one in (1.15). ■

Remark 1.2.3 ([2, Remark 2.4]) Algorithm 1 is obviously a special instance of (1.15) whenever $\sigma \in [0, 1/\sqrt{5}\,)$ by setting $\hat{\sigma} := \sigma$. Next we will show it is not true in general. Let $T : \mathbb{R} \to \mathbb{R}$ be the maximal monotone operator defined by

$$T(z) := z \quad \forall z \in \mathbb{R}. \tag{1.18}$$

Assume that $\sigma \in (\sqrt{2/5}, 1)$, take $z_0 = 1$ and define, for all $k \geq 1$,

$$\begin{aligned} \tilde{z}_k &:= z_k := \left(1 - \sigma^2\right) z_{k-1}, \quad v_k := z_{k-1}, \\ \varepsilon_k &:= \frac{\sigma^4}{2} |z_{k-1}|^2, \quad \lambda_k := \sigma^2. \end{aligned} \tag{1.19}$$

We will show that $(\tilde{z}_k, v_k, \varepsilon_k)$ and $\lambda_k > 0$ in (1.19) satisfy (1.11) but not (1.15) for any choice of $\hat{\sigma} \in [0, 1/\sqrt{5}\,)$. To this end, we first claim that $v_k \in T^{\varepsilon_k}(\tilde{z}_k)$ for all $k \geq 1$. Indeed, using (1.18) and (1.19) we obtain, for all $y \in \mathbb{R}$ and $k \geq 1$,

$$\begin{aligned} (Ty - v_k)(y - \tilde{z}_k) &= (y - z_{k-1})(y - z_{k-1} + \sigma^2 z_{k-1}) \\ &\geq |y - z_{k-1}|^2 - |\sigma^2 z_{k-1}||y - z_{k-1}| \\ &\geq -\frac{|\sigma^2 z_{k-1}|^2}{4} > -\varepsilon_k, \end{aligned}$$

which combined with (1.7) proves our claim. Moreover, it follows from (1.19) that

$$\begin{aligned} |\lambda_k v_k + \tilde{z}_k - z_{k-1}|^2 + 2\lambda_k \varepsilon_k &= |\tilde{z}_k - (1 - \sigma^2) z_{k-1}|^2 + 2\lambda_k \varepsilon_k \\ &= 2\lambda_k \varepsilon_k \\ &= \sigma^2 |\tilde{z}_k - z_{k-1}|^2, \end{aligned} \tag{1.20}$$

which proves that $(\tilde{z}_k, v_k, \varepsilon_k)$ and $\lambda_k > 0$ satisfy the inequality in (1.11). The first and second identities in (1.19) give that they also satisfy (1.12). Altogether, we have that the iteration defined in (1.19) is generated by Algorithm 1 for solving (1.9) with T given in (1.18). On the other hand, it follows from (1.19) and the assumption $\sigma > \sqrt{2/5}$ that

$$\begin{aligned} \sigma^2 |\tilde{z}_k - z_{k-1}|^2 &= \frac{\sigma^2}{2} \left(|\tilde{z}_k - z_{k-1}|^2 + |\lambda_k v_k|^2\right) \\ &> \frac{1}{5} \left(|\tilde{z}_k - z_{k-1}|^2 + |\lambda_k v_k|^2\right). \end{aligned} \tag{1.21}$$

Hence, it follows from (1.20) and (1.21) that the inequality in (1.15) can not be satisfied for any choice of $\hat{\sigma} \in [0, 1/\sqrt{5}\,)$ and so the sequence given is (1.19) is generated by Algorithm 1 but it is not generated by the algorithm described in (1.15).

Remark 1.2.4 ([2, Remark 2.5]) Next we present an example of a monotone inclusion problem for which an instance of (1.15) may fail to converge if we take $\hat{\sigma} \in (1/\sqrt{5}, 1)$. To this end, consider problem (1.9) where the maximal monotone operator $T : \mathbb{R} \to \mathbb{R}$ is defined by

$$T(z) := \alpha z \qquad \forall z \in \mathbb{R}, \tag{1.22}$$

where

$$\alpha := \frac{2\gamma}{\gamma - 2} + 1, \quad \gamma := \frac{1+\theta}{1-\hat{\sigma}^2}. \tag{1.23}$$

($\theta > 0$ is defined in (1.16).) Assuming $\hat{\sigma} \in (1/\sqrt{5}, 1)$ we obtain $5\hat{\sigma}^4 - 6\hat{\sigma}^2 + 1 < 0$, which is clearly equivalent to $\theta > |1 - 2\hat{\sigma}^2|$. Using (1.23) and the latter inequality we conclude that

$$\gamma > 2, \quad \frac{\alpha\gamma}{\alpha + \gamma} > 2. \tag{1.24}$$

Now take $z_0 = 1$ and define, for all $k \geq 1$,

$$\begin{aligned} &(\tilde{z}_k, v_k, \varepsilon_k) := \left(\frac{\gamma}{\alpha+\gamma} z_{k-1}, T(\tilde{z}_k), 0\right), \\ &\lambda_k := 1, \quad z_k := z_{k-1} - \lambda_k v_k. \end{aligned} \tag{1.25}$$

Direct calculation yields, for all $k \geq 1$,

$$|v_k + \tilde{z}_k - z_{k-1}|^2 = \hat{\sigma}^2 \left(|\tilde{z}_k - z_{k-1}|^2 + |v_k|^2\right), \tag{1.26}$$

which, in turn, together with (1.25) imply (1.15). Using (1.22) and (1.25) we find

$$z_k = \left(1 - \frac{\alpha\gamma}{\alpha+\gamma}\right)^k, \quad \forall k \geq 1. \tag{1.27}$$

Using the second inequality in (1.24) and the latter identity we easily conclude that $|z_k| \to +\infty$ as $k \to +\infty$ and so $\{z_k\}$ does not converge to the unique solution $\bar{z} = 0$ of (1.9). This example contrasts with [12, Theorem 4.1(i)], which claims that (1.15) converges for every $\hat{\sigma} \in [0, 1)$.

1.3 On the proximal point method for strongly monotone operators

In this section, we present the convergence rate of the proximal point (PP) algorithm for finding zeroes of strongly monotone inclusions. To this end, we consider the problem

$$0 \in A(z) \tag{1.28}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a μ-strongly maximal monotone operator, for some $\mu > 0$, i.e., A is maximal monotone and there exists $\mu > 0$ such that

$$\langle z - z', v - v' \rangle \geq \mu \|z - z'\|^2 \quad \forall v \in A(z),\ v' \in A(z'). \tag{1.29}$$

Motivated by the main results on the pointwise and ergodic iteration-complexity which have been obtained in [31], nonasymptotic convergence rates of the HPE method for solving strongly monotone inclusions were analyzed in [4].

In this section, we specialize the main result in latter reference, regarding the iteration-complexity of the HPE method for solving strongly monotone inclusions, to the exact PP method for solving (1.28). This is motivated by the fact that under certain conditions, the partial inverse operator T_V – of a maximal monotone operator T – is strongly monotone (see Proposition 3.0.1).

Next, we formally state the PP algorithm (6) with constant stepsize $\lambda > 0$ for solving the (strongly monotone) inclusion (1.28).

Algorithm 2 PPM for solving (1.28)

(0) Let $z_0 \in \mathcal{H}$ and $\lambda > 0$ be given and set $k = 1$.

(1) Compute

$$z_k = (\lambda A + I)^{-1} z_{k-1}. \tag{1.30}$$

(2) Let $k \leftarrow k + 1$ and go to step 1.

The following result establishes the linear convergence of the PP method under the strong monotonicity assumption on the operator A. Although it is a direct consequence of the more general result [4, Proposition 2.2], here we present a short proof for the convenience of the reader.

Proposition 1.3.1 *Assume that the operator A is μ-strongly maximal monotone, for some $\mu > 0$. Let $\{z_k\}$ be generated by* Algorithm 2 *and let $z^* \in \mathcal{H}$ is the (unique) solution of* (1.28)*, i.e., $z^* = A^{-1}(0)$. Then,* *for all* $k \geq 1$*,*

$$\|z_{k-1} - z_k\|^2 \leq \left(1 - \frac{2\lambda\mu}{1+2\lambda\mu}\right)^{k-1} \|z^* - z_0\|^2, \tag{1.31}$$

$$\|z^* - z_k\|^2 \leq \left(1 - \frac{2\lambda\mu}{1+2\lambda\mu}\right)^{k} \|z^* - z_0\|^2. \tag{1.32}$$

Proof: The desired results follow from the following

$$\begin{aligned}
\|z^* - z_{k-1}\|^2 - \|z^* - z_k\|^2 = & \\
&= \|z_{k-1} - z_k\|^2 + 2\lambda\langle \lambda^{-1}(z_{k-1} - z_k), z_k - z^*\rangle \\
&\geq \|z_{k-1} - z_k\|^2 + 2\lambda\mu\|z_k - z^*\|^2 \\
&\geq \frac{2\lambda\mu}{1+2\lambda\mu}\|z^* - z_{k-1}\|^2,
\end{aligned}$$

where we have used that $\lambda^{-1}(z_{k-1} - z_k) \in A(z_k)$, from (1.30), $0 \in A(z^*)$, A is μ-strongly monotone and $\min\{r^2 + as^2 \mid r, s \geq 0, r + s \geq b \geq 0\} = a(a+1)^{-1}b^2$. ■

Chapter 2

An inexact Spingarn's partial inverse method and its applications

In this chapter, we propose and analyze the iteration-complexity of an inexact version of the Spingarn's partial inverse method and its applications to operator splitting and composite convex optimization. As we mentioned in the Introduction, the main results are obtained by viewing the proposed methods within the framework of the HPE method of Solodov and Svaiter [31, 40]. We also introduce a notion of approximate solution to the Spingarn's problem (7) and, as a by-product, we obtain the iteration-complexity of the (exact) Spingarn's partial inverse method (which, up to our knowledge, was not known so far). As applications, we obtain an inexact operator splitting algorithm for solving (13), which generalizes the Spingarn's operator splitting algorithm [44, Chapter 5], and a parallel forward-backward method for solving multi-term composite convex optimization problems.

This chapter is organized as follows. Section 2.1 presents our inexact Spingarn's partial inverse method and its iteration-complexity analysis. In Section 2.2, we present and study an inexact version of the Spingar's operator splitting algorithm. In Section 2.3, we present our parallel forward-backward algorithm for solving multi-term composite convex optimization problems.

2.1 An inexact Spingarn's partial inverse method

In this section, we consider problem (7), i.e., the problem of finding $x, u \in \mathcal{H}$ such that

$$x \in V, \quad u \in V^{\perp} \text{ and } u \in T(x) \tag{2.1}$$

where $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone and V is a closed subspace of $\mathcal{H}$. The solution set of (2.1) is defined by

$$\mathcal{S}^*(V,T) := \{z \in \mathcal{H} \ : \ \exists\, x, u \in \mathcal{H} \text{ satisfying (2.1) s.t. } z = x+u\}, \tag{2.2}$$

which we assume to be nonempty.

One of the main goals of this section is to present and study the iteration-complexity of an inexact version of the Spingarn's partial inverse method (8). Regarding the results on iteration-complexity, we consider the following notion of approximate solution for (2.1): given tolerances $\rho, \epsilon > 0$, find $\bar{x}, \bar{u} \in \mathcal{H}$ and $\bar{\varepsilon} \geq 0$ such that $(x, u) := (\bar{x}, \bar{u})$ and $\varepsilon := \bar{\varepsilon}$ satisfy

$$u \in T^{\varepsilon}(x), \quad \max\{\|x - P_V(x)\|, \|u - P_{V^\perp}(u)\|\} \leq \rho, \quad \varepsilon \leq \epsilon, \tag{2.3}$$

where P_V and $P_{V^\perp}$ stand for the orthogonal projection onto V and $V^{\perp}$, respectively, and $T^{\varepsilon}(\cdot)$ denotes the ε-enlargement of T (see Section 1.2 for more details on notation). For $\rho = \epsilon = 0$, criterion (2.3) gives $\bar{x} \in V$, $\bar{u} \in V^{\perp}$ and $\bar{u} \in T(\bar{x})$, i.e., in this case $\bar{x}, \bar{u}$ satisfy (2.1). Moreover, if $V = \mathcal{H}$ in (2.1), in which case $P_V = I$ and $P_{V^\perp} = 0$, then the criterion (2.3) coincides with one discussed in Section 1.2 for problem (1.9) (see (1.10)).

We next present our inexact version of the Spingarn's partial inverse method.

Algorithm 3 An inexact Spingarn's partial inverse method for (2.1) (I)

(0) Let $x_0 \in \mathcal{H}$ and $\sigma \in [0, 1)$ be given and set $k = 1$.

(1) Compute $(\tilde{x}_k, u_k, \varepsilon_k) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ such that

$$\begin{aligned} & u_k \in T^{\varepsilon_k}(\tilde{x}_k), \\ & \|u_k + \tilde{x}_k - x_{k-1}\|^2 + 2\varepsilon_k \leq \sigma^2 \|P_V(\tilde{x}_k) + P_{V^\perp}(u_k) - x_{k-1}\|^2. \end{aligned} \tag{2.4}$$

(2) Define

$$x_k = x_{k-1} - [P_V(u_k) + P_{V^\perp}(\tilde{x}_k)], \tag{2.5}$$

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

Remarks.

(i) Letting $V = \mathcal{H}$ in (2.1), in which case $P_V = I$ and $P_{V^\perp} = 0$, we obtain that Algorithm 3 coincides with Algorithm 1 with $\lambda_k = 1$ for all $k \geq 1$ for solving (1.9) (or, equivalently, (2.1) with $V = \mathcal{H}$).

(ii) An inexact partial inverse method called *sPIM(ε)* was proposed in [12], Section 4.2, for solving (2.1). The latter method, with a different notation and scaling factor $\eta = 1$, is given according to the iteration:

$$\begin{cases} u_k \in T^{\varepsilon_k}(\tilde{x}_k), \\ \|u_k + \tilde{x}_k - x_{k-1}\|^2 + 2\varepsilon_k \leq \\ \hat{\sigma}^2 \left(\|\tilde{x}_k - P_V(x_{k-1})\|^2 + \|u_k - P_{V^\perp}(x_{k-1})\|^2\right), \\ x_k = x_{k-1} - [P_V(u_k) + P_{V^\perp}(\tilde{x}_k)], \end{cases} \tag{2.6}$$

where $\hat{\sigma} \in [0, 1)$. The convergence analysis given in [12] for the iteration (2.6) relies on the fact (proved in the latter reference) that (2.6) is a special instance of (1.15) (which we observed in Remark 1.2.4 may fail to converge if we consider $\hat{\sigma} \in (1/\sqrt{5}, 1)$). Using the fact just mentioned, the last statement in Proposition 1.2.1 and Proposition 2.1.1 we conclude that (2.6) is a special instance of Algorithm 3 whenever $\hat{\sigma} \in [0, 1/\sqrt{5})$ and it may fail to converge if $\hat{\sigma} > 1/\sqrt{5}$. On the other hand, since, due to Proposition 2.1.1, Algorithm 3 is a special instance of Algorithm 1, it converges for all $\sigma \in [0, 1)$ (see, e.g., [40, Theorem 3.1]). Note that the difference between sPIM(ε) and Algorithm 3 is the inequality in (2.4) and (2.6).

(iii) An inexact version of the Spingarn's partial inverse method for solving (2.1) was also proposed and analyzed in [1]. With a different notation, the latter method (see Theorem 2.4 in [1]) can be given according the iteration:

$$\begin{cases} p_k = (T+I)^{-1}(x_k + u_k) + e_k, \quad r_k = x_k + u_k - p_k, \\ x_{k+1} = x_k - \lambda_k P_V(r_k), \quad u_{k+1} = u_k - \lambda_k P_{V^\perp}(p_k), \end{cases} \tag{2.7}$$

where $\lambda_k \in (0, 2)$, for all $k \geq 1$, $\sum_{k=1}^{+\infty} \lambda_k(2 - \lambda_k) = +\infty$ and $\sum_{k=1}^{+\infty} \lambda_k \|e_k\| < +\infty$. Weak convergence of the iteration (2.7) was obtained in [1] by viewing it in the framework of a variant of a proximal point method previously studied in [15]. We mention that, in part, the methods and results in [1] differ from the corresponding ones obtained in this thesis in that their analysis relies on asymptotic convergence and on a different error criterion.

In what follows, we will prove iteration-complexity results for Algorithm 3 to obtain approximate solutions of (2.1), according to (2.3), as a consequence of the iteration-complexity results from Theorem 1.2.1. To this end, first let $\{\tilde{x}_k\}$, $\{u_k\}$ and $\{\varepsilon_k\}$ be generated by Algorithm 3 and define the *ergodic* sequences associated to them:

$$\begin{aligned} \tilde{x}_k^a &:= \frac{1}{k}\sum_{\ell=1}^{k} \tilde{x}_\ell\,, \quad u_k^a := \frac{1}{k}\sum_{\ell=1}^{k} u_\ell\,, \\ \varepsilon_k^a &:= \frac{1}{k}\sum_{\ell=1}^{k} \left[\varepsilon_\ell + \langle \tilde{x}_\ell - \tilde{x}_k^a, u_\ell - u_k^a\rangle\right]. \end{aligned} \tag{2.8}$$

The approach adopted in the current section for solving (2.1) follows the Spingarn's approach [44], which consists of solving the monotone inclusion

$$0 \in T_V(z) \tag{2.9}$$

where the maximal monotone operator $T_V : \mathcal{H} \rightrightarrows \mathcal{H}$ is the partial inverse of T with respect to the subspace V. In view of (1.8), we have

$$(T_V)^{-1}(0) = \mathcal{S}^*(V, T), \tag{2.10}$$

where the latter set is defined in (2.2). Hence, problem (2.1) is equivalent to the monotone inclusion problem (2.9).

Next proposition shows that Algorithm 3 can be regarded as a special instance of Algorithm 1 for solving (2.9).

Proposition 2.1.1 ([2, Proposition 3.3]) *Let the sequences* $\{\tilde{x}_k\}_{k\geq 1}$, $\{u_k\}_{k\geq 1}$, $\{\varepsilon_k\}_{k\geq 1}$ *and* $\{x_k\}_{k\geq 0}$ *be generated by* Algorithm 3. *Define* $z_0 = x_0$ *and, for all* $k \geq 1$,

$$z_k = x_k, \quad \tilde{z}_k = P_V(\tilde{x}_k) + P_{V^\perp}(u_k), \quad v_k = P_V(u_k) + P_{V^\perp}(\tilde{x}_k). \tag{2.11}$$

Then, for all $k \geq 1$,

$$\begin{aligned} & v_k \in (T_V)^{\varepsilon_k}(\tilde{z}_k), \\ & \|v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2, \\ & z_k = z_{k-1} - v_k, \end{aligned} \tag{2.12}$$

i.e., $(\tilde{z}_k, v_k, \varepsilon_k)$ *and* $\lambda_k \equiv 1$ *satisfy* (1.11) *and* (1.12) *for all* $k \geq 1$. *As a consequence, the sequences* $\{z_k\}_{k\geq 0}$, $\{\tilde{z}_k\}_{k\geq 1}$, $\{v_k\}_{k\geq 1}$ *and* $\{\varepsilon_k\}_{k\geq 1}$ *are generated by* Algorithm 1 (*with* $\lambda_k \equiv 1$) *for solving* (2.9).

Proof: From the inclusion in (2.4), (1.8) with $S = T^{\varepsilon_k}$ and Lemma 1.1.4 we have $P_V(u_k) + P_{V^\perp}(\tilde{x}_k) \in (T_V)^{\varepsilon_k}(P_V(\tilde{x}_k) + P_{V^\perp}(u_k))$ for all $k \geq 1$, which combined with the definitions of $\tilde{z}_k$ and v_k in (2.11) gives the inclusion in (2.12). Direct use of (2.11) and the definition of $\{z_k\}$ yield

$$\begin{aligned} & v_k + \tilde{z}_k + z_{k-1} = u_k + \tilde{x}_k - x_{k-1}, \\ & \tilde{z}_k - z_{k-1} = P_V(\tilde{x}_k) + P_{V^\perp}(u_k) - x_{k-1}, \\ & z_{k-1} - v_k = x_{k-1} - [P_V(u_k) - P_{V^\perp}(\tilde{x}_k)], \end{aligned} \tag{2.13}$$

which combined with (2.4), (2.5) and the definition of $\{z_k\}$ gives the remaining statements in (2.12). The last statement of the proposition follows from (2.12) and Algorithm 1's definition. ■

The following theorem establishes global pointwise and ergodic convergence rates for Algorithm 3, which will be used to prove the iteration-complexity of the latter algorithm, as well as of its special instances.

Theorem 2.1.1 ([2, Theorem 3.1]) *Let* $\{\tilde{x}_k\}$, $\{u_k\}$ *and* $\{\varepsilon_k\}$ *be generated by* Algorithm 3 *and let* $\{\tilde{x}_k^a\}$, $\{u_k^a\}$ *and* $\{\varepsilon_k^a\}$ *be defined in* (2.8). *Let also* $d_{0,V}$ *denote the distance of* x_0 *to the solution set* (2.2). *The following statements hold:*

(a) *For any* $k \geq 1$, *there exists* $j \in \{1, \dots, k\}$ *such that*

$$u_j \in T^{\varepsilon_j}(\tilde{x}_j),$$
$$\sqrt{\|\tilde{x}_j - P_V(\tilde{x}_j)\|^2 + \|u_j - P_{V^\perp}(u_j)\|^2} \leq \frac{d_{0,V}}{\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}}, \tag{2.14}$$
$$\varepsilon_j \leq \frac{\sigma^2 d_{0,V}^2}{2(1-\sigma^2)k}\,;$$

(b) *for any* $k \geq 1$,

$$u_k^a \in T^{\varepsilon_k^a}(\tilde{x}_k^a),$$
$$\sqrt{\|\tilde{x}_k^a - P_V(\tilde{x}_k^a)\|^2 + \|u_k^a - P_{V^\perp}(u_k^a)\|^2} \leq \frac{2d_{0,V}}{k}, \tag{2.15}$$
$$0 \leq \varepsilon_k^a \leq \frac{2(1+\sigma/\sqrt{1-\sigma^2})d_{0,V}^2}{k}\,.$$

Proof: From (2.11), we obtain

$$\tilde{x}_k = P_V(\tilde{z}_k) + P_{V^\perp}(v_k), \quad u_k = P_V(v_k) + P_{V^\perp}(\tilde{z}_k) \quad \forall k \geq 1. \tag{2.16}$$

Direct substitution of the latter identities in $\tilde{x}_k^a$ and u_k^a in (2.8) yields

$$\tilde{x}_k^a = P_V(\tilde{z}_k^a) + P_{V^\perp}(v_k^a), \quad u_k^a = P_V(v_k^a) + P_{V^\perp}(\tilde{z}_k^a) \quad \forall k \geq 1. \tag{2.17}$$

Using (2.16) and (2.17) in the definition of ε_k^a in (2.8) and the fact that the operators P_V and $P_{V^\perp}$ are self-adjoint and idempotent we find

$$\varepsilon_k^a = \frac{1}{\Lambda_k}\sum_{\ell=1}^{k}\lambda_\ell(\varepsilon_\ell + \langle \tilde{z}_\ell - \tilde{z}_k^a, v_\ell - v_k^a\rangle) \quad \forall k \geq 1, \tag{2.18}$$

where $\{\varepsilon_k^a\}$ is defined in (2.8). Now consider the ergodic sequences $\{\Lambda_k\}$, $\{\tilde{z}_k^a\}$ and $\{v_k^a\}$ defined in (1.14) with $\lambda_k := 1$ for all $k \geq 1$. Let d_0 denote the distance of $z_0 = x_0$ to the solution set $(T_V)^{-1}(0)$ of (2.9) and note that $d_0 = d_{0,V}$ in view of (2.10). Based on the above considerations one can use the last statement in Proposition 2.1.1 and Theorem 1.2.1 with $\underline{\lambda} := 1$ to conclude that for any $k \geq 1$ there exists $j \in \{1, \dots, k\}$ such that

$$v_j \in (T_V)^{\varepsilon_j}(\tilde{z}_j),$$
$$\|v_j\| \leq \frac{d_{0,V}}{\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}}, \tag{2.19}$$
$$\varepsilon_j \leq \frac{\sigma^2 d_{0,V}^2}{2(1-\sigma^2)k}\,,$$

and

$$\begin{aligned} v_k^a &\in (T_V)^{\varepsilon_k^a}(\tilde{z}_k^a), \\ \|v_k^a\| &\le \frac{2d_{0,V}}{k}, \\ \varepsilon_k^a &\le \frac{2(1+\sigma/\sqrt{1-\sigma^2})d_{0,V}^2}{k}, \end{aligned} \tag{2.20}$$

where $\{\varepsilon_k^a\}$ is given in (2.8). Using Lemma 1.1.4, the definition in (1.8) (for $S = T^{\varepsilon_k}$), (2.16) and (2.17) we conclude that the equivalence $v \in (T_V)^\varepsilon(\tilde{z}) \iff v \in (T^\varepsilon)_V(\tilde{z}) \iff u \in T^\varepsilon(\tilde{x})$ holds for $(\tilde{z}, v, \varepsilon) = (\tilde{z}_k, v_k, \varepsilon_k)$ and $(\tilde{x}, u, \varepsilon) = (\tilde{x}_k, u_k, \varepsilon_k)$, and $(\tilde{z}, v, \varepsilon) = (\tilde{z}_k^a, v_k^a, \varepsilon_k^a)$ and $(\tilde{x}, u, \varepsilon) = (\tilde{x}_k^a, u_k^a, \varepsilon_k^a)$, for all $k \ge 1$. As a consequence, the inclusions in (2.14) and (2.15) follow from the ones in (2.19) and (2.20), respectively. Since (2.17) gives $v_k^a = P_V(u_k^a) + P_{V^\perp}(\tilde{x}_k^a)$ for all $k \ge 1$, it follows from the definition of $\{v_k\}$ in (2.11) that $(v, u, \tilde{x}) = (v_k, u_k, \tilde{x}_k)$ and $(v, u, \tilde{x}) = (v_k^a, u_k^a, \tilde{x}_k^a)$ satisfy

$$\|v\|^2 = \|P_V(u)\|^2 + \|P_{V^\perp}(\tilde{x})\|^2 = \|u - P_{V^\perp}(u)\|^2 + \|\tilde{x} - P_V(\tilde{x})\|^2$$

for all $k \ge 1$, which, in turn, gives that the inequalities in (2.14) and (2.15) follow from the ones in (2.19) and (2.20), respectively. This concludes the proof. ■

Next result, which is a direct consequence of Theorem 2.1.1(b), gives the iteration-complexity of Algorithm 3 to find $x, u \in \mathcal{H}$ and $\varepsilon \ge 0$ satisfying the termination criterion (2.3).

Theorem 2.1.2 [2, Theorem 3.2](Iteration-complexity) *Let $d_{0,V}$ denote the distance of x_0 to the solution set* (2.2) *and let $\rho, \epsilon > 0$ be given tolerances. Then,* Algorithm 3 *finds $x, u \in \mathcal{H}$ and $\varepsilon \ge 0$ satisfying the termination criterion* (2.3) *in at most*

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{d_{0,V}}{\rho} \right\rceil, \left\lceil \frac{d_{0,V}^2}{\epsilon} \right\rceil\right\}\right) \tag{2.21}$$

iterations.

We now consider a special instance of Algorithm 3 which will be used in Section 2.2 to derive operator splitting methods for solving the problem of finding zeroes of a sum of finitely many maximal monotone operators.

Algorithm 4 An inexact Spingarn's partial inverse method for (2.1) (II)

(0) Let $x_0 \in \mathcal{H}$ and $\sigma \in [0, 1)$ be given and set $k = 1$.

(1) Compute $\tilde{x}_k \in \mathcal{H}$ and $\varepsilon_k \geq 0$ such that

$$u_k := x_{k-1} - \tilde{x}_k \in T^{\varepsilon_k}(\tilde{x}_k), \quad \varepsilon_k \leq \frac{\sigma^2}{2}\|\tilde{x}_k - P_V(x_{k-1})\|^2. \tag{2.22}$$

(2) Define

$$x_k = P_V(\tilde{x}_k) + P_{V^\perp}(u_k), \tag{2.23}$$

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

Remarks.

(i) Letting $\sigma = 0$ in Algorithm 4 and using Proposition 1.1.1(d) we obtain from (2.22) that $x = \tilde{x}_k$ solves the inclusion $0 \in T(x) + x - x_{k-1}$, i.e., $\tilde{x}_k = (T + I)^{-1}x_{k-1}$ for all $k \geq 1$. In other words, if $\sigma = 0$, then Algorithm 4 is the Spingarn's partial inverse method originally presented in [44].

(ii) It follows from Proposition 1.1.1(e) that Algorithm 4 is a generalization to the general setting of inclusions with monotone operators of the *Epsilon-proximal decomposition method scheme (EPDMS)* proposed and studied in [35] for solving convex optimization problems. Indeed, using the identity in (2.22) and some direct computations we find

$$\begin{aligned}\|\tilde{x}_k - P_V(x_{k-1})\|^2 &= \|P_{V^\perp}(\tilde{x}_k) - P_V(u_k)\|^2 \\ &= \|P_{V^\perp}(\tilde{x}_k)\|^2 + \|P_V(u_k)\|^2,\end{aligned}$$

which gives that the right hand side of the inequality in (2.22) is equal to $\sigma^2/2\left(\|P_{V^\perp}(\tilde{x}_k)\|^2 + \|P_V(u_k)\|^2\right)$ (cf. EPDMS method in [35], with a different notation). We also mention that no iteration-complexity analysis was performed in [35].

(iii) Likewise, letting $V = \mathcal{H}$ in Algorithm 4 and using Proposition 1.1.1(e) we obtain that Algorithm 4 generalizes the *IPP-CO fra-*

mework of [33] (with $\lambda_k := 1$ for all $k \geq 1$), for which iteration-complexity analysis was presented in the latter reference, to the more general setting of inclusions problems with monotone operators.

Proposition 2.1.2 ([2, Proposition 3.1]) *The following statements hold true.*

(a) Algorithm 4 *is a special instance of* Algorithm 3.

(b) *The conclusions of Theorem 2.1.1 and Theorem 2.1.2 are still valid with* Algorithm 3 *replaced by* Algorithm 4.

Proof: (a) Let $\{x_k\}$, $\{\tilde{x}_k\}$, $\{\varepsilon_k\}$ and $\{u_k\}$ be generated by Algorithm 4. Firstly, note that the identity in (2.22) yields $u_k + \tilde{x}_k - x_{k-1} = 0$ and, consequently,

$$\begin{aligned}\|\tilde{x}_k - P_V(x_{k-1})\|^2 &= \|P_V(\tilde{x}_k - x_{k-1})\|^2 + \|P_{V^\perp}(\tilde{x}_k)\|^2\\ &= \|P_V(\tilde{x}_k - x_{k-1})\|^2 + \|P_{V^\perp}(u_k - x_{k-1})\|^2\\ &= \|P_V(\tilde{x}_k) + P_{V^\perp}(u_k) - x_{k-1}\|^2,\end{aligned}$$

and

$$\begin{aligned}P_V(\tilde{x}_k) + P_{V^\perp}(u_k) &= (\tilde{x}_k - P_{V^\perp}(\tilde{x}_k)) + P_{V^\perp}(u_k)\\ &= (x_{k-1} - u_k) - P_{V^\perp}(\tilde{x}_k) + P_{V^\perp}(u_k)\\ &= x_{k-1} - \left[P_V(u_k) + P_{V^\perp}(\tilde{x}_k)\right].\end{aligned}$$

Altogether we obtain (a).

(b) This Item is a direct consequence of (a), Theorem 2.1.1 and Theorem 2.1.2. ■

Next, we observe that Proposition 2.1.2(b) and the first remark after Algorithm 4 allow us to obtain the iteration-complexity for the Spingarn's partial inverse method.

Proposition 2.1.3 ([2, Proposition 3.2]) *Let $d_{0,V}$ denote the distance of x_0 to the solution set* (2.2) *and consider* Algorithm 4 *with $\sigma = 0$ or, equivalently, the* Spingarn's partial inverse method of [44]. *For given tolerances $\rho, \epsilon > 0$, the latter method finds*

(a) *$x, u \in \mathcal{H}$ such that $u \in T(x)$,* $\max\{\|x - P_V(x)\|, \|u - P_{V^\perp}(u)\|\} \leq \rho$ *in at most*

$$\mathcal{O}\left(\left\lceil \frac{d_{0,V}^2}{\rho^2} \right\rceil\right) \tag{2.24}$$

iterations.

(b) $x, u \in \mathcal{H}$ *and* $\varepsilon \geq 0$ *satisfying the termination criterion* (2.3) *in at most a number of iterations given in* (2.21).

Proof: (a) The statement in this item is a direct consequence of Proposition 2.1.2(b), Theorem 2.1.1(a) and the fact that $\varepsilon_k = 0$ for all $k \geq 1$ (because $\sigma = 0$ in (2.22)). (b) Here, the result follows from Proposition 2.1.2(b) and Theorem 2.1.2. ■

2.2 An inexact Spingarn's operator splitting algorithm

In this section, we consider the problem (13), i.e., the problem of finding $x \in \mathcal{H}$ such that

$$0 \in \sum_{i=1}^{m} T_i(x) \tag{2.25}$$

where $m \geq 2$ and $T_i : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone for $i = 1, \ldots, m$. As observed in [44], $x \in \mathcal{H}$ satisfies the inclusion (2.25) if and only if there exist $u_1, \ldots, u_m \in \mathcal{H}$ such that

$$u_i \in T_i(x) \quad \text{and} \quad \sum_{i=1}^{m} u_i = 0. \tag{2.26}$$

That said, we consider the (extended) solution set of (2.25) – which we assume nonempty – to be defined by

$$\begin{aligned} \mathcal{S}^*(\Sigma) :=& \{(z_i)_{i=1}^m \in \mathcal{H}^m : \exists\, x, u_1, u_2, ..., u_m \in \mathcal{H} \text{ satisfying (2.26) s.t.} \\ & z_i = x + u_i \;\; \forall i = 1, \ldots, m\}. \end{aligned} \tag{2.27}$$

In this section, we apply the results of Section 2.1 to present and study the iteration-complexity of an inexact version of the Spingarn's operator splitting method [44, Chapter 5] for solving (2.25) and, as a by-product, we obtain the iteration-complexity of the latter method.

To this end, we consider the following notion of approximate solution for (2.25): given tolerances $\rho, \delta, \epsilon > 0$, find $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m \in \mathcal{H}$, $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m \in \mathcal{H}$ and $\bar{\varepsilon}_1, \bar{\varepsilon}_2, \ldots, \bar{\varepsilon}_m \geq 0$ such that $(x_i)_{i=1}^m := (\bar{x}_i)_{i=1}^m$,

$(u_i)_{i=1}^m := (\bar{u}_i)_{i=1}^m$ and $(\varepsilon_i)_{i=1}^m := (\bar{\varepsilon}_i)_{i=1}^m$ satisfy

$$\begin{aligned} & u_i \in T_i^{\varepsilon_i}(x_i) \quad \forall i = 1, \ldots, m, \\ & \left\| \sum_{i=1}^m u_i \right\| \leq \rho, \\ & \|x_i - x_\ell\| \leq \delta \quad \forall i, \ell = 1, \ldots, m, \\ & \sum_{i=1}^m \varepsilon_i \leq \epsilon. \end{aligned} \tag{2.28}$$

For $\rho = \delta = \epsilon = 0$, criterion (2.28) gives $\bar{x}_1 = \bar{x}_2 = \cdots = \bar{x}_m =: \bar{x}$, $\sum_{i=1}^m \bar{u}_i = 0$ and $\bar{u}_i \in T_i(\bar{x})$ for all $i = 1, \ldots, m$, i.e., in this case $\bar{x}, \bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m$ satisfy (2.26).

We next present our inexact version of the Spingarn's operator splitting method [44] for solving (2.25).

Algorithm 5 An inexact Spingarn's operator splitting method for (2.25)

(0) Let $(x_0, y_{1,0}, \ldots, y_{m,0}) \in \mathcal{H}^{m+1}$ such that $y_{1,0} + \cdots + y_{m,0} = 0$ and $\sigma \in [0, 1)$ be given and set $k = 1$.

(1) For each $i = 1, \ldots, m$, compute $\tilde{x}_{i,k} \in \mathcal{H}$ and $\varepsilon_{i,k} \geq 0$ such that

$$\begin{aligned} & u_{i,k} := x_{k-1} + y_{i,k-1} - \tilde{x}_{i,k} \in T_i^{\varepsilon_{i,k}}(\tilde{x}_{i,k}), \\ & \varepsilon_{i,k} \leq \frac{\sigma^2}{2} \|\tilde{x}_{i,k} - x_{k-1}\|^2. \end{aligned} \tag{2.29}$$

(2) Define

$$x_k = \frac{1}{m} \sum_{i=1}^m \tilde{x}_{i,k}, \quad y_{i,k} = u_{i,k} - \frac{1}{m} \sum_{\ell=1}^m u_{\ell,k} \text{ for } i = 1, \ldots, m, \tag{2.30}$$

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

Remarks.

(i) Letting $\sigma = 0$ in Algorithm 5 we obtain the Spingarn's operator splitting method of [44].

(ii) In [12], Section 5, an inexact version of the Spingarn's operator splitting method – called *split-sPIM(ε)* – was proposed for solving (2.25). With a different notation, for $i = 1, \ldots, m$, each iteration of the latter method can be written as:

$$\begin{cases} u_{i,k} \in T_i^{\varepsilon_{i,k}}(\tilde{x}_{i,k}), \\ \|u_{i,k} + \tilde{x}_{i,k} - x_{k-1} - y_{i,k-1}\|^2 + 2\varepsilon_{i,k} \leq \\ \hat{\sigma}^2 \left(\|\tilde{x}_{i,k} - x_{k-1}\|^2 + \|u_{i,k} - y_{i,k-1}\|^2 \right), \\ x_k = x_{k-1} - \dfrac{1}{m} \sum_{i=1}^m u_{i,k}, \\ y_{i,k} = y_{i,k-1} - \tilde{x}_{i,k} + \dfrac{1}{m} \sum_{\ell=1}^m \tilde{x}_{\ell,k}, \end{cases} \tag{2.31}$$

where $\hat{\sigma} \in [0, 1)$. The convergence analysis of [12] consists in analyzing each iteration (2.31) in the framework of the method described in (1.15), whose convergence may fail if we take $\hat{\sigma} > 1/\sqrt{5}$, as we observed in Remark 1.2.4. On the other hand, we will prove in Proposition 2.2.1 that Algorithm 5 can be regarded as a special instance of Algorithm 4, which converges for all $\sigma \in [0, 1)$ (see Proposition 2.1.1, Proposition 2.1.2(b) and [40, Theorem 3.1]). Moreover, we mention that contrary to this work no iteration-complexity analysis is performed in [12].

(iii) An inexact primal-dual composite method of partial inverses was proposed and analyzed in [1] for solving a more general version of (2.25) (in which compositions with linear operators are included in the problem's formulation) together with its dual problem (see Problems 3.5 and 4.2 in [1]). Its (asymptotic) convergence is obtained by viewing it as a special instance of an inexact method of partial inverses, also presented and studied in [1] (see also the third remark after Algorithm 2 for a related discussion).

(iv) In [15, Corollary 2.6], the weak convergence of the (exact) Spingarn's operator splitting method for solving (2.25) with a weighted sum of maximal monotone operators is obtained by exploiting the connections between the latter method and the Douglas-Rachford splitting scheme (see also Remark 2.4 in [15]). We mention that [15] also proposed and studied parallel splitting algorithms for computing the resolvent of a weighted sum of maximal monotone operators as well as applications in convex programming.

For each $i = 1, \ldots, m$, let $\{\tilde{x}_{i,k}\}$, $\{u_{i,k}\}$ and $\{\varepsilon_{i,k}\}$ be generated by Algorithm 5 and define the *ergodic* sequences associated to them:

$$\begin{aligned}
\tilde{x}^a_{i,k} &:= \frac{1}{k}\sum_{\ell=1}^{k} \tilde{x}_{i,\ell}\,, \quad u^a_{i,k} := \frac{1}{k}\sum_{\ell=1}^{k} u_{i,\ell}\,, \\
\varepsilon^a_{i,k} &:= \frac{1}{k}\sum_{\ell=1}^{k} \left[\varepsilon_{i,\ell} + \langle \tilde{x}_{i,\ell} - \tilde{x}^a_{i,k}, u_{i,\ell} - u^a_{i,k}\rangle\right].
\end{aligned} \tag{2.32}$$

Analogously to Section 2.1, in the current section we follow the Spingarn's approach in [44] for solving problem (2.25), which consists of solving the following inclusion in the product space $\mathcal{H}^m$:

$$\mathbf{0} \in \mathbf{T}_{\mathbf{V}}(\mathbf{z}), \tag{2.33}$$

where $\mathbf{T} : \mathcal{H}^m \rightrightarrows \mathcal{H}^m$ is the maximal monotone operator defined by: For all $(x_1, x_2, \ldots, x_m) \in \mathcal{H}^m$,

$$\mathbf{T}(x_1, x_2, \ldots, x_m) := T_1(x_1) \times T_2(x_2) \times \cdots \times T_m(x_m) \tag{2.34}$$

and

$$\mathbf{V} := \{(x_1, x_2, \ldots, x_m) \in \mathcal{H}^m \ : \ x_1 = x_2 = \cdots = x_m\} \tag{2.35}$$

is a closed subspace of $\mathcal{H}^m$ whose orthogonal complement is

$$\mathbf{V}^\perp = \{(x_1, x_2, \ldots, x_m) \in \mathcal{H}^m \ : \ x_1 + x_2 + \cdots + x_m = 0\}\,. \tag{2.36}$$

Based on the above observations, we have that problem (2.25) is equivalent to (2.1) with $\mathbf{T}$ and $\mathbf{V}$ given in (2.34) and (2.35), respectively. Moreover, in this case, the orthogonal projections onto $\mathbf{V}$ and $\mathbf{V}^\perp$ have the explicit formulae:

$$\begin{aligned}
P_{\mathbf{V}}(x_1, x_2, \ldots, x_m) &= \left(\frac{1}{m}\sum_{i=1}^{m} x_i, \ldots, \frac{1}{m}\sum_{i=1}^{m} x_i\right), \\
P_{\mathbf{V}^\perp}(x_1, x_2, \ldots, x_m) &= \left(x_1 - \frac{1}{m}\sum_{i=1}^{m} x_i, \ldots, x_m - \frac{1}{m}\sum_{i=1}^{m} x_i\right).
\end{aligned} \tag{2.37}$$

Next, we show that Algorithm 5 can be regarded as a special instance of Algorithm 4 and, as a consequence, we will obtain that Theorem 2.2.1 follows from results of Section 2.1 for Algorithm 4.

Proposition 2.2.1 ([2, Proposition 4.1]) *Let* $\{x_k\}_{k\geq 0}$ *and, for each* $i = 1,\dots,m$, $\{y_{i,k}\}_{k\geq 0}$, $\{\tilde{x}_{i,k}\}_{k\geq 1}$, $\{u_{i,k}\}_{k\geq 1}$ *and* $\{\varepsilon_{i,k}\}_{k\geq 1}$ *be generated by* Algorithm 5. *Consider the sequences* $\{\mathbf{x}_k\}_{k\geq 0}$, $\{\tilde{\mathbf{x}}_k\}_{k\geq 1}$ *and* $\{\mathbf{u}_k\}_{k\geq 1}$ *in* $\mathcal{H}^m$ *and* $\{\varepsilon_k\}_{k\geq 1}$ *in* $\mathbb{R}_+$ *where*

$$\begin{aligned} &\mathbf{x}_k := (x_k + y_{1,k},\dots,x_k + y_{m,k}), \quad \tilde{\mathbf{x}}_k := (\tilde{x}_{1,k},\dots,\tilde{x}_{m,k}),\\ &\varepsilon_k := \sum_{i=1}^{m} \varepsilon_{i,k}, \quad \mathbf{u}_k := (u_{1,k},\dots,u_{m,k}). \end{aligned} \tag{2.38}$$

Then, for all $k \geq 1$,

$$\begin{aligned} &\mathbf{u}_k \in \left(T_1^{\varepsilon_{1,k}} \times \cdots \times T_m^{\varepsilon_{m,k}}\right)(\tilde{\mathbf{x}}_k), \quad \mathbf{u}_k + \tilde{\mathbf{x}}_k - \mathbf{x}_{k-1} = 0,\\ &\varepsilon_k \leq \frac{\sigma^2}{2}\|\tilde{\mathbf{x}}_k - P_{\mathbf{V}}(\mathbf{x}_{k-1})\|^2,\\ &\mathbf{x}_k = P_{\mathbf{V}}(\tilde{\mathbf{x}}_k) + P_{\mathbf{V}^\perp}(\mathbf{u}_k). \end{aligned} \tag{2.39}$$

As a consequence of (2.39), *the sequences* $\{\mathbf{x}_k\}_{k\geq 0}$, $\{\tilde{\mathbf{x}}_k\}_{k\geq 1}$, $\{\mathbf{u}_k\}_{k\geq 1}$ *and* $\{\varepsilon_k\}_{k\geq 1}$ *are generated by* Algorithm 4 *for solving* (2.9) *with* $\mathbf{T}$ *and* $\mathbf{V}$ *given in* (2.34) *and* (2.35), *respectively.*

Proof: Note that (2.39) follows directly from (2.29), (2.30), (2.38) and definition (1.3) (with $S_i = T^{\varepsilon_{i,k}}$ for $i = 1,\dots,m$). The last statement of the Proposition is a direct consequence of (2.39) and Algorithm 4's definition. ■

Theorem 2.2.1 ([2, Theorem 4.1]) *For each* $i = 1,\dots,m$, *let* $\{\tilde{x}_{i,k}\}$, $\{u_{i,k}\}$ *and* $\{\varepsilon_{i,k}\}$ *be generated by* Algorithm 5 *and let* $\{\tilde{x}^a_{i,k}\}$, $\{u^a_{i,k}\}$ *and* $\{\varepsilon^a_{i,k}\}$ *be defined in* (2.32) . *Let also* $d_{0,\Sigma}$ *denote the distance of* $(x_0 + y_{1,0},\dots,x_0 + y_{m,0})$ *to the solution set* (2.27). *The following statements hold:*

(a) *For any* $k \geq 1$, *there exists* $j \in \{1,\dots,k\}$ *such that*

$$\begin{aligned} &u_{i,j} \in T_i^{\varepsilon_{i,j}}(\tilde{x}_{i,j}) \quad \forall i = 1,\dots,m,\\ &\left\|\sum_{i=1}^{m} u_{i,j}\right\| \leq \frac{\sqrt{m}\, d_{0,\Sigma}}{\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}},\\ &\|\tilde{x}_{i,j} - \tilde{x}_{\ell,j}\| \leq \frac{2\, d_{0,\Sigma}}{\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}} \quad \forall i,\ell = 1,\dots,m,\\ &\sum_{i=1}^{m} \varepsilon_{i,j} \leq \frac{\sigma^2 d_{0,\Sigma}^2}{2(1-\sigma^2)k}\ ; \end{aligned} \tag{2.40}$$

(b) *for any* $k \geq 1$,

$$
\begin{aligned}
& u^a_{i,\,k} \in T_i^{\varepsilon^a_{i,\,k}}(\tilde{x}^a_{i,\,k}) \quad \forall i = 1,\ldots,m,\\
& \left\| \sum_{i=1}^m u^a_{i,\,k} \right\| \leq \frac{2\sqrt{m}\, d_{0,\Sigma}}{k},\\
& \|\tilde{x}^a_{i,k} - \tilde{x}^a_{\ell,k}\| \leq \frac{4\, d_{0,\Sigma}}{k} \quad \forall i,\ell = 1,\ldots,m,\\
& \sum_{i=1}^m \varepsilon^a_{i,\,k} \leq \frac{2(1+\sigma/\sqrt{1-\sigma^2})d^2_{0,\Sigma}}{k}\,.
\end{aligned}
\tag{2.41}
$$

Proof: We start by defining the ergodic sequences associated to the sequences $\{\tilde{\mathbf{x}}_k\}$, $\{\mathbf{u}_k\}$ and $\{\varepsilon_k\}$ in (2.38):

$$
\begin{aligned}
& \tilde{\mathbf{x}}^a_k := \frac{1}{k}\sum_{\ell=1}^k \tilde{\mathbf{x}}_\ell\,, \quad \mathbf{u}^a_k := \frac{1}{k}\sum_{\ell=1}^k \mathbf{u}_\ell\,,\\
& \varepsilon^a_k := \frac{1}{k}\sum_{\ell=1}^k \left[\varepsilon_\ell + \langle \tilde{\mathbf{x}}_\ell - \tilde{\mathbf{x}}^a_k, \mathbf{u}_\ell - \mathbf{u}^a_k\rangle\right].
\end{aligned}
\tag{2.42}
$$

Observe that from (2.2), (2.27), (2.34), (2.35) and (2.36) we obtain $\mathcal{S}^*(\mathbf{V},\mathbf{T}) = \mathcal{S}^*(\Sigma)$ and, consequently, $d_{0,\mathbf{V}} = d_{0,\Sigma}$. That said, it follows from the last statement in Proposition 2.2.1, Proposition 2.1.2(a) and Theorem 2.1.1 that for any $k \geq 1$, there exists $j \in \{1,\ldots,k\}$ such that

$$
\begin{aligned}
& \mathbf{u}_j \in \left(T_1^{\varepsilon_{1,\,j}} \times T_2^{\varepsilon_{2,\,j}} \times \cdots \times T_m^{\varepsilon_{m,\,j}}\right)(\tilde{\mathbf{x}}_j),\\
& \sqrt{\|\tilde{\mathbf{x}}_j - P_{\mathbf{V}}(\tilde{\mathbf{x}}_j)\|^2 + \|\mathbf{u}_j - P_{\mathbf{V}^\perp}(\mathbf{u}_j)\|^2} \leq \frac{d_{0,\Sigma}}{\sqrt{k}}\sqrt{\frac{1+\sigma}{1-\sigma}},\\
& \varepsilon_j \leq \frac{\sigma^2 d^2_{0,\Sigma}}{2(1-\sigma^2)k}\,,
\end{aligned}
\tag{2.43}
$$

and

$$
\begin{aligned}
& \sqrt{\|\tilde{\mathbf{x}}^a_k - P_{\mathbf{V}}(\tilde{\mathbf{x}}^a_k)\|^2 + \|\mathbf{u}^a_k - P_{\mathbf{V}^\perp}(\mathbf{u}^a_k)\|^2} \leq \frac{2d_{0,\Sigma}}{k},\\
& 0 \leq \varepsilon^a_k \leq \frac{2(1+\sigma/\sqrt{1-\sigma^2})d^2_{0,\Sigma}}{k}\,.
\end{aligned}
\tag{2.44}
$$

In particular, we see that Item (a) of Theorem 2.2.1 follows from (2.43), (2.38) and (2.37). Note now that from (2.42), (2.38) and (2.32) we obtain, for all $k \geq 1$,

$$\begin{aligned} \tilde{\mathbf{x}}_k^a &= (\tilde{x}_{1,\,k}^a, \tilde{x}_{2,\,k}^a, \ldots, \tilde{x}_{m,\,k}^a), \\ \mathbf{u}_k^a &= (u_{1,\,k}^a, u_{2,\,k}^a, \ldots, u_{m,\,k}^a), \\ \varepsilon_k^a &= \sum_{i=1}^m \varepsilon_{i,\,k}^a. \end{aligned} \tag{2.45}$$

Hence, the inequalities in (2.41) follow from (2.44), (2.45) and (2.37). To finish the proof of the theorem it suffices to show the inclusions in (2.41) for each $i = 1, \ldots, m$ and all $k \geq 1$. To this end, note that for each $i = 1, \ldots, m$ the desired inclusion is a direct consequence of the inclusions in (2.29), the definitions in (2.32) and Theorem 1.1.1 (with $T = T_i$ for each $i = 1, \ldots, m$). ■

As a consequence of Theorem 2.2.1(b), we obtain the iteration-complexity of Algorithm 5 to find $x_1, x_2, \ldots, x_m \in \mathcal{H}$, $u_1, u_2, \ldots, u_m \in \mathcal{H}$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \geq 0$ satisfying the termination criterion (2.28):

Theorem 2.2.2 [2, Theorem 4.2](Iteration-complexity) *Let $d_{0,\Sigma}$ denote the distance of $(x_0 + y_{1,0}, \ldots, x_0 + y_{m,0})$ to the solution set* (2.27) *and let $\rho, \delta, \epsilon > 0$ be given tolerances.*

Then, Algorithm 5 *finds $x_1, x_2, \ldots, x_m \in \mathcal{H}$, $u_1, u_2, \ldots, u_m \in \mathcal{H}$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \geq 0$ satisfying the termination criterion* (2.28) *in at most*

$$\mathcal{O}\left(\max\left\{\left\lceil\frac{\sqrt{m}\,d_{0,\Sigma}}{\rho}\right\rceil, \left\lceil\frac{d_{0,\Sigma}}{\delta}\right\rceil, \left\lceil\frac{d_{0,\Sigma}^2}{\epsilon}\right\rceil\right\}\right) \tag{2.46}$$

iterations.

Using the first remark after Algorithm 5 and Theorem 2.2.1, we also obtain the *pointwise* and *ergodic* iteration-complexities of Spingarn's operator splitting method [44, Chapter 5].

Theorem 2.2.3 [2, Theorem 4.3](Iteration-complexity) *Let $d_{0,\Sigma}$ denote the distance of $(x_0 + y_{1,0}, \ldots, x_0 + y_{m,0})$ to the solution set* (2.27) *and consider* Algorithm 5 *with $\sigma = 0$ or, equivalently, the* Spingarn's operator splitting method of [44]. *For given tolerances $\rho, \delta, \epsilon > 0$, the latter method finds*

(a) $x_1, x_2, \ldots, x_m \in \mathcal{H}$ *and* $u_1, u_2, \ldots, u_m \in \mathcal{H}$ *such that*

$$\begin{aligned} & u_i \in T_i(x_i) \quad \forall i = 1, \ldots, m, \\ & \left\| \sum_{i=1}^m u_i \right\| \leq \rho, \\ & \|x_i - x_\ell\| \leq \delta\,, \quad \forall i, \ell = 1, \ldots, m, \end{aligned} \tag{2.47}$$

in at most

$$\mathcal{O}\left(\max\left\{ \left\lceil \frac{m\, d_{0,\Sigma}^2}{\rho^2} \right\rceil, \left\lceil \frac{d_{0,\Sigma}^2}{\delta^2} \right\rceil \right\} \right) \tag{2.48}$$

iterations.

(b) $x_1, x_2, \ldots, x_m \in \mathcal{H}$, $u_1, u_2, \ldots, u_m \in \mathcal{H}$ *and* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \geq 0$ *satisfying the termination criterion* (2.28) *in at most the number of iterations given in* (2.46).

Proof: (a) This item follows from Theorem 2.2.1(a) and the fact that $\varepsilon_{i,\,k} = 0$ for each $i = 1, \ldots, m$ and for all $k \geq 1$ (because $\sigma = 0$ in (2.29)). (b) This item follows directly from Theorem 2.2.2. ■

2.3 Applications to multi-term composite convex optimization

In this section, we show how Algorithm 5 and its iteration-complexity results can be used to derive a *parallel forward-backward splitting method* for multi-term composite convex optimization and to study its iteration-complexity. More precisely, consider the minimization problem

$$\min_{x \in \mathcal{H}} \sum_{i=1}^m \left(f_i + \varphi_i\right)(x) \tag{2.49}$$

where $m \geq 2$ and the following conditions are assumed to hold for all $i = 1, \ldots, m$:

(A.1) $f_i : \mathcal{H} \to \mathbb{R}$ is convex, and differentiable with a L_i-Lipschitz continuous gradient, i.e., there exists $L_i > 0$ such that

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq L_i \|x - y\| \quad \forall x, y \in \mathcal{H}; \tag{2.50}$$

(A.2) $\varphi_i : \mathcal{H} \to \overline{\mathbb{R}}$ is proper, convex and closed with an easily computable resolvent $(\lambda \partial \varphi_i + I)^{-1}$, for any $\lambda > 0$;

(A.3) the solution set of (2.49) is nonempty.

We also assume standard regularity conditions [1] on the functions φ_i which make (2.49) equivalent to the monotone inclusion problem (2.25) with $T_i := \nabla f_i + \partial \varphi_i$, for all $i = 1, \ldots, m$, i.e., which make it equivalent to the problem of finding $x \in \mathcal{H}$ such that

$$0 \in \sum_{i=1}^{m} \left(\nabla f_i + \partial \varphi_i \right)(x). \tag{2.51}$$

Analogously to (2.28), we consider the following notion of approximate solution for (2.49): given tolerances $\rho, \delta, \epsilon > 0$, find $\bar{x}_1, \ldots, \bar{x}_m \in \mathcal{H}$, $\bar{u}_1, \ldots, \bar{u}_m \in \mathcal{H}$ and $\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_m \geq 0$ such that $(x_i)_{i=1}^m := (\bar{x}_i)_{i=1}^m$, $(u_i)_{i=1}^m := (\bar{u}_i)_{i=1}^m$ and $(\varepsilon_i)_{i=1}^m := (\bar{\varepsilon}_i)_{i=1}^m$ satisfy (2.28) with $T_i^{\varepsilon_i}$ replaced by $\partial_{\varepsilon_i} f_i + \partial \varphi_i$, for each $i = 1, \ldots, m$. For $\rho = \delta = \epsilon = 0$, this criterion gives $\bar{x}_1 = \bar{x}_2 = \cdots = \bar{x}_m =: \bar{x}$, $\sum_{i=1}^m \bar{u}_i = 0$ and $\bar{u}_i \in (\nabla f_i + \partial \varphi_i)(\bar{x})$ for all $i = 1, \ldots, m$, i.e., in this case $\bar{x}$ solves (2.51).

We will present a parallel forward-backward method for solving (2.49) whose iteration-complexity is obtained by viewing it as a special instance of Algorithm 5. Since problem (2.49) appears in various applications of convex optimization, it turns out that the development of efficient numerical schemes for solving it – specially with $m \geq 2$ very large – is of great importance.

Next is our method for solving (2.49).

[1] see, e.g., [6, Corollary 16.39]

Algorithm 6 A parallel forward-backward splitting method for (2.49)

(0) Let $(x_0, y_{1,0}, \ldots, y_{m,0}) \in \mathcal{H}^{m+1}$ such that $y_{1,0} + \cdots + y_{m,0} = 0$ and $\sigma \in (0,1)$ be given and set $\lambda = \sigma^2 / \max\{L_i\}_{i=1}^m$ and $k = 1$.

(1) For each $i = 1, \ldots, m$, compute

$$\tilde{x}_{i,\,k} = (\lambda \partial \varphi_i + I)^{-1} \left(x_{k-1} + y_{i,\,k-1} - \lambda \nabla f_i(x_{k-1}) \right). \tag{2.52}$$

(2) Define

$$x_k = \frac{1}{m} \sum_{i=1}^{m} \tilde{x}_{i,\,k}, \qquad y_{i,\,k} = y_{i,\,k-1} + x_k - \tilde{x}_{i,\,k} \text{ for } i = 1, \ldots, m, \tag{2.53}$$

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

Remarks.

(i) Since in (2.52) we have a forward step in the direction $-\nabla f_i(x_{k-1})$ and a backward step given by the resolvent of $\partial \varphi_i$, Algorithm 6 can be regarded as a parallel variant of the classical forward-backward splitting algorithm [19].

(ii) For $m = 1$ the above method coincides with the forward-backward method of [33], for which the iteration-complexity was studied in the latter reference.

(iii) A forward-partial-inverse splitting method for solving (2.1) in which T is assumed to be the sum of a point-to-set maximal monotone operator and a cocoercive point-to-point (maximal monotone) operator was proposed and analyzed in [8]. When applied to solve (2.51), for a specific choice of parameters, the method presented in [8, Theorem 5.2] (with a different notation) can be described according to the following iteration, for $k \geq 1$ and $i = 1, \ldots, m$:

$$\begin{cases} \tilde{x}_{i,k} = (\lambda \partial \varphi_i + I)^{-1} \Big(x_{k-1} + \lambda \left[y_{i,k-1} - \frac{1}{m} \sum_{\ell=1}^m \nabla f_\ell(x_{k-1}) \right] \Big), \\ x_k = \frac{1}{m} \sum_{i=1}^m \tilde{x}_{i,\,k}, \\ y_{i,\,k} = y_{i,\,k-1} + \lambda^{-1}(x_k - \tilde{x}_{i,\,k}) \text{ for } i = 1, \ldots, m, \end{cases} \tag{2.54}$$

which is similar to Algorithm 5. We mention that the (asymptotic) convergence of (2.54) as well as other algorithms related to the algorithm of this section can be found in Theorem 5.2 and Corollary 5.3, and Section 6.1 of [8], respectively. Moreover, in contrast to this work, no iteration-complexity analysis is performed in [8].

For each $i = 1, \dots, m$, let $\{x_k\}$, $\{\tilde{x}_{i,k}\}$ be generated by Algorithm 6, $\{u_{i,k}\}$ and $\{\varepsilon_{i,k}\}$ be defined in (2.56) and let $\{\tilde{x}^a_{i,k}\}$, $\{u^a_{i,k}\}$ and $\{\varepsilon^a_{i,k}\}$ be given in (2.32). Define, for all $k \geq 1$,

$$\begin{aligned} u'_{i,k} &:= \frac{1}{\lambda} u_{i,k}, \quad \varepsilon'_{i,k} := \frac{1}{\lambda} \varepsilon_{i,k}, \\ u'^{\,a}_{i,k} &:= \frac{1}{\lambda} u^a_{i,k}, \quad \varepsilon'^{\,a}_{i,k} := \frac{1}{\lambda} \varepsilon^a_{i,k}, \\ \varepsilon''^{\,a}_{i,k} &:= \frac{1}{k} \sum_{\ell=1}^{k} \left[\varepsilon'_{i,\ell} + \langle \tilde{x}_{i,\ell} - \tilde{x}^a_{i,k}, \nabla f_i(x_{\ell-1}) - \frac{1}{k} \sum_{s=1}^{k} \nabla f_i(x_{s-1}) \rangle \right]. \end{aligned} \tag{2.55}$$

Next proposition shows that Algorithm 6 is a special instance of Algorithm 5 for solving (2.25) with $T_i = \nabla(\lambda f_i) + \partial(\lambda \varphi_i)$ for all $i = 1, \dots, m$.

Proposition 2.3.1 ([2, Proposition 4.2]) *Let $\{x_k\}_{k\geq 0}$ and, for $i = 1, \dots, m$, $\{y_{i,k}\}_{k\geq 0}$ and $\{\tilde{x}_{i,k}\}_{k\geq 1}$ be generated by* Algorithm 6. *For $i = 1, \dots, m$, consider the sequences $\{u_{i,k}\}_{k\geq 1}$ and $\{\varepsilon_{i,k}\}_{k\geq 1}$ where, for all $k \geq 1$,*

$$\begin{aligned} u_{i,k} &:= x_{k-1} + y_{i,k-1} - \tilde{x}_{i,k}, \\ \varepsilon_{i,k} &:= \lambda \left[f_i(\tilde{x}_{i,k}) - f_i(x_{k-1}) - \langle \nabla f_i(x_{k-1}), \tilde{x}_{i,k} - x_{k-1} \rangle \right]. \end{aligned} \tag{2.56}$$

Then, for all $k \geq 1$,

$$\nabla(\lambda f_i)(x_{k-1}) \in \partial_{\varepsilon_{i,k}}(\lambda f_i)(\tilde{x}_{i,k}), \tag{2.57}$$

$$u_{i,k} - \nabla(\lambda f_i)(x_{k-1}) \in \partial(\lambda \varphi_i)(\tilde{x}_{i,k}), \tag{2.58}$$

$$u_{i,k} \in \left(\partial_{\varepsilon_{i,k}}(\lambda f_i) + \partial(\lambda \varphi_i) \right)(\tilde{x}_{i,k}), \tag{2.59}$$

$$0 \leq \varepsilon_{i,k} \leq \frac{\sigma^2}{2} \|\tilde{x}_{i,k} - x_{k-1}\|^2, \tag{2.60}$$

$$x_k \text{ and } y_{i,k} \text{ satisfy (2.30).} \tag{2.61}$$

As a consequence of (2.56)–(2.61)*, the sequences* $\{x_k\}_{k\geq 0}$*,* $\{y_{i,\,k}\}_{k\geq 1}$*,* $\{\tilde{x}_{i,\,k}\}_{k\geq 0}$*,* $\{\varepsilon_{i,k}\}_{k\geq 1}$ *and* $\{u_{i,\,k}\}_{k\geq 1}$ *are generated by* Algorithm 5 *for solving* (2.25) *with*

$$T_i = \nabla(\lambda f_i) + \partial(\lambda \varphi_i) \quad \forall i = 1, \ldots, m.$$

Proof: Inclusion (2.57) follows from Lemma 1.1.1 with $(f, x, \tilde{x}, v, \varepsilon) = (\lambda f_i, x_{k-1}, \tilde{x}_{i,\,k}, \nabla(\lambda f_i)(x_{k-1}), \varepsilon_{i,\,k})$, where $\varepsilon_{i,\,k}$ is given in (2.56). Inclusion (2.58) follows from (2.52), the first identity in (2.56) and Lemma 1.1.2(a). Inclusion (2.59) is a direct consequence of (2.57) and (2.58). The inequalities in (2.60) follow from assumption (A.1), the second identity in (2.56), Lemma 1.1.3 and the definition of $\lambda > 0$ in Algorithm 6. The fact that x_k satisfies (2.30) follows from the first identities in (2.30) and (2.53). Direct use of (2.53) and the assumption that $y_{1,0} + \cdots + y_{m,0} = 0$ in step 0 of Algorithm 6 gives $\sum_{\ell=1}^{m} y_{\ell,\,k} = 0$ for all $k \geq 0$, which, in turn, combined with the second identity in (2.53) and the first identity in (2.56) proves that $y_{i,\,k}$ satisfies the second identity in (2.30). Altogether, we obtain (2.61). The last statement of the proposition follows from (2.56)–(2.61) and Proposition 1.1.1(b; e). ■

Next, we prove (global) convergence rates for Algorithm 6.

Theorem 2.3.1 ([2, Theorem 4.4]) *For each* $i = 1, \ldots, m$*, let* $\{\tilde{x}_{i,\,k}\}$ *be generated by* Algorithm 6 *and* $\{\tilde{x}^{\,a}_{i,\,k}\}$ *be given in* (2.32)*. Let* $\{u'_{i,\,k}\}$*,* $\{\varepsilon'_{i,\,k}\}$*,* $\{u'^{\,a}_{i,\,k}\}$*,* $\{\varepsilon'^{\,a}_{i,\,k}\}$ *and* $\{\varepsilon''^{\,a}_{i,\,k}\}$ *be given in* (2.55)*. Let also* $d_{0,\Sigma}$ *denote the distance of* $(x_0 + y_{1,0}, \ldots, x_0 + y_{m,0})$ *to the solution set* (2.27) *in which* $T_i := \nabla(\lambda f_i) + \partial(\lambda \varphi_i)$ *for* $i = 1, \ldots, m$*, and define* $L_\Sigma := \max\{L_i\}_{i=1}^m$*. The following hold:*

(a) *For any* $k \geq 1$*, there exists* $j \in \{1, \ldots, k\}$ *such that*

$$\begin{aligned} & u'_{i,j} \in \left(\partial_{\varepsilon'_{i,j}} f_i + \partial \varphi_i\right)(\tilde{x}_{i,j}) \quad \forall i = 1, \ldots, m, \\ & \left\| \sum_{i=1}^{m} u'_{i,j} \right\| \leq \frac{\sqrt{m}\, L_\Sigma\, d_{0,\Sigma}}{\sigma^2 \sqrt{k}} \sqrt{\frac{1+\sigma}{1-\sigma}}, \\ & \|\tilde{x}_{i,j} - \tilde{x}_{\ell,j}\| \leq \frac{2\, d_{0,\Sigma}}{\sqrt{k}} \sqrt{\frac{1+\sigma}{1-\sigma}} \quad \forall i, \ell = 1, \ldots, m, \\ & \sum_{i=1}^{m} \varepsilon'_{i,j} \leq \frac{L_\Sigma\, d_{0,\Sigma}^2}{2(1-\sigma^2)k}\ ; \end{aligned} \tag{2.62}$$

(b) *for any $k \geq 1$,*

$$
\begin{aligned}
&u'^{\,a}_{i,\,k} \in \left(\partial_{\varepsilon''^{\,a}_{i,\,k}} f_i + \partial_{\left(\varepsilon'^{\,a}_{i,\,k} - \varepsilon''^{\,a}_{i,\,k}\right)} \varphi_i\right)(\tilde{x}^a_{i,\,k}) \quad \forall i = 1, \ldots, m, \\
&\left\| \sum_{i=1}^m u'^{a}_{i,k} \right\| \leq \frac{2\sqrt{m}\, L_\Sigma\, d_{0,\Sigma}}{\sigma^2 k}, \\
&\|\tilde{x}^a_{i,\,k} - \tilde{x}^a_{\ell,k}\| \leq \frac{4 d_{0,\Sigma}}{k} \quad \forall i, \ell = 1, \ldots, m, \\
&\sum_{i=1}^m \varepsilon'^{\,a}_{i,\,k} \leq \frac{2(1 + \sigma/\sqrt{1-\sigma^2})\, L_\Sigma\, d^2_{0,\Sigma}}{\sigma^2 k}.
\end{aligned}
\tag{2.63}
$$

Proof:

From the last statement of Proposition 2.3.1, the fact that

$$\left(\sum_{i=1}^m [\nabla f_i + \partial \varphi_i]\right)^{-1}(0) = \left(\sum_{i=1}^m [\nabla(\lambda f_i) + \partial(\lambda \varphi_i)]\right)^{-1}(0)$$

and Theorem 2.2.1 we obtain that (2.40) and (2.41) hold. As a consequence of the latter fact, (2.59), (2.55), Lemma 1.1.3(b), the fact that $\lambda = \sigma^2/L_\Sigma$ and some direct calculations we obtain (2.62) and the inequalities in (2.63). To finish the proof, it suffices to prove the inclusion in (2.63). To this end, note first that from (2.57), (2.32), the last identity in (2.55), Lemma 1.1.2(b) and Theorem 1.1.1(b) we obtain, for each $i = 1, \ldots, m$,

$$\frac{1}{k} \sum_{s=1}^k \nabla f_i(x_{s-1}) \in \partial_{\varepsilon''^{\,a}_{i,\,k}} f_i(\tilde{x}^{\,a}_{i,\,k}) \quad \forall k \geq 1. \tag{2.64}$$

On the other hand, it follows from (2.58), Lemma 1.1.2(a), (2.55), Theorem 1.1.1(b) and some direct calculations that, for each $i = 1, \ldots, m$,

$$u'^{\,a}_{i,\,k} - \frac{1}{k} \sum_{s=1}^k \nabla f_i(x_{s-1}) \in \partial_{\left(\varepsilon'^{\,a}_{i,\,k} - \varepsilon''^{\,a}_{i,\,k}\right)} \varphi_i(\tilde{x}^{\,a}_{i,\,k}) \quad \forall k \geq 1, \tag{2.65}$$

which, in turn, combined with (2.64) gives the inclusion in (2.63). ∎

The following theorem is a direct consequence of Theorem 2.3.1.

Theorem 2.3.2 [2, Theorem 4.5](iteration-complexity) *Let $d_{0,\Sigma}$ denote the distance of $(x_0 + y_{1,0}, \ldots, x_0 + y_{m,0})$ to the solution set (2.27) in which $T_i := \nabla(\lambda f_i) + \partial(\lambda \varphi_i)$, for $i = 1, \ldots, m$, and let $\rho, \delta, \epsilon > 0$ be given tolerances. Let $L_\Sigma := \max\{L_i\}_{i=1}^m$. Then,* Algorithm 6 *finds*

(a) $x_1, x_2, \ldots, x_m \in \mathcal{H}$, $u_1, u_2, \ldots, u_m \in \mathcal{H}$ *and* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \geq 0$ *satisfying the termination criterion* (2.28) *with* $T_i^{\varepsilon_i}$ *replaced by* $\partial_{\varepsilon_i} f_i + \partial \varphi_i$ *in at most*

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{m\, L_\Sigma^2\, d_{0,\Sigma}^2}{\rho^2} \right\rceil, \left\lceil \frac{d_{0,\Sigma}^2}{\delta^2} \right\rceil, \left\lceil \frac{L_\Sigma\, d_{0,\Sigma}^2}{\epsilon} \right\rceil\right\}\right) \tag{2.66}$$

iterations.

(b) $x_1, x_2, \ldots, x_m \in \mathcal{H}$, $u_1, u_2, \ldots, u_m \in \mathcal{H}$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \geq 0$ *and* $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_m \geq 0$ *such that*

$$\begin{aligned}
& u_i \in (\partial_{\varepsilon_i} f_i + \partial_{\hat{\varepsilon}_i} \varphi_i)(x_i) \quad \forall i = 1, \ldots, m, \\
& \left\| \sum_{i=1}^m u_i \right\| \leq \rho, \\
& \|x_i - x_\ell\| \leq \delta\,, \quad \forall i, \ell = 1, \ldots, m, \\
& \sum_{i=1}^m (\varepsilon_i + \hat{\varepsilon}_i) \leq \epsilon
\end{aligned} \tag{2.67}$$

in at most

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{\sqrt{m}\, L_\Sigma\, d_{0,\Sigma}}{\rho} \right\rceil, \left\lceil \frac{d_{0,\Sigma}}{\delta} \right\rceil, \left\lceil \frac{L_\Sigma\, d_{0,\Sigma}^2}{\epsilon} \right\rceil\right\}\right) \tag{2.68}$$

iterations.

Chapter 3

On the convergence rate of the SPDG algorithm

In this chapter, we show that the scaled proximal decomposition on the graph of a maximal monotone operator (SPDG) algorithm [25] can be analyzed within the Spingarn's partial inverse framework. We analyze the global (nonasymptotic) convergence rate of the SPDG algorithm under the assumptions that the operator T in (7) is strongly monotone and Lipschitz continuous [25]. We will prove, in particular, that under such conditions on T, the partial inverse T_V is strongly monotone as well, which allows one to employ recent results on the convergence and iteration-complexity of proximal point type methods for strongly monotone operators (see Proposition 1.3.1).

By showing that the (scaled) partial inverse of T – with respect to V– is strongly monotone, we obtain a potentially faster convergence to the SPDG algorithm when compared to the one proved in [25] by means of fixed point techniques. Moreover, the convergence rates obtained in this chapter allows one to measure the convergence speed of the SPDG algorithm on three different measures of approximate solution to the problem (7) (see Theorem 3.0.2 and the remarks right below it).

We consider (again) problem (7), i.e., the problem of finding $x, u \in \mathcal{H}$ such that

$$x \in V, \quad u \in V^{\perp} \text{ and } u \in T(x) \tag{3.1}$$

where now the following hold:

A1) V is a closed vector subspace of $\mathcal{H}$.

A2) $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is (maximal) η-strongly monotone, i.e., T is maximal monotone and there exists $\eta > 0$ such that

$$\langle z - z', v - v' \rangle \geq \eta \|z - z'\|^2 \qquad \forall v \in T(z), v' \in T(z'). \quad (3.2)$$

A3) $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is L-Lipschitz continuous, i.e., there exists $L > 0$ such that

$$\|v - v'\| \leq L\|z - z'\| \qquad \forall v \in T(z), v' \in T(z'). \quad (3.3)$$

In order to study the iteration-complexity of the SPDG algorithm, let us recall the following notion of approximate solution of (3.1), obtained in Chapter 2: for a given tolerance $\rho > 0$, find $x, u \in \mathcal{H}$ such that

$$u \in T(x), \quad \max\{\|x - P_V(x)\|, \gamma\|u - P_{V^\perp}(u)\|\} \leq \rho, \quad (3.4)$$

where $\gamma > 0$. For $\rho = 0$, criterion (3.4) gives $x \in V$, $u \in V^\perp$ and $u \in T(x)$, i.e., in this case the pair (x, u) is a solution of (3.1). We mention that criterion (3.4) naturally appears in different settings and has not been considered in [25].

Next, we formally present the scaled proximal decomposition on the graph of a maximal monotone operator (SPDG) algorithm, already briefly discussed (with a different notation) in (15).

Algorithm 7 SPDG algorithm for solving (3.1) [25, Algorithm 3]

(0) Let $x_0 \in V$, $y_0 \in V^\perp$, $\gamma > 0$ be given and set $k = 1$.

(1) Compute

$$\begin{aligned} \tilde{x}_k &= (\gamma T + I)^{-1}(x_{k-1} + \gamma y_{k-1}), \\ u_k &= \gamma^{-1}(x_{k-1} + \gamma y_{k-1} - \tilde{x}_k). \end{aligned} \quad (3.5)$$

(2) If $\tilde{x}_k \in V$ and $u_k \in V^\perp$, then stop. Otherwise, define

$$x_k = P_V(\tilde{x}_k), \quad y_k = P_{V^\perp}(u_k), \quad (3.6)$$

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

Remarks.

(i) Algorithm 7 was originally proposed and studied in [25]. When $\gamma = 1$, it reduces to the Spingarn's partial inverse method for solving (3.1). The authors of the latter reference emphasize the importance of introducing the scaling $\gamma > 0$ in order to speed up the convergence of the SPDG algorithm, specially when solving ill-conditioned problems.

(ii) As we mentioned earlier, one of the contributions of this thesis is to show that similar results (actually potentially better) to the one obtained in [25] regarding the convergence rate of Algorithm 7 can be proved by means of the Spingarn's partial inverse framework, instead of fixed point techniques.

The following result appears (with a different notation) inside the proof of [25, Theorem 4.2].

Theorem 3.0.1 (inside the proof of [25, Theorem 4.2]) *If T is η–strongly monotone and Lipschitz continuous with constant L, then the convergence of the sequence $\{(x_k, \gamma y_k)\}$ is linear, in the sense that*

$$\|x^* - x_k\|^2 + \gamma^2\|u^* - y_k\|^2 \le \left(1 - \frac{2\gamma\eta}{(1+\gamma L)^2}\right)^k d_0^2 \qquad \forall k \ge 1, \tag{3.7}$$

where (x^*, u^*) is the (unique) solution of (3.1) *and*

$$d_0 := \sqrt{\|x^* - x_0\|^2 + \gamma^2\|u^* - y_0\|^2}. \tag{3.8}$$

Remarks.

(i) The optimal convergence speed is achieved by letting $\gamma = 1/L$ in (3.7), in which case (see p. 461 in [25])

$$\|x^* - x_k\|^2 + \gamma^2\|u^* - y_k\|^2 \le \left(1 - \frac{\eta}{2L}\right)^k d_0^2 \qquad \forall k \ge 1. \tag{3.9}$$

(ii) It follows from (3.9) that, for a given tolerance $\rho > 0$, Algorithm 7 finds $x, y \in \mathcal{H}$ such that $\|x^* - x\|^2 + \gamma^2\|u^* - y\|^2 \le \rho$ after performing at most

$$2 + \log\left(\frac{2L}{2L-\eta}\right)^{-1} \log\left(\frac{d_0^2}{\rho}\right) \tag{3.10}$$

iterations. In the third remark after Theorem 3.0.2, we show that our approach provides a potentially better upper bound on the number of iterations needed by the SPDG algorithm to achieve prescribed tolerances, specially for ill-conditioned problems.

A direct consequence of the next proposition is that, in contrast to the reference [25], it is possible to analyze the SPDG algorithm within the original Spingarn's partial inverse framework. We show that under assumptions A2) and A3), the partial inverse operator T_V is strongly monotone.

Proposition 3.0.1 ([3, Proposition 2.2]) *Under the assumptions* A2) *and* A3) *on the maximal monotone operator T, its partial inverse T_V with respect to V is μ-strongly (maximal) monotone with*

$$\mu = \frac{\eta}{1+L^2} > 0. \tag{3.11}$$

Proof: Take $v \in T_V(z)$, $v' \in T_V(z')$ and note that, from (1.8), we have

$$\begin{aligned} P_V(v) + P_{V^\perp}(z) &\in T(P_V(z) + P_{V^\perp}(v)), \\ P_V(v') + P_{V^\perp}(z') &\in T(P_V(z') + P_{V^\perp}(v')), \end{aligned} \tag{3.12}$$

which, in turn, combined with the assumption A2) and after some direct calculations yields

$$\begin{aligned} \langle z - z', v - v' \rangle &= \langle P_V(z-z') + P_{V^\perp}(z-z'), P_V(v-v') + P_{V^\perp}(v-v') \rangle \\ &= \langle P_V(z-z') + P_{V^\perp}(v-v'), P_V(v-v') + P_{V^\perp}(z-z') \rangle \\ &\geq \eta \| P_V(z-z') + P_{V^\perp}(v-v') \|^2 \\ &= \eta \left(\| P_V(z-z') \|^2 + \| P_{V^\perp}(v-v') \|^2 \right) \qquad (3.13) \\ &\geq \eta \| P_V(z-z') \|^2. \qquad (3.14) \end{aligned}$$

On the other hand, assumption A3) and (3.12) imply

$$\begin{aligned} &\| [P_V(v) + P_{V^\perp}(z)] - [P_V(v') + P_{V^\perp}(z')] \| \leq \\ &\qquad L \| [P_V(z) + P_{V^\perp}(v)] - [P_V(z') + P_{V^\perp}(v')] \|, \end{aligned}$$

which, in particular, gives

$$\begin{aligned} &\| P_V(z-z') \|^2 + \| P_{V^\perp}(v-v') \|^2 \geq \\ &\qquad \geq \frac{1}{L^2} \left(\| P_V(v-v') \|^2 + \| P_{V^\perp}(z-z') \|^2 \right) \\ &\qquad \geq \frac{1}{L^2} \| P_{V^\perp}(z-z') \|^2. \end{aligned} \tag{3.15}$$

Using (3.14) and combining (3.13) and (3.15) we find, respectively,

$$\langle z - z', v - v' \rangle \geq \eta \|P_V(z-z')\|^2, \quad L^2\langle z - z', v - v' \rangle \geq \eta \|P_{V^\perp}(z-z')\|^2.$$

The desired result now follows by adding the above inequalities and by using the definition of $\mu > 0$ in (3.11). ■

Next, we show that Algorithm 7 is a special instance of Algorithm 2.

Proposition 3.0.2 ([3, Proposition 2.3]) *Let* $\{x_k\}$ *and* $\{y_k\}$ *be generated by* Algorithm 7 *and define*

$$z_k = x_k + \gamma y_k \qquad \forall k \geq 0. \tag{3.16}$$

Then, for all $k \geq 1$*,*

$$z_{k-1} - z_k = P_V(\gamma u_k) + P_{V^\perp}(\tilde{x}_k), \tag{3.17}$$

$$z_k = ((\gamma T)_V + I)^{-1} z_{k-1}. \tag{3.18}$$

As a consequence of (3.18)*, we have that* Algorithm 7 *is a special instance of* Algorithm 2 *with* $\lambda = 1$ *for solving* (1.28) *with* $A = (\gamma T)_V$*.*

Proof: Using (3.5), we obtain $\gamma u_k \in (\gamma T)(\tilde{x}_k)$ and, as a consequence, from (1.8), we have

$$P_V(\gamma u_k) + P_{V^\perp}(\tilde{x}_k) \in (\gamma T)_V(P_V(\tilde{x}_k) + P_{V^\perp}(\gamma u_k)). \tag{3.19}$$

From the second identity in (3.5), we have $\gamma u_k + \tilde{x}_k = x_{k-1} + \gamma y_{k-1}$, which combined with (3.6) gives

$$P_V(\gamma u_k) = x_{k-1} - x_k, \qquad P_{V^\perp}(\tilde{x}_k) = \gamma(y_{k-1} - y_k),$$

which, in turn, is equivalent to (3.17). Using (3.19), (3.17), (3.16) and (3.6) we find $z_{k-1} - z_k \in (\gamma T)_V(z_k)$, which is clearly equivalent to (3.18). The last statement of the proposition follows directly from (3.18) and Algorithm 2's definition. ■

In the next theorem, we present one of the main contributions of this thesis, namely, convergence rates for the SPDG algorithm, obtained within the original Spingarn's partial inverse framework.

Theorem 3.0.2 ([3, Theorem 2.4]) *Let* $\{x_k\}$, $\{y_k\}$, $\{\tilde{x}_k\}$ *and* $\{u_k\}$ *be generated by* Algorithm 7, *let* (x^*, u^*) *be the (unique) solution of* (3.1) *and let* d_0 *be as in* (3.8). *Then, for all* $k \geq 1$,

$$\|x_{k-1} - x_k\|^2 + \gamma^2 \|y_{k-1} - y_k\|^2 \leq \left(1 - \frac{2\gamma\eta}{(1+\gamma L)^2 - 2\gamma(L-\eta)}\right)^{k-1} d_0^2, \tag{3.20}$$

$$\|\tilde{x}_k - P_V(\tilde{x}_k)\|^2 + \gamma^2 \|u_k - P_{V^\perp}(u_k)\|^2 \leq \left(1 - \frac{2\gamma\eta}{(1+\gamma L)^2 - 2\gamma(L-\eta)}\right)^{k-1} d_0^2, \tag{3.21}$$

$$\|x^* - x_k\|^2 + \gamma^2 \|u^* - y_k\|^2 \leq \left(1 - \frac{2\gamma\eta}{(1+\gamma L)^2 - 2\gamma(L-\eta)}\right)^{k} d_0^2. \tag{3.22}$$

Proof: First, note that, from (1.8), (x^*, u^*) is a solution of (3.1) if and only if $x^* + \gamma u^* =: z^* \in (\gamma T)_V^{-1}(0)$. Using the last statement in Proposition 3.0.2, Proposition 3.0.1 to the operator γT (which is $(\gamma\eta)$-strongly monotone and (γL)-Lipschitz continuous) and Proposition 1.3.1, we conclude that the inequalities (1.31) and (1.32) hold with z^* as above, $\lambda = 1$, z_k as in (3.16) and

$$\mu = \frac{\gamma\eta}{1 + (\gamma L)^2}.$$

Direct calculations yield (recall that $\lambda = 1$)

$$\frac{2\mu}{1+2\mu} = \frac{2\gamma\eta}{(1+\gamma L)^2 - 2\gamma(L-\eta)}.$$

Hence, (3.20) and (3.22) follow from (1.31) and (1.32), respectively. To finish the proof, it remains to prove (3.21). To this end, note that it follows from (3.16), (3.17), (3.20) and the facts that $P_{V^\perp} = I - P_V$ and $P_V = I - P_{V^\perp}$. ■

Remarks.

(i) Analogously to first remark after Theorem 3.0.1, one can easily verify that $\gamma = 1/L$ provides the best convergence speed in (3.20)–

(3.22), in which case we find, respectively,

$$\|x_{k-1} - x_k\|^2 + \gamma^2 \|y_{k-1} - y_k\|^2 \leq \left(1 - \frac{\eta}{\eta + L}\right)^{k-1} d_0^2, \quad (3.23)$$

$$\|\tilde{x}_k - P_V(\tilde{x}_k)\|^2 + \gamma^2 \|u_k - P_{V^\perp}(u_k)\|^2 \leq \left(1 - \frac{\eta}{\eta + L}\right)^{k-1} d_0^2, \quad (3.24)$$

$$\|x^* - x_k\|^2 + \gamma^2 \|u^* - y_k\|^2 \leq \left(1 - \frac{\eta}{\eta + L}\right)^{k} d_0^2. \quad (3.25)$$

(ii) Since $L \geq \eta$, it follows that the above estimates are potentially better (specially to ill-conditioned problems) than the optimal one obtained to the SPDG algorithm in [25] via fixed point techniques, namely (3.9). The same remark applies to the rates in (3.20)–(3.22), when compared to the corresponding one in (3.7). See Figure 3.1.

(iii) Note that (3.24) (resp. (3.25)) imply that, for a given tolerance $\rho > 0$, the SPDG algorithm finds a pair (x, u) (resp. $(x, \gamma y)$) satisfying the termination criterion (3.4) (resp. $\|x^* - x\|^2 + \gamma^2 \|u^* - y\|^2 \leq \rho$) after performing no more than

$$2 + \log\left(\frac{\eta + L}{L}\right)^{-1} \log\left(\frac{d_0^2}{\rho}\right) \quad (3.26)$$

iterations.

(iv) By taking $L = 57$ and $\eta = 9$ (see p. 462 in [25]), we find $\log\left(\frac{2L}{2L - \eta}\right)^{-1} \approx 13$ and $\log\left(\frac{\eta + L}{L}\right)^{-1} \approx 7$ in (3.10) and (3.26), respectively. This shows that the upper bound (3.26), obtained in this thesis, is more accurate than the corresponding one (3.10), obtained in [25].

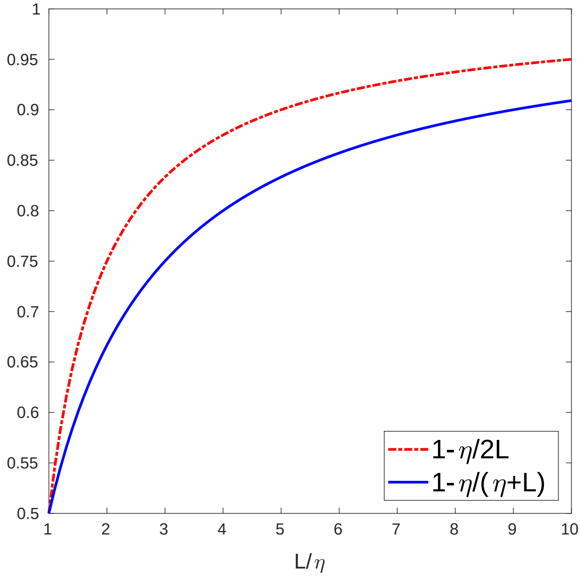

Figure 3.1: Solid line: see the convergence rates (3.23)–(3.25); dotted line: see (3.9).

Chapter 4

Final remarks

In this thesis, we proposed and analyzed the iteration-complexity of an inexact version of the Spingar's partial inverse method and, as a consequence, we obtained the iteration-complexity of an inexact version of the Spingarn's operator splitting method as well as of a parallel forward-backward method for multi-term composite convex optimization. We proved that our method falls in the framework of the hybrid proximal extragradient (HPE) method, for which the iteration-complexity has been obtained recently by Monteiro and Svaiter. We introduced a notion of approximate solution for the Spingarn's problem (which generalizes the one introduced by Monteiro and Svaiter for monotone inclusions) and proved the iteration-complexity for the above mentioned methods based on this notion of approximate solution. We also proved that the SPDG algorithm can alternatively be analyzed within the Spingarn's partial inverse framework, instead of the fixed point approach proposed in [25]. We proved that under the assumptions of [25], namely strong monotonicity and Lipschitz continuity, the Spingarn's partial inverse of the underlying maximal monotone operator is strongly monotone as well. This allowed us to employ recent developments in the convergence analysis and iteration-complexity of proximal point type methods for strongly monotone operators. By doing this, we additionally obtained a potentially better convergence speed for the SPDG algorithm as well as a better upper bound on the number of iterations needed to achieve prescribed tolerances.

References

[1] Alghamdi, M. A., Alotaibi, A., Combettes, P. L., Shahzad, N.: A primal-dual method of partial inverses for composite inclusions. Optim. Lett. 8(8), 2271–2284 (2014).

[2] Alves, M. Marques, Lima, S. C.: An inexact Spingarn's partial inverse method with applications to operator splitting and composite optimization. J. Optim. Theory Appl., 175(3): 818–847, 2017 (doi:10.1007/s10957-017-1188-y).

[3] Alves, M. Marques, Lima, S. C.: On the convergence rate of the scaled proximal decomposition on the graph of a maximal monotone operator (SPDG) algorithm. *arXiv preprint arXiv:1711.09959*, 2017.

[4] Alves, M. Marques, Monteiro, R. D. C., Svaiter, B. F.: Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds. SIAM J. Optim., 26(4): 2730–2743 (2016).

[5] Bauschke, H. H., Boţ, R. I., Hare, W. L., Moursi, W. M.: Attouch-Théra duality revisited: paramonotonicity and operator splitting. J. Approx. Theory 164(8), 1065–1084 (2012).

[6] Bauschke, H. H., Combettes, P. L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011).

[7] Boţ, R. I., Csetnek, E. R.: A hybrid proximal-extragradient algorithm with inertial effects. Numer. Funct. Anal. Optim. 36(8), 951–963 (2015).

[8] Briceño-Arias, L. M.: Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions. Optimization 64(5), 1239–1261 (2015).

[9] Briceño-Arias, L. M.: Forward-partial inverse-forward splitting for solving monotone inclusions inverse method for solving monotone inclusions. J. Optim. Theory Appl., vol. 166, pp. 391-413, (2015).

[10] Briceño-Arias, L. M., Combettes, P. L.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. 21(4), 1230–1250 (2011).

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

[12] Burachik, R. S., Sagastizábal, C.A., Scheimberg, S.: An inexact method of partial inverses and a parallel bundle method. Optim. Methods Softw. 21(3), 385–400 (2006).

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

[14] Ceng, L. C., Mordukhovich, B. S., Yao, J. C.: Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. J. Optim. Theory Appl. 146(2), 267–303 (2010).

[15] Combettes, P. L.: Iterative construction of the resolvent of a sum of maximal monotone operators. J. Convex Anal. 16(3-4), 727–748 (2009).

[16] Eckstein, J., Silva, P. J. S.: A practical relative error criterion for augmented Lagrangians. Math. Program. 141(1-2, Ser. A), 319–348 (2013).

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

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

[19] Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems Vol. I. Springer-Verlag, New York (2003).

[20] Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. Manuscript, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA, January 2016.

[21] He, Y., Monteiro, R. D. C.: An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. SIAM J. Optim. 26(1), 29–56 (2016).

[22] Idrissi, H., Lefebvre, O., Michelot, C.: Applications and numerical convergence of the partial inverse method, in Optimization, Lecture Notes in Math. 1405, S. Dolecki. Springer-Verlag, New York: 39–54 (1989).

[23] Iusem, A. N., Sosa, W.: On the proximal point method for equilibrium problems in Hilbert spaces. Optimization 59(8), 1259–1274 (2010).

[24] Lotito, P. A., Parente, L. A. Solodov, M. V.: A class of variable metric decomposition methods for monotone variational inclusions. J. Convex Anal. 16(3-4), 857–880 (2009).

[25] Mahey P., Oualibouch S., Tao P. D.: Proximal decomposition on the graph of a maximal monotone operator. SIAM J. Optim., 5(2): 454–466 (1995).

[26] Martinez-Legaz, J.-E., Théra, M.: ϵ-subdifferentials in terms of subdifferentials. Set-Valued Anal. 4(4), 327–332 (1996).

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

[28] Monteiro, R. D. C., Ortiz, C., Svaiter, B. F.: Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems. Comput. Optim. Appl. 57(1), 45–69 (2014).

[29] Monteiro, R. D. C., Ortiz, C., Svaiter, B. F.: An adaptive accelerated first-order method for convex optimization. Comput. Optim. Appl. 64(1), 31–73 (2016).

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

[31] Monteiro, R. D. C., Svaiter, B. F.: On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. SIAM Journal on Optimization 20, 2755–2787 (2010).

[32] Monteiro, R. D. C., 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 Journal on Optimization 21, 1688–1720 (2010).

[33] Monteiro, R. D. C., Svaiter, B. F.: Convergence rate of inexact proximal point methods with relative error criteria for convex optimization. *Manuscript*, 2010.

[34] Nesterov, Y.: Introductory Lectures on Convex Optimization. A Basic Course. Kluwer Academic Publishers, Boston (2004).

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

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

[37] Rockafellar, R. T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper., 1(2): 97–116 (1976).

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

[39] Rockafellar, R. T., Wets R.: Variational Analysis. Springer Verlag, Berlin 1998.

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

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

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

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

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