UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA PURA E APLICADA

Raul Tintaya Marcavillaca

On relative-error inertial-relaxed inexact proximal algorithms for monotone inclusion problems

Florianópolis
2020

Raul Tintaya Marcavillaca

On relative-error inertial-relaxed inexact proximal algorithms for monotone inclusion problems

Tese apresentada ao Curso de Pós- Graduação em Matemática Pura e Aplicada do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina, para a obtenção do grau de Doutor em Matemática, com Área Concentração em Matemática Aplicada.
Orientador. Prof. Maicon Marques Alves, Dr.

Florianópolis
2020

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

Tintaya Marcavillaca, Raul
On relative-error inertial-relaxed inexact proximal algorithms for monotone inclusion problems / Raul Tintaya Marcavillaca ; orientador, Maicon Marques Alves, 2021.
104 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Programa de Pós-Graduação em Matemática Pura e Aplicada, Florianópolis, 2021.

Inclui referências.

1. Matemática Pura e Aplicada. 2. Otimização convexa. Operadores monótonos maximais. 3. Método ponto proximal inexato. 4. Métodos projectivos de decomposição. 5. Inercia e relaxação. I. Marques Alves, Maicon. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Matemática Pura e Aplicada. III. Título.

Raul Tintaya Marcavillaca

On relative-error inertial-relaxed inexact proximal algorithms for monotone inclusion problems

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Alfredo Noel Iusem, Dr.
Instituto de Matemática Pura e Aplicada - IMPA

Prof. Paulo José da Silva e Silva, Dr.
Universidade Estadual de Campinas - UNICAMP

Prof. Antonio Carlos Gardel Leitão, Dr.
Universidade Federal de Santa Catarina - UFSC

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

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Matemática, com área de concentração em Matemática Aplicada.

ICP-Edu
Documento assinado digitalmente
Daniel Goncalves
Data: 12/02/2021 11:29:59-0300
CPF: 017.022.459-77

Prof. Daniel Gonçalves, Dr.
Coordenador do Programa de Pós-Graduação

ICP-Edu
Documento assinado digitalmente
Maicon Marques Alves
Data: 14/02/2021 14:33:50-0300
CPF: 967.631.910-49

Prof. Maicon Marques Alves, Dr.
Orientador

Florianópolis, 2020

“Faça as coisas o mais simples que puder,
porém não as mais simples”.

Albert Einstein.

Agradecimentos

Foram muitas pessoas que contribuiram para eu chegar até aqui:

Esta tese é dedicada à minha família: meus pais, Eusebio e Gerarda; irmãos, Eber, Yackelin, Charmely e Lisbeth; minha tia Plácida e prima Miriam. Obrigado por me motivar e incentivar a continuar crescendo acadêmica e profissionalmente, por todo o apoio emocional, e por todos os momentos que passamos juntos.

O meu mais profundo agradecimento ao meu orientador Prof. Dr. Maicon Marques Alves, pelo seu tempo e paciência; sem ele, esta tese não teria sido possível. Seu valioso feedback contribuiu muito para desenvolvimento de este trabalho. Agradeço a ele pela amizade, por suas percepções e sugestões que ajudaram a moldar minhas habilidades de pesquisa.

Agradeço aos professores membros da comissão examinadora: Alfredo Iusem, Antonio Gardel Leitão, Paulo José da Silva e Douglas Soares Gonçalves. Obrigado pelos comentários úteis e sugestões que ajudaram a melhorar esta tese. Meu especial agradecimento ao Professor Alfredo por todo o apoio e as recomendações.

Agradeço aos meus colegas da sala 107: Marina, Joel, Javi e Rubens pelas boas conversas e momentos de descontração nas horas do café. Em especial à minha irmã acadêmica Marina, pela amizade e por todas as boas e agradavéis conversas durante esse período de doutorado.

Agradeço a todos os amigos colombianos que conheci em Florianópolis: Cindy, Sebastian, Paula (catalã - colombiana), Felipe e Alex "Pelucas" por me fazerem sentir como em casa apesar de estar longe dela. Em especial, sou grato à minha namorada Marduk, por todo o amor, incentivo e companheirismo durante esses anos.

Agradeço aos meus amigos e professores da facultade da UNSAAC no Perú, Guido Alvarez, Alejandro Tito, José Moso, Ignacio Velasques e Patricio Choque pelos primeiros ensinamentos em matemática e pela acolhida atenciosa que sempre me mostram cada vez que volto lá. Não posso deixar mecionar aos meus queridos amigos e colegas Alain, Gino, Carmen, Lisbeth, Shirley e Edison Niño, obrigado pela amistade e momentos agradáveis.

Por fim, gostaria de agradecer a todos os professores e funcionários do departamento de matemática da UFSC. Em particular, gostaria de agradecer ao Professor Ruy Charão pela amizade e motivação para superar as adversidades e continuar estudando. Agradeço também a Elisa e Erica pela forma amável e competente com que sempre me receberam na secretaria da pos-graduação.

A todos aqueles que de alguma forma contribuíram para para o desenvolvimento de este trabalho, que por algum motivo não foram mencionados aqui, meus sinceros agradecimentos.

Agradeço à CAPES pelo apoio financeiro.

Resumo

Neste trabalho, propomos e estudamos uma versão inercial subrelaxada e com erro relativo do método proximal extragradiente (HPE) de Sodolov e Svaiter para resolver problemas de inclusão monótono. Estudamos a convergência assintótica do método, bem como suas taxas de convergência não-assintótica global em termos de complexidade computacional em número de iterações. Analisamos o novo método sob condições mais flexíveis do que as existentes na literatura, tanto nos parâmetros de extrapolação quanto de erro relativo. A nova abordagem é aplicada a dois tipos de métodos do tipo "forward-backward" para resolver inclusões monótonas com determinada estrutura.

Além disso, para resolver problemas de inclusão monótono envolvendo soma finita de operadores monótonos maximais, propomos e estudamos uma versão inercial relaxada om erro relativo do método "projective splitting method (PSM)" de Eckstein e Svaiter. O algoritmo proposto se beneficia de uma combinação de efeitos inerciais e de relaxação, controlada por parâmetros dentro de uma determinada faixa. Propomos condições suficientes sobre esses parâmetros (também estudamos a interação entre eles) para garantir a convergência fraca das sequências geradas por nosso algoritmo. Como uma aplicação do algoritmo proposto, derivamos um algoritmo inercial semelhante ao método "alternating direction method of multipliers method (ADMM)" com multiplos blocos.

Palavras-chaves:Métodos inerciais e relaxados. Método de ponto proximal inexato. Método HPE. Algoritmos de decomposição. Algoritmos projetivos de decomposição. Algoritmos do tipo "forward-backward". ADMM multibloco. Erro-relativo. Complexidade em iteração. Taxas de convergência.

Resumo Expandido

Introdução

Seja $\mathcal{H}$ um espaço de Hilbert com produto interno $\langle \cdot, \cdot \rangle$ e $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ a norma induzida pelo produto interno. 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 monótono maximal. Devido à generalidade matemática dos operadores monótonos maximais, o problema (1) é muito inclusivo e serve como um modelo unificado para muitos problemas de importância fundamental, tais como, problemas de ponto fixo, problemas de desigualdade variacional problemas de minimização convexa, e suas extensões.

O método iterativo mais popular para resolver (1) aproximadamente é o método de ponto proximal (PP), proposto inicialmente por Martinet(1970) no contexto de otimização convexa, e posteriormente generalizado por Rockafellar(1976) para um contexto mais geral de operadores monótonos maximais. Os métodos PP com erro-relativo são variantes inexactas do método PP que permite uma tolerância de erro-relativo na solução aproximada de subproblemas proximais. O Método Híbrido Proximal Extragradiente (HPE) proposto por Solodov e Svaiter (1999) corresponde a essa família de métodos. Monteiro e Svaiter (2010) estabeleceram a complexidade computacional para o método HPE. Em muitas aplicações, os métodos PP na forma clássica não são muito eficientes. Os avanços que visam acelerar a convergência de métodos proximais enfocam, entre outras abordagens, as formas de incorporar informações de segunda ordem para alcançar uma convergência mais rápida. Para esse fim, Álvarez e Attouch (2001) propuseram o método PP inercial obtido pela discretização no tempo de um sistema dinâmico dissipativo de segunda ordem. A partir desse trabalho, os métodos inerciais têm sido um foco de estudo explorado por muitos pesquisadores.

Neste trabalho também consideramos problemas de inclusão monótono da forma

$$0 \in G_1^* T_1(G_1 z) + \ldots + G_n^* T_n(G_n z) \tag{2}$$

onde (para cada $i = 1, \ldots, n$), $T_i : \mathcal{H} \rightrightarrows \mathcal{H}$ são operadores monótonos maximais e $G_i : \mathcal{H} \to \mathcal{H}$ são operadores lineares e contínuos. Problemas do tipo (2) aparece em diferentes campos da matemática aplicada e otimização, incluindo aprendizado de máquinas, problemas inversos e processamento de imagens. Uma estratégia muito popular para encontrar soluções aproximadas de (2) são os métodos de decomposição (ou de divisão) que remonta ao desenvolvimento de alguns esquemas numéricos bem conhecidos como o método de Douglas-Rachord, método das inversas parciais e entre outras. A principal característica desta abordagem é que em cada iteração é usado a informação individual de cada operador T_i e G_i. A família de métodos projetivos de decomposição (SPM) que foi introduzida recentemente por Eckstein e Svaiter (2009) tem ganhado notoriedade nos últimos anos, devido a sua flexibilidade comparado com outro tipo de métodos de decomposição, no que diz respeito aos parâmetros e à ativação de cada operador T_i (usando o resolvente de cada uma de elas) e G_i durante o processo iterativo. O método proposto é baseado na reformulação de (2) como um problema de viabilidade convexo definido por um convexo fechado (conjunto de soluções estendidas do problema (2)) para o qual um hiperplano separador é construído por avaliação individual do resolvente de cada operador.

Objetivos

- Propor uma versão inercial relaxada do método HPE para resolver (1); estudar sua convergência assintótica e não assintótica(taxas de convergência não assintótica e termos da complexidade computacional) sobre certas condições nos parámetros de inercia e relaxação. Como uma aplicação, se busca usar a nova abordagem para estudar dois tipos de métodos do tipo forward-backward para quando T é considerado como soma de um monótono ponto-ponto cocoercivo (ou Lipschitz contínuo) com outro operador ponto-conjunto monótono maximal.

- Propor uma versão inercial relaxada do método projetivo de decomposição (SPM) para resolver um problema de inclusão monótono do tipo (2); propor condições suficientes sobre os parâmetros de inercia e relaxação, assim como também estudar a interação entre eles, para analisar a convergência fraca do novo método proposto. Como uma aplicação se busca derivar um algoritmo inercial semelhante ao método de direção alternada de multiplicadores (ADMM) com múltiplos blocos.

Metodologia

Através de uma revisão bibliográfica, observamos que: (i) A combinação de efeitos de inércia e relaxação que é uma estratégia puramente algébrico tem sido bastante estudado nos últimos anos, pois fornecem uma forma de aceleração de métodos numéricos para problemas de minimização convexa e problemas de inclusão monótono. Mais precisamente, as técnicas de relaxação combinan a saída da operação(por exemplo, o operador (etapa) proximal ou gradiente para o caso de minimização convexa) com a iteração anterior controlado por um parâmetro de relaxação, em que a sobre-relaxação é conhecido por melhorar a convergência do algoritmo; enquanto a estratégia inercial é dado antes da operação mediante um passo de extrapolação usando como informação os últimos dois iterados. (ii) A solução de sub-problemas de forma inexacta usando critérios de erro-relativo podem melhorar substancialmente o desempenho do método comparado com a solução de forma exacta. Cabe mencionar que o método HPE se caracteriza por usar esse tipo de critérios na solução aproximada de subproblemas internos, e ganharam notoriedade devido à sua robustez como um framework para o projeto e análise de vários algoritmos concretos de inclusão monótona, desigualdades variacionais, ponto de sela e problemas de otimização convexa. Motivado pela discussão acima, neste trabalho buscamos combinar esses três ingredientes, inércia, relaxação e erro-relativo para propor métodos iterativos inexatos con inércia e relaxação na procura de soluções aproximadas para problemas do tipo (1) e (2).

Resultados e Considerações Finais

Como resultado de nossa pesquisa, propomos e estudamos a convergência do método HPE inercial e sub-relaxado para encontrar soluções aproximadas de (1). Estabelecemos a complexidade computacional em termos do número de iterações(número de iterações necessárias para alcançar tolerâncias prescritas) para nosso método proposto. Além disso, desde que o problema (2) pode ser visto como um problema de viabilidade convexa como mencionamos anteriormente, propomos um método iterativo inercial que resolve este problema de viabilidade convexa, e estudamos sua convergência; como aplicação de este último, propusemos um método inercial com relaxação e erro-relativo que resolve (2). Finalmente como subproducto derivamos um método com inércia e relaxação parecido ao método de direção alternada de multiplicadores (ADM) com múltiplos blocos. Também foram estabelecidas condições suficientes sobre os parâmetros para garantir a convergência de esses algoritmos propostos. Cabe mencionar que todos os resultados mostrados neste trabalho com a excepção de aquelas citados, são originais e contribuem ao desenvolvimento de novos métodos que sejam rápidos no sentido de acelerar a convergência, flexíveis, e de fácil implementação.

Palavras-chave: Métodos inerciais e relaxados. Método de ponto proximal inexato. Método HPE. Algoritmos de decomposição. Algoritmos projetivos de decomposião. Algoritmos de tipo forward-backward. ADMM multibloco. Erro-relativo. Complexidade em iteração. Taxas de convergência.

Abstract

We propose and study an inertial under-relaxed version of the relative-error hybrid proximal extragradient (HPE) method of Sodolov and Svaiter for solving monotone inclusion problems. We study the asymptotic convergence of the method, as well as its nonasymptotic global convergence rates in terms of iteration-complexity. We analyze the new method under more flexible assumptions than existing ones, both on the extrapolation and on the relative-error parameters. The approach is applied to two types of forward-backward methods for solving structured monotone inclusions.

For solving structured monotone inclusion problems involving the sum of finitely many maximal monotone operators, we propose and study an inertial-relaxed version of Eckstein and Svaiter projective splitting method. The proposed algorithm benefits from a combination of inertial and relaxation effects, which are both controlled by parameters within a certain range. We propose sufficient conditions on these parameters (as well as we study the interplay between them) to ensure weak convergence of sequences generated by our algorithm. As an application of the proposed algorithm we derive an inertial algorithm resembling the multi-block alternating direction method of multipliers (ADMM).

Keywords: Inertial and relaxed methods. Inexact proximal point methods. HPE method. Splitting algorithms. Projective splitting method. Forward-backward type algorithms. Multi-block ADMM. Relative-error. Iteration-complexity. Convergence rates.

Table of contents

Basic notation and terminology

$\mathcal{H}, \mathcal{H}_0, \ldots, \mathcal{H}_n$: real Hilbert spaces

$\boldsymbol{\mathcal{H}} := \mathcal{H}_0 \times \ldots \times \mathcal{H}_{n-1}$: product space

$\langle \cdot, \cdot \rangle$: inner product

$\| \cdot \|$: norm induced by the inner product $\langle \cdot, \cdot \rangle$

$\mathbb{R}^n$: the n-dimensional Euclidian space

$T : \mathcal{H} \rightrightarrows \mathcal{H}$: set-valued or point-to-set operator

$\operatorname{dom} T$: the domain of T

$\operatorname{im} T$: the image or range of T ($\operatorname{range} T$)

$G(T)$: the graph of T

$f : \mathcal{H} \to (-\infty, +\infty]$: extended-real valued function

$\operatorname{dom}(f)$: the domain of f

∂f: the subdifferential of a convex function f

f^*: Fenchel-Legendre conjugate of f

∇f: gradient of a differentiable function f

$N_C(x)$: the normal cone of a set C at a point x

$\iota_C(x)$: the indicator function of a set C

$ri(D)$: relative interior of a set D

L^*: the adjoint operator of a linear and bounded operator L

M^T: the transpose of a matrix M

P_C: the projection operator onto a nonempty subset C of $\mathcal{H}$.

Introduction

Maximal monotone operators appear in several branches of applied mathematics such as optimization, partial differential equations, control theory, mathematical economics and variational analysis. These operators have been object of intense research between 1960s and 1980s, when Kachurovskii, Brezis, Browder, Minty and Rockafellar established the fundamental results about them. The notion of monotone operator was first formulated and studied in [72, 86]. Much of the initial work was done in the context of functional analysis and partial differential equations (see e.g., [28, 29, 73]), but it was soon noticed the relevance of the theory in convex analysis and convex optimization (see [87, 108, 110]).

A monotone inclusion problem (MIP) consists in finding the root of a maximal monotone operator, i.e., $0 \in T(z)$ where T is maximal monotone. The MIP is motivated by the fact that optimality conditions for convex optimization problems that meet a regularity condition can be expressed as monotone inclusion problems. Furthermore, the investigations performed in this more general setting of (maximal) monotone operators bring new insights when considering the problem of solving complicated nondifferentiable convex optimization problems involving finite sums, compositions with linear operators or infimal convolutions. Moreover, due to its applications in the theory of nonlinear partial differential equations, variational inequalities and specially in optimization theory (see e.g., [20, 26, 55, 73]), the study of monotone inclusion problems continues to attract a large group from the mathematical community.

One of the most popular algorithms for finding approximate solution for monotone inclusion problems is the proximal point algorithm (PPA) (or proximal point (PP) method), proposed by Martinet [84] (1970) in the context of monotone variational inequalities(with point-to-point operators) and then popularized by Rockafellar in [112] for general maximal monotone operators (1976). Actually, the term "proximal point" was originally coined early by Moreau in 1962 (see [94]), but it gained notoriety after Rockafellar's work. Although the PPA has good global convergence properties [112], the major drawback is that it requires the evaluation of the resolvent mapping $(\lambda T + I)^{-1}$. The difficulty lies in the fact that inverting the operator $\lambda T + I$, which can be as difficult as solving the original problem. One alternative to attenuate this drawback is to decompose the operator as a sum of two or more maximal monotone operators in such a way that their resolvents are relatively easier to calculate. Thereby, one can create methods that use independently these proximal mappings.

Among other results, Rockafellar in [112] proposed an inexact version of the PPA based on a summable absolute error criterion. The error criterion considered by Rockafellar involves a sequence of errors whose sum is finite, which can present some practical disadvantages in specific problems, as were pointed in [33, 39, 106] because this error criterion does not indicate how they are chosen. Hence, it turns out to be relevant to develop error conditions for approximating proximal subproblems that can be computable during the progress of the iterates. In the last two decades, as an alternative to inexact Rockafellar's PPA, *relative-error methods* have deserved the attention of several researchers (see, e.g., [7, 22, 34, 92, 120, 121]). This type of inexact PPA, which started with the pioneering works of Solodov and Svaiter [118, 117], is widely used both in the theory and practice of numerical optimization (see, e.g., [34, 60, 64, 91, 92]). The first method proposed by Solodov and Svaiter is the hybrid projection proximal point (HPP) method [118]. The main characteristic of the HPP method is the combination of a relative-error criterion and projective steps. Indeed, at each iteration, a hyperplane is constructed that strictly separates the current iterate from the solution set (which is assumed to be nonempty) and then the current iterate is projected onto this hyperplane (that is constructed using points in the graph of the operator). The second type of inexact PPA proposed by Solodov and Svaiter is the hybrid proximal extragradient (HPE) method [117]. This general framework has a different mechanism of iteration than the HPP method and combines steps of the proximal and Korpolevich's extragradient [74] methods. One of the most important characteristics of these two approaches is that they allow a significant relaxation of tolerance requirements imposed on the solution of proximal subproblems. This yields a more practical framework based on the proximal algorithms.

We mention now that if the MIP consists in finding zeros of a sum with more than two maximal monotone operators, since in general there exists no closed formula for the resolvent of the sum of operators in terms of their resolvents, then it follows that the PPA applied directly to this sum is not suitable from an implementation point of view. The operator splitting algorithms overcome this drawback, where the term "splitting" is used in order to stress out that in the iterative scheme the operators involved are evaluated separately. In particular, if we consider a MIP as a sum of two maximal monotone operators, we mention the Peaceman-Rachford [101], Douglas-Rachford [67, 76, 49], forward-backward [76, 100] and Tseng's forward-backward-forward [128] methods as the most popular algorithms to solve this type of problems. We also mention that very recently, David and Yin proposed a three-operator splitting method in [46]. For any arbitrary sum of maximal monotone operators, we mention the operator splitting method proposed by Spingarn [123], where at each iteration, the solution of subproblems is performed in parallel.

The primal-dual splitting methods are also splitting methods that deal with inclusion problems where some complex structure of monotone operators is involved, such as mixtures of linearly composed and parallel-sums. The main difficulties in applying directly these methods are due to the fact that the resolvent of such compositions cannot be expressed in a closed manner (except in some very restrictive cases), see [24, 27, 35, 44, 45] for further considerations concerning this class of algorithms. The key feature of these algorithms is that they are fully decomposable, in the sense that each of the operators is evaluated separately in the algorithm, either via forward or via backward steps. It is also worth mentioning that primal-dual algorithms solve simultaneously a (primal) monotone inclusion problem and its dual monotone inclusion problem in the sense of Attouch-Thera [18]. For instance, [27] deals with monotone inclusion problems involving sums of compositions with bounded linear operators by rewriting the original monotone inclusion problem as the sum of a maximal monotone operator and a linear and skew-adjoint operator in an appropriate product space, and then applies Tseng's splitting algorithm [128] to develop a new algorithmic framework.

Recently, a new family of operator splitting methods for solving MIPs involving sums of maximal monotone operators was introduced in [51, 52] by Eckstein and Svaiter. The proposed method is based on reformulating the MIP as a convex feasibility problem defined by a closed convex (extended solution set) for which a separating hyperplane is constructed by individual evaluation of the resolvent of each operator. The resulting algorithm is essentially a projective method, in the sense that in each iteration a hyperplane is constructed separating the currently iterate to the extended solution set (defined in an appropriate product space) and then the next iterate is calculated by projecting the current iterate onto this hyperplane. Nowadays this family of operator splitting methods is known as *projective splitting methods* (PSM). We refer the reader to [2, 42, 48, 68, 71, 70, 69, 79, 80] for some recent contributions on this subject.

Inertial methods for solving monotone inclusion and optimization problems, with roots in (time) implicit discretization of second-order differential equations, gained a lot of attention in nowadays research (see, e.g., [3, 9, 11, 13, 16] and the references therein). After discretizing the second-order dynamical system one obtains an iterative scheme where the next iterate depends of the two previous iterates. The early example of such methods is due to Polyak [105], who introduced the so-called *heavy ball method* for minimizing a strongly convex quadratic function which can greatly improve upon the convergence speed of the gradient method (see also [104], p. 65). In [5], Alvarez and Attouch translated the idea of the heavy ball method to the general setting of a maximal monotone operators using the scheme of the PPA, resulting in an algorithmic scheme named *inertial proximal point algorithm* (IPPA). Since then, we notice an increasing interest of the optimization community in the class of first-order proximal algorithms with inertial effects, including inertial versions of ADMM, Douglas-Rachford, forward-backward and Tseng's modified forward-backward methods (see, [10, 36, 38, 77, 96]), as well as inexact versions of them (see, e.g., [4, 6, 8, 22]). The intense research activity on inertial methods in the last years is in part due to its links with fast first-order methods for convex (also for non-convex) optimization problems (see, e.g., [13, 17, 21, 99]). Let us mention also the fast gradient method of Nesterov [97, 98] and the FISTA [21] of Beck and Teboulle, which are iterative schemes involving inertial steps and have fast convergence rates in function values (in the worst case).

Motivation and goals

In the theory of monotone operators, an important topic is the study of iterative methods for solving monotone inclusion problems

$$0 \in T(z). \tag{3}$$

In this thesis, we are concerned with the monotone inclusion problem (3). We are also interested in the case where T can be written as a sum of maximal monotone operators composed with bounded linear operators

$$0 \in \sum_{i=1}^{n} G_i^* T_i(G_i z). \tag{4}$$

It is known that a way to speed up the convergence rate of algorithms in optimization and monotone inclusions consists in constructing the next iterate by combining the information of the previous iterations by introducing relaxation and inertia.

Relaxation and Inertia. Relaxation techniques have proven to be an essential ingredient in the formulation of algorithms for monotone inclusions, as they provide more flexibility to the iterative scheme (see [20, 49]) and have the property of speeding up the algorithm similar to the inertia. An important issue is the study of the interplay between relaxation and inertia, which will be one of the main topics of this thesis. Without using inertia, over-relaxation provides a natural way to speed up algorithms. By contrast, for the solution of monotone inclusions by inertial methods, we will see that under-relaxation allows to balance inertial and relaxation effects (see also [4, 11, 10, 12, 16, 81, 66]).

Motivated by the above discussion, our main goal in this thesis is to propose inexact methods with inertial and relaxation effects for finding zeros of the inclusion problems (3) and (4). We will focus our attention in the HPE method [92] and the PSM [51, 52], for solving (3) and (4) respectively, by introducing a new and extra step, named extrapolated or inertial step at each iteration of such algorithms. The proposed algorithmic frameworks unify the basic ideas of the IPPA with the HPE method and PSM, respectively.

Main contributions. We summarize the main contributions of this thesis are as follows.

(a) Contributions in Chapter 2. We propose and study an inertial and relaxed version of the HPE method of Sodolov and Svaiter [117] for solving the inclusion monotone problem (3).We also study the asymptotic convergence and nonasymptotic global $\mathcal{O}(1/\sqrt{k})$ *pointwise* and $\mathcal{O}(1/k)$ *ergodic* convergence rates (iteration-complexity) of the proposed algorithm (Algorithm 2). As applications, we established asymptotic convergence and pointwise and ergodic iteration-complexity of inertial under-relaxed versions of the Tseng's modified forward-backward method (Algorithm 4) and forward-backward method (Algorithm 5) for find zeros of sum of two maximal monotone, i.e., $0 \in F(z)+B(z)$ under the assumption that F is monotone and either Lipschitz continuous or cocoercive and B is maximal monotone operator.

(b) Contributions in chapter 3. We propose an inertial-relaxed and inexact version of the projective splitting method introduced by Eckstein and Svaiter in [51, 52] for solving (4). We establish asymptotic convergence and nonasymptotic pointwise global convergence rates for the proposed algorithm (Algorithm 7). The analysis is established by viewing Algorithm 7 within a general framework (Algorithm 6) for solving the (feasibility) problem of finding points in closed convex subsets of Hilbert spaces, where the joint adjustment of inertia and relaxation parameters plays a central role. We propose sufficient conditions on these parameters (as well as we study the interplay between them) to ensure weak convergence of sequences generated by the framework. Finally, as an application of Algorithm 7 we derive an inertial algorithm (Algorithm 8) resembling the multi-block ADMM.

Outline of the thesis. This thesis is organized in four chapters as follows. Chapter 1 contains five sections. In Section 1.1, we present notations and basis results. Section 1.2 reviews some definitions and facts on maximal monotone operators and convex analysis that will be used along this work. In Section 1.3, we briefly review monotone inclusions, proximal point and splitting methods, and in Section 1.5 we give a brief overview on inertial methods. Chapter 2 is devoted to the HPE method with inertial effects for solving monotone inclusion problems. It contains three sections as follows. In Section 2.1, we present some preliminaries and basic results, review some existing algorithms and discuss in detail the main contributions of this chapter. The inertial under-relaxed HPE method (Algorithm 2) is presented in Section 2.2; asymptotic convergence and iteration-complexity analysis are discussed in this section. Sections 2.3 is devoted to present and study the inertial versions of the Tseng's modified forward-backward and forward-backward algorithms; convergence results are also presented. Chapter 3 presents an inertial and inexact version of the PSM for solving structured monotone inclusion problems involving the sum of (finite) many maximal monotone operators. It is divided into four sections. Section 3.1 reviews some important facts on such problems and provides a brief motivation for the projective splitting method. In section 3.2, it is stated an inertial-relaxed separator-projector method for solving the (feasibility) problem of finding points in closed convex subsets of a Hilbert space (Algorithm 6) and its weak convergence is analyzed. In Section 3.3, it is presented the main contribution of this chapter, the inertial-relaxed (inexact) projective splitting method (Algorithm 7) as a specialization of Algorithm 6. We also prove weak convergence, exploiting the analysis in the preceding section. Finally, in Section 8, we state an application of the preceding section; we derived an inertial multi-block ADMM-like method (Algorithm 8) for a structured and constrained convex optimization problem. Finally, in Chapter 4, we will discuss the main results and contributions of this thesis to (inexact) inertial-relaxed methods for solving monotone inclusion problems and we will present some proposals for future works. More details on goals, results achieved and contributions will be also discussed in each chapter.

Chapter 1

Preliminaries

In this chapter, we present some basic results. First, we review some important facts and definitions of convex analysis, maximal monotone operators and ε–enlargements. We also address iterative methods for solving structured monotone inclusions problems, as well as its motivations, and finally we give an overview of the inertial methods. We also refer the reader to the literature on these topics [10, 20, 51, 52, 103, 129] and references therein.

1.1 Notation and basic results

Throughout this thesis we will denote by $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ real Hilbert spaces and let $\langle \cdot, \cdot \rangle_i$ and $\| \cdot \|_i = \sqrt{\langle \cdot, \cdot \rangle_i}$ be the inner product and its induced norm, for $i = 0, 1, \ldots, n$. For simplicity we will write $\| \cdot \| := \| \cdot \|_i$ and $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_i$, for all $i = 0, \ldots, n$. Assume that $\mathcal{H}_0 = \mathcal{H}_n$ and let $\boldsymbol{\mathcal{H}} := \mathcal{H}_0 \times \ldots \times \mathcal{H}_{n-1}$ be endowed with the inner product and norm defined, respectively, as follows (for some $\gamma > 0$):

$$\langle (z, w), (z', w') \rangle_\gamma = \gamma \langle z, z' \rangle + \sum_{i=1}^{n-1} \langle w_i, w_i' \rangle, \quad \|(z, w)\|_\gamma^2 = \gamma \|z\|^2 + \sum_{i=1}^{n-1} \|w_i\|^2,$$

where $z, z' \in \mathcal{H}_0$ and $w := (w_1, \ldots, w_{n-1}), w' := (w_1', \ldots, w_{n-1}') \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_{n-1}$.

The symbols $\rightharpoonup$ and $\to$ denote weak and strong convergence, respectively. For a bounded linear operator $L : \mathcal{H} \to \mathcal{G}$, where $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, the operator $L^* : \mathcal{G} \to \mathcal{H}$ denotes the adjoint operator of L. The norm of L is given by

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|_{\mathcal{G}}}{\|x\|_{\mathcal{H}}}.$$

For C a nonempty, convex and closed subset of $\mathcal{H}$, we define the orthogonal projection $P_C(x)$ of x onto C as the unique point $P_C(x)$ in C such that

$$\|P_C(x) - x\| \leq \|x - y\| \quad \text{for all} \quad y \in C,$$

i.e.,

$$P_C(x) = \arg\min\{\|x - y\| : y \in C\}.$$

The following is a well-known fact on the orthogonal projection:

$$z = P_C(x) \Leftrightarrow z \in C \ \text{ and } \ \langle x - z, y - z \rangle \leq 0, \quad \forall y \in C. \tag{1.1}$$

1.2 Monotone operators and convex analysis

An operator (or mapping) $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is called set-valued operator or point-to-set operator, if T maps every point $x \in \mathcal{H}$ to a subset of $\mathcal{H}$, i.e.

$$T : \mathcal{H} \rightrightarrows \mathcal{H} : \quad x \mapsto T(x) \subset \mathcal{H}.$$

The graph of T is defined by

$$G(T) := \{(x, v) \in \mathcal{H} \times \mathcal{H} \,|\, v \in T(x)\}.$$

The domain and range (or image) of T are defined respectively, as

$$\operatorname{dom} T := \{x \in \mathcal{H} \mid T(x) \neq \emptyset\} \quad \text{and} \quad \operatorname{im} T := \{v \in \mathcal{H} \mid \exists\, x \in \mathcal{H} : v \in T(x)\}.$$

The inverse of T is $T^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ such that $v \in T(x)$ if and only if $x \in T^{-1}(v)$, i.e.,

$$T^{-1}(v) := \{x \in \mathcal{H} \mid v \in T(x)\}.$$

For any $\gamma > 0$, the operator $\gamma T : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by

$$(\gamma T)x := \gamma T(x) := \{\gamma v \,|\, v \in T(x)\}.$$

The resolvent of $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined as

$$J_{\lambda T} := (I + \lambda T)^{-1} \quad \lambda > 0,$$

where $I : \mathcal{H} \to \mathcal{H}$ is the identity operator in $\mathcal{H}$.

Definition 1.2.1. A Set-valued operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is said to be

(i) monotone if, for any $(x, v),\ (x', v') \in G(T)$,

$$\langle x - x', v - v' \rangle \geq 0;$$

(ii) maximal monotone if it is monotone and maximal in the family of monotone operators in $\mathcal{H}$, with respect to the partial order of the inclusion, that is, if $S : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone and $G(T) \subseteq G(S)$, then $T = S$.

If T is maximal monotone, then $J_{\lambda T} : \mathcal{H} \to \mathcal{H}$ is single-value and maximal monotone, [20, Proposition 23.8, Corollary 23.11].

Definition 1.2.2. Let $f : \mathcal{H} \to (-\infty, +\infty]$ be an extended-real valued function.

(a) the domain and epigraph of f are defined, respectively, as

$$\operatorname{dom} f := \{x \in \mathcal{H} \mid f(x) < +\infty\}$$

and

$$\operatorname{epi} f := \{(x, \mu) \in \mathcal{H} \times \mathbb{R} \mid \mu \geq f(x)\}.$$

(b) f is proper if $\operatorname{dom} f \neq \emptyset$.

(c) f is convex, if for all $x, y \in \mathcal{H}$ and $t \in [0,1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

(d) f is lower-semicontinuos (or closed) if $\operatorname{epi}(f)$ is closed in $\mathcal{H} \times \mathbb{R}$.

The set of proper, lower-semicontinuous and convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$ is denoted by $\Gamma_0(\mathcal{H})$.

The subdifferential of a proper function $f : \mathcal{H} \to (-\infty, +\infty]$ is the set-valued (or point-to-set) operator $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ defined as

$$\partial f(x) := \begin{cases} \{v : f(y) \geq f(x) + \langle v, y - x \rangle \quad \forall y \in \mathcal{H}\} & \text{if} \quad x \in \operatorname{dom} f \\ \emptyset & \text{otherwise.} \end{cases}$$

The vector $v \in \mathcal{H}$ is called a subgradient of f at $x \in \mathcal{H}$ if $v \in \partial f(x)$. Let us mention that if f is proper, convex, and Gâteaux differentiable at $x \in \mathcal{H}$, then $\partial f(x) = \{\nabla f(x)\}$ (see e.g. [20, Proposition 17.31]). The operator ∂f is trivially monotone if f is convex and proper. In addition, if $f \in \Gamma_0(\mathcal{H})$ then ∂f is maximal monotone (see [110] or [20, Theorem 20.25]).

Given $\epsilon \geq 0$, the $\epsilon-$subdifferential of $f : \mathcal{H} \to (-\infty, +\infty]$ is defined at every $z \in \mathcal{H}$ as

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

The Fenchel-Legendre conjugate (or Fenchel conjugate, or Legendre–Fenchel transform) of a convex function f, denoted by $f^* : \mathcal{H} \to (-\infty, +\infty]$, is defined as

$$f^*(u) = \sup_{z \in \mathcal{H}} \left\{ \langle u, z \rangle - f(z) \right\}.$$

It is simple to see that f^* is convex and closed function. Furthermore. if f is proper, closed and convex, f^* is a proper function and holds the property:

$$u \in \partial f(x) \Leftrightarrow x \in \partial f^*(u).$$

The Fermat's rule in the nonsmooth case underlines the usefulness of the subdifferential: if f is proper, then for $x \in \operatorname{dom} f$ one have

$$x \in \operatorname{argmin} f \quad \text{if and only if} \quad 0 \in \partial f(x). \tag{1.2}$$

The proximal mapping $\operatorname{prox}_{\lambda f} : \mathcal{H} \to \mathcal{H}$ is defined as follows:

$$\operatorname{prox}_{\lambda f}(x) := \underset{\xi \in \mathcal{H}}{\operatorname{argmin}} \left\{ f(\xi) + \frac{1}{2\lambda} \|x - \xi\|^2 \right\} \quad \lambda > 0. \tag{1.3}$$

Writing the optimality condition (1.2)(Fermat's rule) for (1.3), we have that

$$\operatorname{prox}_{\lambda f}(x) + \lambda \partial f \left(\operatorname{prox}_{\lambda f}(x) \right) \ni x,$$

i.e.,

$$\operatorname{prox}_{\lambda f}(x) = (I + \lambda \partial f)^{-1}(x). \tag{1.4}$$

Thus, $\operatorname{prox}_{\lambda f}$ is the resolvent of index $\lambda > 0$ of the maximal monotone operator ∂f.

Example 1. The normality operator of a convex and closed set C, denoted by N_C, is defined as follows

$$N_C(x) := \begin{cases} \{w \in \mathcal{H} : \langle w, x - z \rangle \geq 0, \quad \forall z \in C\} & \text{,if} \quad x \in C \\ \emptyset & \text{,if} \quad x \notin C. \end{cases}$$

It can be checked that $N_C(x) = \partial \iota_C(x)$ (see e.g. [20, Example 16.13]), where ι_C is the indicator function of C, which is defined as

$$\iota_C(x) := \begin{cases} 0 & \text{if} \quad x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, N_C is a maximal monotone operator (see [20, Example 20.26]). Observe that from definition of N_C: for $x^* \in C$ and $w^* \in \mathcal{H}$, one has

$$0 \in w^* + N_C(x^*) \quad \text{if and only if} \quad \langle w^*, x - x^* \rangle \geq 0 \quad \text{for all} \quad x \in C. \tag{1.5}$$

Definition 1.2.3 (Cocoercive operator)**.** An operator $T : \mathcal{H} \to \mathcal{H}$ is cocoercive with constant $\beta > 0$ if

$$\langle x - x', T(x) - T(x') \rangle \geq \beta \|T(x) - T(x')\|^2 \quad \forall x, x' \in \mathcal{H}.$$

The above inequality implies that T is $\frac{1}{\beta}$-Lipschitz continuous.

The ε-enlargement of a maximal monotone operator was introduced in [30] by Burachik, Iusem and Svaiter.

Definition 1.2.4. Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator and $\varepsilon \geq 0$. The ε-enlargement of T is the operator $T^\varepsilon : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$T^\varepsilon(z) := \{v \in \mathcal{H} \mid \langle z - z', v - v' \rangle \geq -\varepsilon \;\; \forall (z', v') \in G(T)\} \quad z \in \mathcal{H}. \tag{1.6}$$

It is clear that $T(z) \subset T^\varepsilon(z)$ for all $z \in \mathcal{H}$.

The following summarizes some useful properties of T^ε (see, e.g., [32, Lemma 3.1 and Proposition 3.4(b)]).

Proposition 1.2.5. 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 T is maximal monotone, $\{(\tilde{z}_k, v_k, \varepsilon_k)\}$ is such that $v_k \in T^{\varepsilon_k}(\tilde{z}_k)$, for all $k \geq 1$, $w - \lim_{k\to\infty} \tilde{z}_k = z$, $\lim_{k\to\infty} v_k = v$ and $\lim_{k\to\infty} \varepsilon_k = \varepsilon$, then $v \in T^\varepsilon(z)$[1].

[1] $w - \lim$ denotes weak limit.

We now state the *weak transportation formula* [31] for computing points in the graph of T^{ε}. This formula will be useful in the complexity analysis of some ergodic iterates generated by the algorithms studied in the Chapter 2.

Theorem 1.2.6. ([31, Theorem 2.3]) Suppose $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone and let $\tilde{z}_\ell, v_\ell \in \mathcal{H}$, $\varepsilon_\ell, \alpha_\ell \in \mathbb{R}_+$, for $\ell = 1, \ldots, k$, be such that

$$v_\ell \in T^{\varepsilon_\ell}(\tilde{z}_\ell), \quad \ell = 1, \ldots, k, \quad \sum_{\ell=1}^{k} \alpha_\ell = 1,$$

and define

$$\tilde{z}_k^a := \sum_{\ell=1}^{k} \alpha_\ell \, \tilde{z}_\ell \, , \quad v_k^a := \sum_{\ell=1}^{k} \alpha_\ell \, v_\ell \, , \quad \varepsilon_k^a := \sum_{\ell=1}^{k} \alpha_\ell \left(\varepsilon_\ell + \langle z_\ell - \tilde{z}_k^a, v_\ell - v_k^a \rangle \right).$$

Then, the following hold:

(a) $\varepsilon_k^a \geq 0$ and $v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$.

(b) If, in addition, $T = \partial f$ for some proper, convex and closed function f and $v_\ell \in \partial_{\varepsilon_\ell} f(\tilde{z}_\ell)$ for $\ell = 1, \ldots, k$, then $v_k^a \in \partial_{\varepsilon_k^a} f(\tilde{z}_k^a)$.

1.3 Monotone inclusion problems

Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator. A monotone inclusion problem (MIP) consists in

$$\text{find} \quad z \in \mathcal{H} \quad \text{such that} \quad 0 \in T(z). \tag{1.7}$$

A zero of the monotone operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is any point $z \in \mathcal{H}$ satisfying (1.7). We denote the set of all zeros of T or the solution set of the MIP by $T^{-1}(0)$ or $\text{zer}(T)$.

Due to the mathematical generality of maximal monotone operators, problem (1.7) is very inclusive and serves as an unified model for many problems such as optimization, variational inequalities, saddle-point, equilibrium problems, etc.

Example 2 (Minimization problem)**.** A basic example of monotone inclusion is the convex minimization problem:

$$\min_{z \in \mathcal{H}} f(z). \tag{1.8}$$

Solving (1.8) is equivalent to $\nabla f(z) = 0$ if f is a differentiable function, or $0 \in \partial f(z)$ in the nondifferentiable case (see the Fermat's rule (1.2)). Therefore, the minimizing problem (1.8) is equivalent to MIP (1.7).

Example 3 (Variational inequality problem)**.** Let C be a nonempty closed and convex subset of a real Hilbert space $\mathcal{H}$ and $T : \mathcal{H} \to \mathcal{H}$ a single value operator. The variational inequality problem for T and C, denoted by $\text{VIP}(T, C)$ is as follows: Find $x^* \in C$ such that there exist $w^* \in T(x^*)$ satisfying

$$\langle w^*, x - x^* \rangle \geq 0 \quad \forall x \in C.$$

Observe that, the last relation and (1.5) immediately yields that $0 \in T(x^*) + N_C(x^*)$. Therefore, any VIP can be viewed as a MIP.

Other important instance of the MIP is the saddle-point problem, see e.g., [114] (see also [111] and references therein for other special instances).

1.4 Proximal Point and operator splitting methods

In this subsection, we provide a brief overview of some operator splitting methods. These methods include the proximal point, forward–backward, Tseng's forward–backward and projective primal–dual splitting methods.

1.4.1 Proximal Point algorithm

The proximal point algorithm (PPA) (or proximal point (PP) method) is the iterative scheme, which can be described as follows

$$z_{k+1} = (\lambda_k T + I)^{-1}(z_k), \quad \forall\, k \geq 0 \tag{1.9}$$

where $\{\lambda_k\} \subset \mathbb{R}_{++}$ is a sequence of regularization parameters (or step size parameters) and $(\lambda T + I)^{-1}$ is the resolvent of T, see Section 1.2. When considering numerical schemes of this type, the usual terminology is that we perform a *backward step*, meaning that the set-valued operator is evaluated via its resolvent. The PPA was primarily proposed by Martinet in [84] inspired in the earlier work of Moreau [95] and a few years later popularized and generalized for general maximal monotone operators by Rockafellar in [112].

The resolvent map is a nonexpansive mapping (see [26, Proposition 2.2]) and furthermore

$$J_{\lambda T}(z) = z \quad \text{if and only if} \quad 0 \in T(z).$$

The asymptotic analysis concerning PPA reveals that the sequence generated by (1.9) converges weakly to a solution of (1.7), provided the set of solutions $T^{-1}(0) \neq \emptyset$ and the step size parameter λ_k is bounded away from zero (see [84] in the context of variational inequalities on bounded sets and [112] in the general case).

The proximal point algorithm can be viewed as an implicit one-step discretization method for the evolution differential inclusion problem

$$\dot{z}(t) + T(z(t)) \ni 0 \quad t > 0. \tag{1.10}$$

As examples let us consider two particular cases for T in the above differential inclusion:

- $T = \nabla f$ as the gradient of a differentiable function $f : \mathcal{H} \to \mathbb{R}$ whose gradient is Lipschitz continuous. Then, for each $z_0 \in \mathcal{H}$,

$$\begin{cases} \dot{z}(t) = -\nabla f(z(t)) & t > 0 \\ z(0) = z_0, \end{cases} \tag{1.11}$$

is the so-called *steepest descent* differential equation. The *Cauchy-Lipschitz-Picard Theorem* (see for example [122, Theorem 54]) ensures that (1.11) has a unique solution. Observe also that the stationary points of (1.11) are exactly the critical points of f (zeroes of ∇f). Moreover, it was proved in [103, Section 6.1] that

$$z(t) \rightharpoonup z^* \in S \quad \text{and} \quad \lim_{t\to+\infty} f(z(t)) = \inf f,$$

whenever $S := \operatorname{argmin} f \neq \emptyset$ and f is convex. It is worth to mention that the properties of the trajectories (solution) of (1.11) are expected to continue inheriting the same properties when it is discretized at the time. An implicit finite-difference scheme for (1.11) gives

$$\frac{z_{k+1} - z_k}{\lambda_k} = -\nabla f(z_{k+1}) \Longleftrightarrow z_{k+1} - z_k + \lambda_k \nabla f(z_{k+1}) = 0.$$

Since f is convex, the latter relation is equivalent to the following variational problem:

$$z_{k+1} = \operatorname*{argmin}_{z\in\mathcal{H}} \left\{ f(z) + \frac{1}{2\lambda_k}\|z - z_k\|^2 \right\} = \operatorname{prox}_{\lambda_k f}(z_k), \tag{1.12}$$

which is exactly the proximal point algorithm for convex minimizing problem (see (1.4)). In contrast, considering an explicit discretization of (1.11):

$$\frac{z_{k+1} - z_k}{\lambda_k} = -\nabla f(z_k) \Longleftrightarrow z_{k+1} = z_k - \lambda_k \; \nabla f(z_k). \tag{1.13}$$

We recover the *gradient method* (or *forward step method*) for the unconstrained minimizing problem, originally devised by Cauchy in 1847.

- If $T = \partial f$, with $f \in \Gamma_0(\mathcal{H})$, it is possible to prove that the steepest descent differential inclusion

$$\begin{cases} \dot{z}(t) + \partial f(z(t)) \ni 0, \quad t > 0 \\ z(0) = z_0, \end{cases} \tag{1.14}$$

has similar properties of (1.11): for each $z_0 \in \overline{\operatorname{dom} f}$, the steepest descent differential inclusion (1.14) has a unique absolutely continuous solution $z : [0, +\infty) \to \mathcal{H}$ and satisfies $\lim_{t\to\infty} f(z(t)) = \inf f$. Further, if $S = \operatorname{argmin} f \neq \emptyset$, then $z(t)$ converges weakly to a point in S (see [26, 115]). Similar implicit and explicit finite-difference discretization in (1.11) can be applied to the nonsmooth case in (1.14), resulting (1.12) for implicit discretization and the so-called *subgradient method* [116] for explicit discretization in (1.14).

Despite the difficulty inherent to the implementation of the implicit rule, the proximal point method (1.12) has remarkable stability properties (compared with the explicit method (1.13), see e.g., [103, Sections 6.1 and 6.2] for a more detailed explanation).

1.4.2 Forward-backward and Tseng's forward-backward algorithms

We now assume that we are interested in solving the problem

$$0 \in F(z) + B(z) \tag{1.15}$$

where B and F are maximal monotone operators, and F is assumed to be point-to-point.

An important special case is when $B = N_C$, the normal cone of a nonempty closed convex subset C of $\mathcal{H}$. In this case, (1.15) reduces to a variational inequality problem (see Example 3).

Firstly, we assume that $F : \mathcal{H} \to \mathcal{H}$ is a (single-valued) β-cocoercive operator ($\beta > 0$). The *forward-backward algorithm*, proposed by Lions and Mercier [76] and Passty [100] and whose its roots is in the projected gradient algorithm for convex optimization, is one of the most popular numerical algorithms for solving the structured monotone inclusion problem (1.15), having numerous applications in modern applied mathematics (see, e.g., [20, 107]). It can be described as follows: for all $k \geq 0$,

$$z_{k+1} := (\lambda_k B + I)^{-1}(z_k - \lambda_k F(z_k)), \tag{1.16}$$

where $\lambda_k > 0$ is a stepsize parameter and z_k is the current iterate. The sequence $\{z_k\}$ generated in (1.16) is weakly convergent to a solution of (1.15) whenever the set of solutions to (1.15) is not empty set and $\lambda_k < 2\beta$ (see, e.g., [20]). The terminology forward-backward is justified by the fact that the set-valued operator is evaluated through a backward step and the single-valued one via a forward step. The eqref eq: fb.mth iterative scheme can be understood as an explicit time discretization of step size equal to 1 of the first order dynamic system

$$\begin{cases} \dot{z}(t) + z(t) = J_{\lambda B}(I - \lambda F)z(t), \quad t > 0, \\ z(0) = z_0. \end{cases}$$

For more about dynamical systems of implicit type associated to monotone inclusions and convex optimization problems, see e.g. [1, 9].

Let us suppose now that the cocoercivity of F is relaxed to monotonicity and Lipschitz-continuity. In the seminal paper [128], Tseng proposed and studied the following modification of (1.16) - known as the *Tseng's modified forward-backward method* or as the *forward-backward-forward method*: for all $k \geq 0$,

$$\begin{cases} \tilde{z}_k := (\lambda_k B + I)^{-1}(z_k - \lambda_k F(z_k)), \\ z_{k+1} := \tilde{z}_k - \lambda_k(F(\tilde{z}_k) - F(z_k)). \end{cases} \tag{1.17}$$

It is clear that (1.17) generalizes (1.16) by performing an additional forward step to define the next iterate z_{k+1}. This is crucial to obtain convergence under the (weaker than cocoercivity) assumption of *Lipschitz continuity* on F (see, e.g., [20, 128]). If we assume that the set of solutions to (1.17) is nonempty and $\lambda_k < 1/L$ (with L being the Lipschitz constant of F), the sequences generated by (1.17) converges to a solution of (1.15). Despite the fact that the Tseng's modified forward-backward method requires an additional sequence to be computed, this numerical scheme opened the gate towards the development of the *primal-dual algorithms* that are able to solve highly structured monotone inclusion problems (see [27]). Like the *forward-backward method*, the *forwad-backward-forward method* can be seen as a (implicit) discretization of the following dynamical systems governed by maximally monotone operators:

$$\begin{cases} \tilde{z}(t) = J_{\lambda B}\left(I - \lambda F\right) z(t), \\ \dot{z}(t) + z(t) = \tilde{z}(t) - \lambda\left(F\tilde{z}(t) - Fz(t)\right), \\ z(0) = z_0. \end{cases}$$

Asymptotic analysis of the dynamical system above can be found in [19].

1.4.3 Projective splitting method (PSM)

We consider the monotone inclusion problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* T_i G_i(z) \tag{1.18}$$

where $n \geq 2$, and for each $i = 1, \ldots, n$, the operator $T_i : \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ is (set-valued) maximal monotone and $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ is a bounded linear operator. Moreover it is assumed that (1.18) has at least one solution.

A very popular strategy to find approximate solutions of (1.18) is that of (monotone) operator splitting algorithms, which traces back to the development and analysis of well-known numerical algorithms like the forward-backward, Douglas/Paceman-Rachford and many others [46, 76, 85, 128]. These operator splitting techniques are, in turn a special instance of the *Krasnoselskii-Mann iteration* for finding fix points of nonexpansive operators [75, 82].

A different class of operator splitting algorithms is the family of *projective splitting algorithms*(PSM), which has deserved a lot of attention in nowadays research, mainly due to its flexibility (when compared to other classes of operator splitting algorithms) regarding parameters and the activation of T_i and G_i separately during the iterative process. This class of algorithms has a different convergence mechanism based on projection on separating sets and does not reduces to the *Krasnoselskii-Mann iteration*. The first instances of projective splitting algorithms appeared in two papers by Eckstein and Svaiter, namely [51, 52]: in the first one, the authors addressed the problem of find zeros of the sum of two operators based on the pioneering works [65, 118, 119], and in the second paper, they generalized their first paper to more than two operators (with G_i being the identity operator for $i = 1, \ldots, n$ in (1.18)). We mention some recent developments on the projective splitting algorithms:

- In [2], Alotaibi, Combettes and Shahzad introduced a technique that combines proximal and projective steps for handling compositions of linear and monotone operators for solving systems of composite monotone inclusions, i.e., $n = 2$ with G_1 being equal to the identity operator in (1.18). This approach served to propose new primal-dual splitting algorithms for solving systems of inclusions involving sums of linearly composed maximally monotone operators.

- In [42], Combettes and Eckstein extended the approaches taken in the earlier works [2, 52] incorporating block-iterative and asynchronous features. The block-iterative operation means that not all operators are activated at every iteration but only a subset of them, and asynchronous if, at any iteration, it has the ability to incorporate the result of calculations initiated at earlier iterations, allowing lags in the operator processing (this approach is called "incremental" in the optimization literature).

- Recently, Jonhstone and Eckstein (see [68, 71]) showed that it is possible to process Lipschitz-continuous operators using forward steps rather than the customary resolvent or backward step, i.e., there exist a subset $I_F \subset \{1, \ldots, n\}$ such that for each $i \in I_F$, T_i is Lipschitz-continuous and for each of these operators are processed forward type steps. The same authors, in [70], studied the case where $T_i = A_i + B_i$ in (1.18) with A_i maximal and B_i cocoercive (for $i = 1, \ldots, n$), where in order to construct a semispace containing the "extended solution", B_i and A_i are processed by forward and backward steps, respectively.

In order to motivate the projective splitting algorithms, for simplicity, we consider (3.1) without the linear operators G_i, i.e., find $z \in \mathcal{H}_0$

$$0 \in T_1(z) + \cdots + T_n(z), \tag{1.19}$$

which in turn is clearly equivalent to find a point in the *extended solution set* of (1.19):

$$\mathcal{S} := \left\{ (z, w_1, \ldots, w_{n-1}) \in \mathcal{H}_0 \times \cdots \times \mathcal{H}_{n-1} \mid w_i \in T_i(z),\ i = 1, \ldots, n-1,\ -\sum_{i=1}^{n-1} w_i \in T_n(z) \right\}. \tag{1.20}$$

Since $\mathcal{S}$ is nonempty (by the initial assumption), closed and convex in $\boldsymbol{\mathcal{H}}$ (see [68, Lemma 3] for a more general case), it follows that problem (1.18) reduces to the task of finding a point in $\mathcal{S}$. Hence the family of splitting projective algorithms can be seen as a particular case of the *separator-projector method* for finding points in convex closed sets.

Note now that, if we pick $y_i^k \in T_i(x_i^k)$ $(i = 1, \ldots, n)$, then from the monotoncity of T_i and the inclusions in (1.20) we have

$$\langle z - x_i^k, w_i - y_i^k \rangle \geq 0 \qquad i = 1, \ldots, n,$$

where $w_n := -\sum_{i=1}^{n-1} w_i$. This mean that

$$\sum_{i=1}^{n} \langle z - x_i^k, y_i^k - w_i \rangle \leq 0 \qquad \forall (z, w_1, \ldots, w_{n-1}) \in \mathcal{S}. \tag{1.21}$$

The latter inequality means, in particular, that the set $\{(x_i^k, y_i^k)\}_{i=1}^n$ defines a function of $p = (z, w_1, \ldots, w_{n-1})$ which is negative in $\mathcal{S}$, namely

$$\varphi_k(z, w_1, \ldots, w_{n-1}) := \sum_{i=1}^{n} \langle z - x_i^k, y_i^k - w_i \rangle.$$

It can be proved that this function is actually affine and, as a consequence, we conclude from (1.21) that it defines a semispace containing the extended solution set $\mathcal{S}$, say $H_k := \{p \in \boldsymbol{\mathcal{H}} \mid \varphi_k(p) \leq 0\} \supset \mathcal{S}$, see Figure 1.1.

Based on the exposed above, it follows that the main mechanism behind the idea of projective splitting algorithms is basically: at the iteration $p^k := (z^k, w_1^k, \ldots, w_{n-1}^k)$, pick, for each $i = 1, \ldots, n$, pairs (x_i^k, y_i^k) in the graph of T_i in such a way that $\varphi_k(p^k)$ is positive if $p^k \notin \mathcal{S}$, then update the current iterate to $p^{k+1} := (z^{k+1}, w_1^{k+1}, \ldots, w_{n-1}^{k+1})$ by projecting p^k onto the semispace H_k. In fact, since the extended solution is entirely on the other side of this semispace, then the projection of the current point makes progress towards the solution. Computation of (x_i^k, y_i^k) are in general performed by activating the (approximate) resolvent $(\rho T_i + I)^{-1}$ ($\rho > 0$) operator of each T_i to guarantee, in particular, that the current iterate $(z^k, w_1^k, \ldots, w_{n-1}^k)$ belongs to the positive side of the corresponding hyperplane $\widehat{H}_k := \{p \in \mathcal{H} \mid \varphi_k(p) = 0\}$, i.e. $\varphi_k(p^k) = \sum_{i=1}^n \langle z^k - x_i^k, y_i^k - w_i^k \rangle > 0$ (see Figure 1.1).

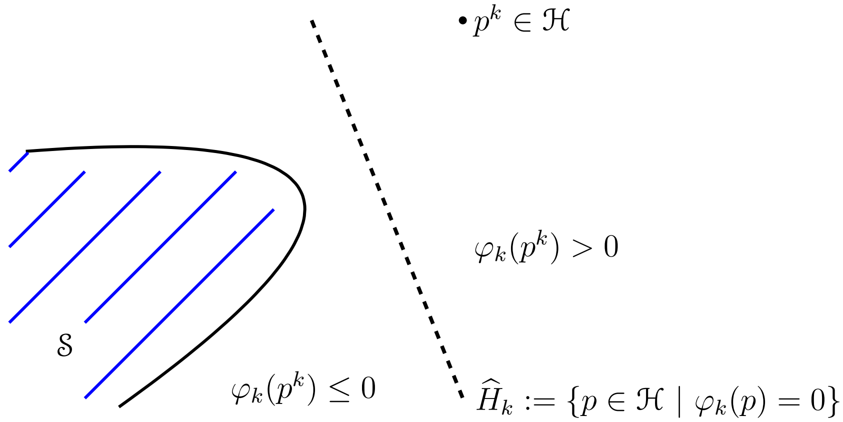

Figure 1.1: Geometrical interpretation of the splitting projective method.

The decomposition properties of the PSM appear from the particular way in which the separating hyperplanes are constructed (affine function φ_k) through individual calculations on each operator T_i, which we will discuss in more details in Chapter 3.

1.5 Inertial methods

In [105], Polyak introduced the so-called *heavy ball method*, in order to speed-up the gradient algorithm. It is a two-step iterative method for minimizing a differentiable (convex) function $f : \mathcal{H} \to \mathbb{R}$, which has the following form:

$$\begin{cases} w_k := z_k + \alpha_k(z_k - z_{k-1}), \\ z_{k+1} := w_k - \lambda_k \nabla f(z_k). \end{cases} \tag{1.22}$$

where $\alpha_k \in [0, 1)$ is an extrapolation factor and $\lambda_k > 0$ is a step-size parameter that has to be sufficiently small. The difference when compared to the gradient method (1.13) is that, in each iteration, the extrapolate term $w_k := z_k + \alpha_k(z_k - z_{k-1})$ is used instead of z_k. The acceleration is explained by the fact that the next iterate is computed by taking a step which is a combination of the direction $z_k - z_{k-1}$ and the current anti-gradient direction $-\nabla f(z_k)$. This modification increases the speed of gradient descent, especially when the objective is strongly convex (see [104]). The heavy ball method can also be interpreted as an explicit finite differences discretization of the so-called *heavy ball with friction* (HBF) second order dynamical system

$$\ddot{z}(t) + \gamma \dot{z}(t) + \nabla f(z(t)) = 0, \tag{1.23}$$

which is a nonlinear oscillator with damping $\gamma > 0$ and potential $f : \mathcal{H} \to \mathbb{R}$. When $\mathcal{H} = \mathbb{R}^2$, this system is a simplified version of a differential equation describing the motion of a heavy ball that keeps rolling over the graph of the function f under its own inertia until friction stops it at a stationary point of f. The three terms in (1.23) can be interpreted, respectively, as inertial, friction and gravity forces. We also mention that the dynamical system (1.23) has been considered by several authors in the context of minimizing the function f, these investigations being either concerned with the asymptotic convergence of the generated trajectories to a critical point of f

as well as the convergence of the function value along the trajectories to its global minimum value (see e.g., [3, 9, 14, 61] and references therein).

In 1983, in the seminal paper [97], Nesterov proposed a modification of the heavy ball method in other to improved the convergence rate for differentiable convex functions with L-Lipschitz gradient as follows: for all $k \geq 0$

$$\begin{cases} w_k := z_k + \alpha_k(z_k - z_{k-1}), \\ z_{k+1} := w_k - \lambda_k \nabla f(w_k). \end{cases} \tag{1.24}$$

where $\lambda_k = 1/L$ and the inertial parameter $\{\alpha_k\}$ behaves like $1 - \frac{3}{k}$ as $k \to \infty$. Currently, the resulting method is known in the literature as *Nesterov's fast gradient method*. The difference to the method proposed by Polyak is that the gradient is evaluated in the extrapolated term w_k instead to z_k. This iterative scheme exhibits (in the worst case) the convergence rate in functional values $f(z_k) - \min f = \mathcal{O}(\frac{1}{k^2})$, which for first order methods is known to be optimal. It is important to mention that when $\gamma(t) = \frac{3}{t}$, Su, Boyd and Candes, in [124], have shown that HBF can be seen as a continuous version of the above Nesterov's fast gradient method. Observe also that $\alpha_k \to 1$, as $k \to \infty$. This is a key property for obtaining fast convergent methods, in line of Nesterov's method (1.24). This motivated the development of iterative schemes with extrapolation steps for general maximal monotone operators for obtaining fast methods (in the line of Nesterov's methods) which was also one of the motivations of this thesis. We mention some contributions on this subject [6, 8, 10, 11, 16, 43, 83].

In [5], Alvarez and Attouch translated the idea of the heavy ball method to the setting of a general maximal monotone operators using the framework of the PP method (1.9). The resulting algorithm is called the inertial proximal point algorithm (IPPA). To motivated it, we consider the implicit discretization of (HBF) as follows:

$$\frac{z_{k+1} - 2z_k + z_{k-1}}{h^2} + \gamma \frac{z_{k+1} - z_k}{h} + \nabla f(z_{k+1}) = 0.$$

Rearranging the above equality one has

$$\left(\frac{1 + \gamma h}{h^2}\right) z_{k+1} - \left(\frac{1 + \gamma h}{h^2}\right) z_k + \frac{1}{h^2}(z_k - z_{k-1}) + \nabla f(z_{k+1}) = 0.$$

Setting $\lambda := \frac{h^2}{1+\gamma h}$ and $\alpha := \frac{1}{1+\gamma h}$, the preceding equality becomes

$$z_{k+1} = z_k + \alpha(z_k - z_{k-1}) - \lambda \nabla f(z_{k+1}).$$

Note that $0 < \alpha < 1$ and $\lambda > 0$. In terms of resolvents, $J_{\lambda \nabla f} = (I + \lambda \nabla f)^{-1}$, the latter relation can be written as

$$z_{k+1} = (\lambda \nabla f + I)^{-1} (z_k + \alpha(z_k - z_{k-1})),$$

or equivalently (whenever f convex, see (1.3)),

$$z_{k+1} = \underset{z \in \mathcal{H}}{\operatorname{argmin}} \left\{ f(z) - \frac{1}{2\lambda} \|z - (z_k - \alpha(z_k + z_{k-1}))\|^2 \right\}. \tag{1.25}$$

Note that (1.25) is nothing but a proximal point step applied to the extrapolated point $w_k = z_k + \alpha(z_k - z_{k-1})$, rather than z_k as in the classical *PP method* (1.12). Convergence analysis of (1.25) in a more general context, $f \in \Gamma_0(\mathcal{H})$ (not necessarily differentiable) was established by Alvarez in [3]. For the general maximal monotone operator T instead of ∇f, (1.25) is written as follows: $\forall k \geq 0$,

$$\begin{cases} w_k := z_k + \alpha_k(z_k - z_{k-1}), \\ z_{k+1} := (\lambda_k T + I)^{-1}(w_k). \end{cases} \tag{1.26}$$

where $\{\alpha_k\}$ is the extrapolation parameters (or inertial parameter) and $\{\lambda_k\}$ is the step size parameter. As we already mentioned, the iterative scheme (1.26) is called *Inertial Proximal Point Algorithm* (IPPA), or *Inertial PP method.* Note that if $\alpha_k \equiv 0$, then it follows that (1.26) reduces to the Rockafellar's PPA (1.9). Inertial PP-type methods deserve a lot of attention in nowadays research due the possibility of extending this methodology to different practical algorithms and, in part, as we mentioned earlier, due to its connections with fast first-order methods in convex programming. Asymptotic (weak) convergence of $\{z_k\}$ generated in (1.26) to a solution of (1.7) was first obtained by Alvarez and Attouch in [5, Theorem 2.1] under the conditions

$$\begin{aligned} &\underline{\lambda} := \inf_{k\geq 0} \lambda_k > 0, \\ &\forall k \geq 0,\ \alpha_k \in [0,1) \quad \text{and} \quad \alpha := \sup_{k\geq 0} \alpha_k < 1, \\ &\sum_{k=0}^{\infty} \alpha_k \|z_k - z_{k-1}\|^2 < +\infty. \end{aligned} \tag{1.27}$$

One way to ensure the third condition in (1.27) in practice is to determine $\{\alpha_k\}$ adaptively, or in particular as in ([5, Proposition 2.1]):

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < 1/3, \qquad \forall k \geq 0. \tag{1.28}$$

The above upper bound 1/3 on $\{\alpha_k\}$ has become standard in the analysis of inertial-like proximal algorithms (see, e.g., [36, 38, 77, 96]). It seems that (1.28) was first improved by Alvarez in [4, Proposition 2.5] in the setting of projective-proximal point-type methods and, more recently, by Attouch-Cabot in [11] and our contributions [8], with relaxation playing a central role.

Chapter 2

On inexact relative-error hybrid proximal extragradient, forward-backward and Tseng's modified forward-backward methods with inertial effects

This chapter is dedicated to the formulation of an inertial under-relaxed version of the relative-error hybrid proximal extragradient (HPE) method. We study its asymptotic convergence as well as nonasymptotic global convergence rates in terms of iteration-complexity. We analyze the new method under more flexible assumptions than the existing ones, both on the extrapolation and relative-error parameters. As special instances of the proposed inertial HPE method, we derive two types of forward-backward methods for solving structured monotone inclusions, namely inertial under-relaxed versions of the Forward-Backward (FB) and Tseng's modified FB methods.

This chapter is organized as follows. In Section 2.1, we present some preliminaries and basic results, and review some algorithms previously stated in Section 1.4 . Section 2.2, presents our inertial under-relaxed HPE method (Algorithm 2), being the main results: Theorems 2.2.8 (asymptotic convergence), and 2.2.11 and 2.2.13 (iteration-complexity). Finally, Section 2.3 is devoted to present and study the inertial versions of the Tseng's modified forward-backward and forward-backward algorithms; the main results are Theorems 2.3.3 and 2.3.7.

The material presented in this chapter is published in [8].

2.1 Preliminaries, basic results and general notation

Notation: Throughout this chapter $\mathcal{H}$ denotes a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

2.1.1 Problem statement

Consider the general monotone inclusion problem (MIP) of finding $z \in \mathcal{H}$ such that

$$0 \in T(z) \tag{2.1}$$

as well as the *structured* MIP

$$0 \in F(z) + B(z) \tag{2.2}$$

where T and B are (set-valued) maximal monotone operators on $\mathcal{H}$ and $F : \text{dom}(F) \subset \mathcal{H} \to \mathcal{H}$ is a (point-to-point) monotone operator which is either *Lipschitz continuous* or *cocoercive* (see Subsections 2.3.1 and 2.3.2 for the precise statement). Problems (2.1) and (2.2) appear in different fields of applied mathematics and optimization including convex optimization, signal processing, PDEs, inverse problems, among others (see, e.g.,[20, 59]). It is worth mentioning that under mild conditions on the operators F and B, problem (2.2) becomes a special instance of (2.1) with $T := F + B$.

2.1.2 The Alvarez–Attouch's inertial proximal point method

As we mentioned earlier, the *proximal point* (PP) *method* is an iterative scheme for seeking approximate solutions of (2.1). In its exact formulation, an iteration of the PP method can be described by

$$z_k := (\lambda_k T + I)^{-1}(z_{k-1}) \qquad \forall k \geq 1, \tag{2.3}$$

where $\lambda_k > 0$ is known as a stepsize parameter (or proximal parameter) and z_{k-1} is the current iterate.

The *inertial* PP *method* is a modification of (2.3) proposed and studied by Alvarez and Attouch in [5] as follows: for all $k \geq 1$,

$$\begin{cases} w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}), \\ z_k := (\lambda_k T + I)^{-1}(w_{k-1}), \end{cases} \tag{2.4}$$

where $\{\alpha_k\}$ is a sequence of extrapolation parameters and w_k is the inertial term.

One of the main goals of this chapter is the analysis of an inertial under-relaxed HPE-type method under the assumption (actually more general than) (1.28) on $\{\alpha_k\}$; see Assumption (**A**).

2.1.3 The hybrid proximal extragradient (HPE) method of Solodov and Svaiter

Since the exact computation of $z_k = (\lambda_k T + I)^{-1}(z_{k-1})$ can be difficult or even impossible in practice, the use of approximate solutions is essential for devising implementable algorithms. This motivated Rockafellar [112] to propose and analyze an inexact version of the PP method (2.3) based on a summable error criterion. More precisely if, at each iteration $k \geq 1$, z_k is computed satisfying

$$\|z_k - (\lambda_k T + I)^{-1}(z_{k-1})\| \leq e_k, \quad \sum_{k=1}^{\infty} e_k < \infty, \tag{2.5}$$

and $\{\lambda_k\}$ is bounded away from zero, then $\{z_k\}$ converges (weakly) to a solution of (2.1). This result has found important applications in the design and analysis of many practical algorithms for solving challenging problems in optimization and related fields. Many modern inexact versions of the PP method (2.3), as opposed to the summable error criterion (2.5), use *relative-error tolerances* for solving the associated subproblems. The first methods of this type were proposed by Solodov and Svaiter in [117, 118] and subsequently studied in [91, 92, 93, 120, 121]. The key idea consists of observing that (2.3) can be decoupled as

$$v_k \in T(z_k), \quad \lambda_k v_k + z_k - z_{k-1} = 0, \tag{2.6}$$

and then relaxing (2.6) within relative-error tolerance criteria. Among these new methods, the hybrid proximal extragradient (HPE) method [117] has been shown to be very effective as a framework for the design and analysis of many concrete algorithms (see, e.g., [22, 34, 50, 62, 64, 78, 88, 89, 90, 93, 117, 120, 121]).

It can be described as follows:

Algorithm 1. HPE method

Input: $z_0 \in \mathcal{H}$ and $\sigma \in [0, 1)$.

1: **for** $k = 1, 2, \ldots,$ **do**

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

$$v_k \in T^{\varepsilon_k}(\tilde{z}_k), \quad \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2. \tag{2.7}$$

3: Define

$$z_k := z_{k-1} - \lambda_k v_k. \tag{2.8}$$

Here, $T^{\varepsilon}(\cdot)$ is the ε-enlargement of T.

Note that if $\sigma = 0$, then (2.7) and (2.8) imply that $z_k = \tilde{z}_k$ and $\varepsilon_k = 0$, which combined with the fact that $T^0 = T$ (see Proposition 1.2.5(d)) implies that the *HPE method* reduces to the exact PP method (2.3).

Recently, Bot and Csetnek proposed an inertial version of the HPE method in [22]. We emphasize that the proposed method (Algorithm 2) in this chapter differs from the inertial HPE type method of Bot and Csetnek since it is based on a different mechanism of iteration. They have proved asymptotic convergence of their method under the assumption: $\alpha(5 + 4\sigma^2) + \sigma^2 < 1$, where $\sigma \in [0, 1[$ is as in (2.7), and $0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < 1$ for all $k \geq 1$ (cf. (1.28)). The condition $\alpha(5 + 4\sigma^2) + \sigma^2 < 1$ enforces $\alpha \approx 0$ whenever $\sigma \approx 1$. This would, in particular, degenerate the desired inertial effect in many important applications of HPE-type methods for which $\sigma = 0.99$ is known (experimentally) to be the best choice among all possible $\sigma \in [0, 1[$ (see, e.g., [50, 54, 88, 89]).

In the next section, we propose an inertial under-relaxed HPE-type method (Algorithm 2) with guarantee of asymptotic convergence and iteration-complexity (both pointwise and ergodic) under the assumption (actually more general than) (1.28) on $\{\alpha_k\}$; see Assumption (**A**). The price to pay is to perform, in addition to inertial, under-relaxed steps.

2.1.4 Forward-backward and Tseng's modified forward-backward methods

We will briefly summarize (in the notation of this chapter) the forward-backward and Tseng's modified forward-backward splitting methods for solving (2.2). The *forward-backward method* (see, e.g., [76, 100]) can be described as follows: for all $k \geq 1$,

$$z_k := (\lambda_k B + I)^{-1}(z_{k-1} - \lambda_k F(z_{k-1})), \tag{2.9}$$

where $\lambda_k > 0$ is a stepsize parameter and z_{k-1} is the current iterate and $F : \mathcal{H} \to \mathcal{H}$ is a cocoercive operator.

The *Tseng's modified forward-backward splitting method* can be described as: for all $k \geq 1$,

$$\begin{cases} \tilde{z}_k := (\lambda_k B + I)^{-1}(z_{k-1} - \lambda_k F(z_{k-1})), \\ z_k := \tilde{z}_k - \lambda_k (F(\tilde{z}_k) - F(z_{k-1})), \end{cases} \tag{2.10}$$

where $F : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuos.

Since both forward-backward and Tseng's modified forward-backward methods are known to be special instances of the HPE method (Algorithm 1) for solving (2.2) (see, e.g., [92, 117, 125]), we have managed to propose and study inertial under-relaxed versions of (2.9) and (2.10) – namely, Algorithms 4 and 5 , respectively – as special instances of the proposed inertial under-relaxed HPE method (Algorithm 2). We discuss some existing inertial/relaxed variants of (2.9) and (2.10) as well as how they are related to Algorithms 4 and 5 (see the remarks following them).

2.2 An inertial under-relaxed hybrid proximal extragradient (HPE) method

Consider the monotone inclusion problem (2.1), i.e., the problem of finding $z \in \mathcal{H}$ such that

$$0 \in T(z) \tag{2.11}$$

where T is a maximal monotone operator on $\mathcal{H}$ for which $T^{-1}(0) \neq \emptyset$.

In this section, we propose and study the asymptotic convergence and nonasymptotic global convergence rates (iteration-complexity) of an inertial under-relaxed hybrid proximal extragradient (HPE) method (Algorithm 2) for solving (2.11).

Regarding the iteration-complexity analysis, we consider the following notion of approximate solution for (2.11): given tolerances $\rho, \epsilon > 0$, find $z, v \in \mathcal{H}$ and $\varepsilon \geq 0$ such that

$$v \in T^{\varepsilon}(z), \quad \|v\| \leq \rho, \quad \varepsilon \leq \epsilon. \tag{2.12}$$

Note that $\rho = \epsilon = 0$ in (2.12) gives $0 \in T(z)$, i.e., in this case $z \in \mathcal{H}$ is a solution of (2.11) (for a more detailed discussion on (2.12), see, e.g., [92]).

We now present one of the first contributions of this chapter: a relaxed and inertial version of the HPE method, which combines the ideas of Alvarez-Attouch [5] and Sodolov-Svaiter [118].

Algorithm 2. An inertial under-relaxed HPE method for solving (2.11)

Input: $z_0 = z_{-1} \in \mathcal{H}$ and $0 \le \alpha, \sigma < 1$ and $0 < \tau \le 1$.

1: **for** $k = 1, 2, \dots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}). \tag{2.13}$$

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

$$v_k \in T^{\varepsilon_k}(\tilde{z}_k), \quad \|\lambda_k v_k + \tilde{z}_k - w_{k-1}\|^2 + 2\lambda_k \varepsilon_k \le \sigma^2 \|\tilde{z}_k - w_{k-1}\|^2. \tag{2.14}$$

4: Define

$$z_k := w_{k-1} - \tau \lambda_k v_k. \tag{2.15}$$

Remark 2.2.1. We now make some remaks regarding Algorithm 2.

(i) Algorithm 2 clearly combines the inertial PP and the HPE methods (2.4) and (2.7)-(2.8), respectively. It reduces to (2.4) when $\sigma = 0$ and $\tau = 1$. Indeed, in this case, using (2.14), (2.15) and Proposition 1.2.5(d), we find $0 \in \lambda_k T(z_k) + z_k - w_{k-1}$ for all $k \ge 1$ (cf. iteration $(\mathcal{A}_0)$–$(\mathcal{A}_2)$ in [5]), i.e., $z_k = (\lambda_k T + I)^{-1}(w_{k-1})$, which is exactly (2.4).

(ii) A similar inertial relaxed relative-error PP algorithm was proposed and analyzed by Alvarez in [4]. We emphasize that in contrast to Algorithm 2, the algorithm proposed by Alvarez is a projective-type algorithm (see, e.g., [118]) and it is based on a different mechanism of iteration.

(iii) Algorithm 2 generalizes the HPE method of Solodov and Svaiter [92] and (a special instance of) the under-relaxed HPE method of Svaiter [126]. Indeed, the HPE method (2.7) is obtained by letting $\alpha = 0$ and $\tau = 1$, in which case $w_{k-1} = z_{k-1}$, while the under-relaxed HPE method (with $t_k \equiv \tau$, in the notation of the latter reference) appears whenever $\alpha = 0$ in Algorithm 2.

(iv) As we mentioned in Subsection 2.1.3, an inertial HPE-type method was recently proposed and studied by Bot and Csetnek in [22]. We refer the reader to Subsection 2.1.3 for a discussion of the contributions of this paper in the light of the latter reference, regarding the HPE-type methods.

(v) We emphasize that, in contrast to the analysis presented in this work (see Theorems 2.2.11 and 2.2.13), in all cases of inertial-type algorithms which were mentioned in remarks (i)-(iv) no iteration-complexity analysis has been obtained.

(vi) Step 3 of Algorithm 2 does not specify how to compute $\lambda_k > 0$ and the triple $(\tilde{z}_k, v_k, \varepsilon_k)$ satisfying (2.14), their computation depending on the instance of the method under consideration. In this regard, Proposition 2.3.6 shows, in particular, how the evaluation of a cocoercive (monotone) point-to-point operator naturally produces such triples.

2.2.1 Convergence analysis

In this subsection we establish our main convergence result of Algorithm 2, see Theorem 2.2.6 and Theorem 2.2.8. The next four results, especially Proposition 2.2.5, will be useful for proving the main results on the convergence and iteration-complexity of Algorithm 2. The following lemma was proved in [126, Lemma 2.1]. Here, we present a short and direct proof for the convenience of the reader.

Lemma 2.2.2. ([126, Lemma 2.1]) Let $\tilde{z}, v, w \in \mathcal{H}$ and $\lambda > 0$, $\varepsilon \geq 0$ and $\sigma \in [0, 1[$ be such that

$$v \in T^{\varepsilon}(\tilde{z}), \quad \|\lambda v + \tilde{z} - w\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|\tilde{z} - w\|^2. \tag{2.16}$$

Let $\tau \in [0, 1]$ and define $z_+ := w - \tau\lambda v$. Then, the following hold:

(a) For any $z \in \mathcal{H}$,

$$\|w - z\|^2 - \|z_+ - z\|^2 \geq (1 - \sigma)^2 \tau \|\tilde{z} - w\|^2 + 2\tau\lambda \left(\varepsilon + \langle \tilde{z} - z, v\rangle\right) + \tau(1 - \tau)\|\lambda v\|^2.$$

(b) For any $z^* \in T^{-1}(0)$,

$$\|w - z^*\|^2 - \|z_+ - z^*\|^2 \geq (1 - \sigma^2)\tau \|\tilde{z} - w\|^2 + \tau(1 - \tau)\|\lambda v\|^2.$$

Proof. (a) Using the inequality in (2.16) and some algebraic manipulations we find, for any $z \in \mathcal{H}$,

$$\begin{aligned} \|w - z\|^2 - \|(w - \lambda v) - z\|^2 &= \|\tilde{z} - w\|^2 - \|\lambda v + \tilde{z} - w\|^2 + 2\lambda\langle \tilde{z} - z, v\rangle \\ &\geq (1 - \sigma^2)\|\tilde{z} - w\|^2 + 2\lambda \left(\varepsilon + \langle \tilde{z} - z, v\rangle\right). \end{aligned} \tag{2.17}$$

The fact that $z_+ = (1 - \tau)w + \tau(w - \lambda v)$ and (A.7) (see Lemma A.2.1) yield

$$\begin{aligned} \|z_+ - z\|^2 &= (1 - \tau)\|w - z\|^2 + \tau\|(w - \lambda v) - z\|^2 - \tau(1 - \tau)\|\lambda v\|^2 \\ &= \|w - z\|^2 - \tau\left(\|w - z\|^2 - \|(w - \lambda v) - z\|^2\right) - \tau(1 - \tau)\|\lambda v\|^2. \end{aligned}$$

Multiplying (2.17) by $\tau \in [0, 1]$ and using the latter identity we obtain the desired inequality in (a).

(b) This is a direct consequence of item (a), (1.6), the inclusion in (2.16) and the fact that $0 \in T(z^*)$. □

Proposition 2.2.3. Let $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{w_k\}$ be generated by *Algorithm 2* and define, for all $k \geq 1$,

$$s_k := \max\left\{\eta\|z_k - w_{k-1}\|^2, (1 - \sigma^2)\tau\|\tilde{z}_k - w_{k-1}\|^2\right\} \tag{2.18}$$

where

$$\eta := \eta(\sigma, \tau) := \frac{2}{(1 + \sigma)\tau} - 1 > 0. \tag{2.19}$$

Then, for any $z^* \in T^{-1}(0)$,

$$\|z_k - z^*\|^2 + s_k \leq \|w_{k-1} - z^*\|^2 \qquad \forall k \geq 1. \tag{2.20}$$

Proof. Using (2.14), (2.15) and Lemma 2.2.2(b) we obtain

$$\|z_k - z^*\|^2 + (1-\sigma^2)\tau\|\tilde{z}_k - w_{k-1}\|^2 + \tau(1-\tau)\|\lambda_k v_k\|^2 \le \|w_{k-1} - z^*\|^2. \tag{2.21}$$

Note now that from (2.15) and (2.14) we have

$$\begin{aligned}\tau^{-1}\|z_k - w_{k-1}\| = \|\lambda_k v_k\| &\le \|\lambda_k v_k + \tilde{z}_k - w_{k-1}\| + \|\tilde{z}_k - w_{k-1}\| \\ &\le (1+\sigma)\|\tilde{z}_k - w_{k-1}\|,\end{aligned}$$

which, in turn, gives

$$(1-\sigma^2)\tau\|\tilde{z}_k - w_{k-1}\|^2 \ge \frac{(1-\sigma)}{\tau(1+\sigma)}\|z_k - w_{k-1}\|^2. \tag{2.22}$$

On the other hand, (2.15) yields

$$\tau(1-\tau)\|\lambda_k v_k\|^2 = \tau^{-1}(1-\tau)\|\tau\lambda_k v_k\|^2 = \tau^{-1}(1-\tau)\|z_k - w_{k-1}\|^2. \tag{2.23}$$

To finish the proof, note that (2.20) is a direct consequence of (2.18), (2.21)–(2.23) and (2.19). □

Lemma 2.2.4. Let $\{z_k\}$, $\{w_k\}$ and $\{\alpha_k\}$ be generated by *Algorithm 2* and let $z \in \mathcal{H}$. Then, for all $k \ge 1$,

$$\|w_{k-1} - z\|^2 = (1+\alpha_{k-1})\|z_{k-1} - z\|^2 - \alpha_{k-1}\|z_{k-2} - z\|^2 + \alpha_{k-1}(1+\alpha_{k-1})\|z_{k-1} - z_{k-2}\|^2.$$

Proof. From (2.13) we have

$$z_{k-1} - z = \frac{1}{1+\alpha_{k-1}}(w_{k-1} - z) + \frac{\alpha_{k-1}}{1+\alpha_{k-1}}(z_{k-2} - z),$$

and $w_{k-1} - z_{k-2} = (1+\alpha_{k-1})(z_{k-1} - z_{k-2})$, which combined with Lemma A.2.1 in Appendix A.1 yields the desired identity. □

Proposition 2.2.5. Let $\{z_k\}$, $\{w_k\}$ and $\{\alpha_k\}$ be generated by *Algorithm 2* and let $\{s_k\}$ be as in (2.18). Let also $z^* \in T^{-1}(0)$ and define

$$(\forall k \ge -1) \quad h_k := \|z_k - z^*\|^2 \quad \text{and} \quad (\forall k \ge 1) \quad \delta_k := \alpha_{k-1}(1+\alpha_{k-1})\|z_{k-1} - z_{k-2}\|^2. \tag{2.24}$$

Then, $h_0 = h_{-1}$ and

$$h_k - h_{k-1} + s_k \le \alpha_{k-1}(h_{k-1} - h_{k-2}) + \delta_k \qquad \forall k \ge 1, \tag{2.25}$$

i.e., the sequences $\{h_k\}$, $\{s_k\}$, $\{\alpha_k\}$ and $\{\delta_k\}$ satisfy the assumptions of Lemma A.1.1 below.

Proof. Using Lemma 2.2.4 with $z = z^*$ and (2.24) we obtain, for all $k \ge 1$,

$$\|w_{k-1} - z^*\|^2 = (1+\alpha_{k-1})h_{k-1} - \alpha_{k-1}h_{k-2} + \delta_k,$$

which combined with Proposition 2.2.3 and the definition of h_k in (2.24) yields (2.25). The identity $h_0 = h_{-1}$ follows from the fact that $z_0 = z_{-1}$ and the first definition in (2.24). □

Next we present the first result on the asymptotic convergence of Algorithm 2.

Theorem 2.2.6 (First result on the weak convergence of *Algorithm* 2)**.** Let $\{z_k\}$, $\{\lambda_k\}$ and $\{\alpha_k\}$ be generated by *Algorithm* 2. If the following holds

$$\sum_{k=0}^{\infty} \alpha_k \|z_k - z_{k-1}\|^2 < +\infty \tag{2.26}$$

and, additionally, $\lambda_k \geq \underline{\lambda} > 0$, for all $k \geq 1$, then the sequence $\{z_k\}$ converges weakly to a solution of the monotone inclusion problem (2.11).

Proof. Using Proposition 2.2.5, (2.26), the fact that $\alpha_k \leq \alpha$ for all $k \geq 1$ and Lemma A.1.1 below, one concludes that

(i) $\lim_{k\to\infty} \|z_k - z^*\|$ exist for every $z^* \in \mathcal{S} := T^{-1}(0)$. In particular, $\{z_k\}$ is bounded;

(ii) $\sum_{k=1}^{\infty} s_k < +\infty$, and therefore $\lim_{k\to\infty} s_k = 0$, where $\{s_k\}$ is as in (2.18).

Using (ii), (2.14)–(2.18) and the assumption $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$, we find

$$\lim_{k\to\infty} \|z_k - w_{k-1}\| = \lim_{k\to\infty} \|\tilde{z}_k - w_{k-1}\| = \lim_{k\to\infty} \|v_k\| = \lim_{k\to\infty} \varepsilon_k = 0. \tag{2.27}$$

Now let $z^\infty \in \mathcal{H}$ be a weak cluster point of $\{z_k\}$ (recall that it is bounded by (i)) and let $\{z_{k_j}\}$ be such that $z_{k_j} \rightharpoonup z^\infty$. Using (2.27) and the inclusion in (2.14) we obtain

$$(\forall j \geq 1)\;\; v_{k_j} \in T^{\varepsilon_{k_j}}(\tilde{z}_{k_j}), \;\; \lim_{j\to\infty} v_{k_j} = 0, \;\; \lim_{j\to\infty} \varepsilon_{k_j} = 0 \;\text{ and }\; w - \lim_{j\to\infty} \tilde{z}_{k_j} = z^\infty, \tag{2.28}$$

which, in turn, combined with Proposition 1.2.5(e) yields $z^\infty \in \mathcal{S} = T^{-1}(0)$. We have proved that all sequential cluster point of $\{z_k\}$ belong to $\mathcal{S} = T^{-1}(0)$. Since the two assumptions of Lemma A.2.2 are verified, it follows that $\{z_k\}$ converges weakly to a point in $\mathcal{S}$, i.e., $\{z_k\}$ converges weakly to a solution of the monotone inclusion problem (2.11). □

Remark. Condition (2.26) appeared for the first time in [5], and since then it has become a standard assumption in the asymptotic convergence analysis of different inertial PP-type algorithms.

Next, we present a sufficient condition on the input parameters (α, σ, τ) in Algorithm 2 to ensure that (2.26) holds (see Theorems 2.2.8, 2.2.11 and 2.2.13).

Assumption (A): $(\alpha, \sigma, \tau) \in [0, 1[\times [0, 1[\times]0, 1]$ and $\{\alpha_k\}$ satisfy the following (for some $\beta > 0$):

$$0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < \beta < 1 \qquad \forall k \geq 1 \tag{2.29}$$

and

$$\tau = \tau(\sigma, \beta') := \frac{2(\beta' - 1)^2}{(1+\sigma)\left[2(\beta'-1)^2 + 3\beta' - 1\right]}, \tag{2.30}$$

where

$$\beta' := \max\left\{\beta, \frac{2(1-\sigma)}{3 - \sigma + \sqrt{9 + 2\sigma - 7\sigma^2}}\right\} \in \left[\frac{2(1-\sigma)}{3 - \sigma + \sqrt{9 + 2\sigma - 7\sigma^2}}, 1\right[. \tag{2.31}$$

Remark 2.2.7. We make some remarks regarding **Assumption (A)**.

(i) Conditions (2.29)–(2.31) will be crucial to prove convergence and iteration-complexity of the algorithms presented and studied in this chapter; see, e.g., Theorems 2.2.8, 2.2.11 and 2.2.13, and Section 2.3.

(ii) Note that by letting $\sigma = 0$, which by Remark 2.2.1(i) means that it reduces to an under-relaxed version of the (exact) Alvarez–Attouch's inertial PP method, we obtain that (2.29)–(2.31) are now simply given by: $0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < \beta < 1$, for all $k \geq 1$, and

$$\tau = \tau(\beta') := \frac{2(\beta'-1)^2}{2(\beta'-1)^2 + 3\beta' - 1}, \qquad \beta' := \max\{\beta, 1/3\} \in [1/3, 1[\,. \tag{2.32}$$

In particular, in this case, we have $\tau = \tau(0, 1/3) = 1$ whenever $\beta = 1/3$ in (2.29), which corresponds to the standard upper bound on $\{\alpha_k\}$ which has been used in different works in the current literature (see Subsection 2.1.2 and Section 1.5 for a discussion). Hence, even in the setting of *exact* inertial PP methods, conditions (2.29)–(2.31) generalize the usual assumption (1.28). See Figure 2.1.

(iii) As we mentioned earlier, an inertial HPE-type method was proposed and studied by Bot and Csetnek in [22], where asymptotic convergence is proved under the assumption $\alpha(5 + 4\sigma^2) + \sigma^2 < 1$ on $\alpha, \sigma \in [0, 1[$. Note that, in this case, $\alpha \approx 0$ whenever $\sigma \approx 1$. This contrasts to the conditions (2.29)–(2.31), which, in particular yield $\tau = \tau(\sigma, 1/3) = 1/(1+\sigma) > 0.5$ (uniformly on σ) when $\beta = 1/3$ in (2.29). This may become especially useful in numerical implementations of Algorithm 2, since $\sigma = 0.99$ has been usually employed in the recent literature on HPE-type methods (see, e.g., [50, 54, 88, 89]). Further, (2.29)–(2.31) allow the upper bound α on $\{\alpha_k\}$ to be chosen arbitrarily close to 1, at the price of performing under-relaxed steps with the explicitly computed $\tau = \tau(\sigma, \beta)$ as in (2.30). See Figure 2.1.

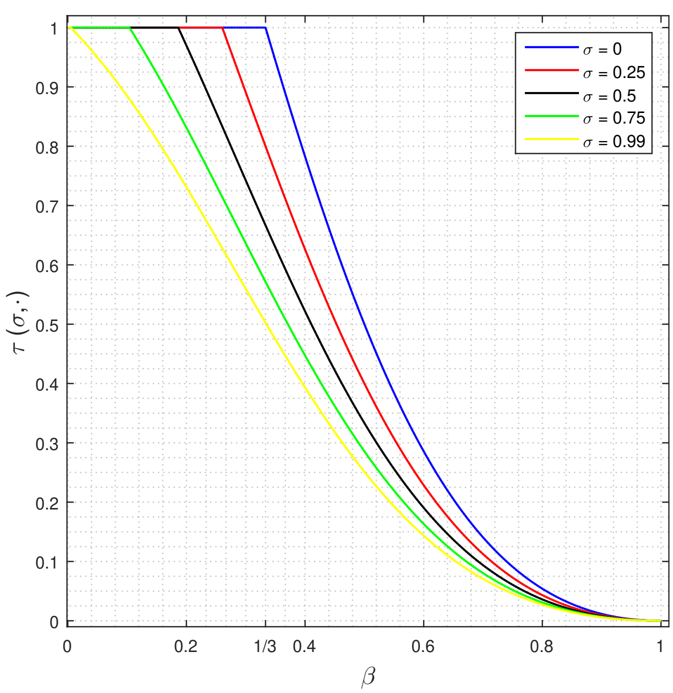

Figure 2.1: Function $]0,1[\ni \beta \mapsto \tau(\sigma, \beta) \in]0,1[$ as in (2.30) for $\sigma \in \{0, 0.25, 0.5, 0.75.0.99\}$.

Theorem 2.2.8 (Second result on the weak convergence of Algorithm 2)**.** Under the *Assumption* (**A**) on *Algorithm 2*, let $\eta > 0$ be as in (2.19) and define the quadratic real function:

$$q(\alpha') := (\eta - 1)\alpha'^{\,2} - (1 + 2\eta)\alpha' + \eta \quad \forall \alpha' \in \mathbb{R}. \tag{2.33}$$

Then, $q(\alpha) > 0$ and, for every $z^* \in T^{-1}(0)$,

$$\sum_{j=1}^{k} \|z_j - z_{j-1}\|^2 \leq \frac{2\,\|z_0 - z^*\|^2}{(1-\alpha)q(\alpha)} \qquad \forall k \geq 1. \tag{2.34}$$

As a consequence, it follows that under the assumption (**A**) the sequence $\{z_k\}$ generated by Algorithm 2 converges weakly to a solution of the monotone inclusion problem (2.11) whenever $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$.

Proof. Using (2.13), the Cauchy-Schwarz inequality and the Young inequality $2ab \leq a^2 + b^2$ with $a := \|z_k - z_{k-1}\|$ and $b := \|z_{k-1} - z_{k-2}\|$ we find

$$\begin{aligned} \|z_k - w_{k-1}\|^2 &= \|z_k - z_{k-1}\|^2 + \alpha_{k-1}^2\|z_{k-1} - z_{k-2}\|^2 - 2\alpha_{k-1}\langle z_k - z_{k-1}, z_{k-1} - z_{k-2}\rangle \\ &\geq \|z_k - z_{k-1}\|^2 + \alpha_{k-1}^2\|z_{k-1} - z_{k-2}\|^2 - \alpha_{k-1}\left(2\|z_k - z_{k-1}\|\|z_{k-1} - z_{k-2}\|\right) \\ &\geq (1 - \alpha_{k-1})\|z_k - z_{k-1}\|^2 - \alpha_{k-1}(1 - \alpha_{k-1})\|z_{k-1} - z_{k-2}\|^2, \end{aligned}$$

which combined with (2.25), and after some algebraic manipulations, yields

$$h_k - h_{k-1} - \alpha_{k-1}(h_{k-1} - h_{k-2}) - \gamma_{k-1}\|z_{k-1} - z_{k-2}\|^2 \leq -\eta(1 - \alpha_{k-1})\|z_k - z_{k-1}\|^2 \quad \forall k \geq 1, \tag{2.35}$$

where

$$\gamma_k := (1 - \eta)\alpha_k^2 + (1 + \eta)\alpha_k \geq 0 \qquad \forall k \geq 0. \tag{2.36}$$

Define,

$$\mu_0 := (1 - \alpha_0)h_0 \geq 0, \qquad \mu_k := h_k - \alpha_{k-1}h_{k-1} + \gamma_k\|z_k - z_{k-1}\|^2 \quad \forall k \geq 1, \tag{2.37}$$

where h_k is as in (2.24). Using (2.33), the assumption that $\{\alpha_k\}$ is nondecreasing (see (2.29)) and (2.35)–(2.37) we obtain, for all $k \geq 1$,

$$\begin{aligned} \mu_k - \mu_{k-1} &\leq \left[h_k - h_{k-1} - \alpha_{k-1}(h_{k-1} - h_{k-2}) - \gamma_{k-1}\|z_{k-1} - z_{k-2}\|^2\right] + \gamma_k\|z_k - z_{k-1}\|^2 \\ &\leq \left[\gamma_k - \eta(1 - \alpha_k)\right]\|z_k - z_{k-1}\|^2 \\ &= -\left[(\eta - 1)\alpha_k^2 - (1 + 2\eta)\alpha_k + \eta\right]\|z_k - z_{k-1}\|^2 \\ &= -q(\alpha_k)\|z_k - z_{k-1}\|^2. \end{aligned} \tag{2.38}$$

Note now that from (2.30) and Lemma A.1.2 we have

$$\beta' = \frac{4 - 2(1 + \sigma)\tau}{4 - (1 + \sigma)\tau + \sqrt{(1 + \sigma)\tau\,[16 - 7(1 + \sigma)\tau]}},$$

which, in turn, combined with the definition of $\eta > 0$ in (2.19), and after some algebraic calculations, gives

$$\beta' = \frac{2\eta}{2\eta + 1 + \sqrt{8\eta + 1}}.$$

The latter identity implies, in particular, that β' is either the smallest or the largest root of the quadratic function $q(\cdot)$. Hence, from (2.29) and the fact that $\beta' \geq \beta$ (see (2.31)) we obtain

$$q(\alpha_k) \geq q(\alpha) > q(\beta') = 0.$$

The above inequalities combined with (2.38) yield

$$\|z_k - z_{k-1}\|^2 \leq \frac{1}{q(\alpha)}(\mu_{k-1} - \mu_k), \quad \forall k \geq 1, \tag{2.39}$$

which, in turn, combined with (2.29) and the definition of μ_k in (2.37), gives

$$\begin{aligned} \sum_{j=1}^{k} \|z_j - z_{j-1}\|^2 &\leq \frac{1}{q(\alpha)}(\mu_0 - \mu_k), \\ &\leq \frac{1}{q(\alpha)}(\mu_0 + \alpha h_{k-1}) \quad \forall k \geq 1. \end{aligned} \tag{2.40}$$

Note now that (2.39), (2.29) and (2.37) also yield

$$\begin{aligned} \mu_0 \geq \ldots \geq \mu_k =& h_k - \alpha_{k-1} h_{k-1} + \gamma_k \|z_k - z_{k-1}\|^2 \\ \geq& h_k - \alpha h_{k-1}, \quad \forall k \geq 1, \end{aligned}$$

and so,

$$h_k \leq \alpha^k h_0 + \frac{\mu_0}{1-\alpha} \leq h_0 + \frac{\mu_0}{1-\alpha} \qquad \forall k \geq 0. \tag{2.41}$$

Hence, (2.34) follows directly from (2.40), (2.41), the definition of μ_0 in (2.37) and the definition of h_0 in (2.24). On the other hand, the second statement of the theorem follows from (2.34) and Theorem 2.2.6 (recall that $\alpha_k \leq \alpha < 1$ for all $k \geq 0$). □

Remark 2.2.9.

i) A quadratic function similar to $q(\cdot)$, as defined in (2.33), was also considered by Alvarez in [4]. As we mentioned in Remark 2.2.1(ii), the algorithm studied in the later reference is different of the corresponding algorithm presented in this work, namely Algorithm 2. Moreover, note that if $\eta = 1$, then $q(\alpha') = 1 - 3\alpha'$ (cf. [5]).

ii) The assumption $z_{-1} = z_0$ in the input of Algorithm 2 was imposed to ensure the inequality $\mu_0 \geq 0$, which is used in the proof of Theorem 2.2.8. An alternative to the latter assumption on z_0 is to choose $\alpha_0 = 0$ (see Step 2 of Algorithm 2), in which case the assumption $z_{-1} = z_0$ is not necessary.

Corollary 2.2.10. Under the *Assumption* (**A**) on *Algorithm* 2, let $\eta > 0$ and $q(\cdot)$ be as in (2.19) and (2.33), respectively, and let $z^* \in T^{-1}(0)$. Then, for all $k \geq 1$,

$$\|z_k - z^*\|^2 + \sum_{j=1}^{k} \tau\Big(\max\left\{\eta\tau\|\lambda_j v_j\|^2, (1-\sigma^2)\|\tilde{z}_j - w_{j-1}\|^2\right\}\Big) \leq \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\|z_0 - z^*\|^2.$$

Proof. Using Proposition 2.2.5 and Lemma A.1.1(a) we conclude that (A.2) in Appendix A.1 holds, with $\{s_k\}$, $\{h_k\}$ and $\{\delta_k\}$ as in (2.18) and (2.24), which gives that the desired result follows from (A.2) and (2.34). □

2.2.2 Complexity analysis

We now study the pointwise and ergodic iteration-complexity of Algorithm 2. In order to study the *ergodic* iteration-complexity of Algorithm 2, we need to define 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 $\{\varepsilon_k\}$, respectively, as follows: For $k \geq 1$,

$$\begin{aligned} &\Lambda_k := \sum_{j=1}^{k} \lambda_j, \qquad \tilde{z}_k^{\,a} := \frac{1}{\Lambda_k} \sum_{j=1}^{k} \lambda_j\, \tilde{z}_j \qquad v_k^{\,a} := \frac{1}{\Lambda_k} \sum_{j=1}^{k} \lambda_j\, v_j, \\ &\varepsilon_k^{\,a} := \frac{1}{\Lambda_k} \sum_{j=1}^{k} \lambda_j(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^{\,a}, v_j - v_k^{\,a}\rangle) = \frac{1}{\Lambda_k} \sum_{j=1}^{k} \lambda_j(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^{\,a}, v_j\rangle). \end{aligned} \tag{2.42}$$

In what follows we present the first result on nonasymptotic global convergence rates/iteration-complexity of Algorithm 2.

Theorem 2.2.11 (global $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate of *Algorithm* 2)**.** Under the *Assumption* (**A**) on Algorithm 2, let $\eta > 0$ and $q(\cdot)$ be as in (2.19) and (2.33), respectively, and let d_0 denote the distance of z_0 to $T^{-1}(0)$. Assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Then, for every $k \geq 1$, there exists $i \in \{1, \dots, k\}$ such that

$$v_i \in T^{\varepsilon_i}(\tilde{z}_i), \tag{2.43}$$

$$\|v_i\| \leq \frac{d_0}{\underline{\lambda}\tau\sqrt{k}} \sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}, \tag{2.44}$$

$$\varepsilon_i \leq \frac{\sigma d_0^2}{2(1-\sigma^2)\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right). \tag{2.45}$$

Proof. Let $z^* \in T^{-1}(0)$ be such that $d_0 = \|z_0 - z^*\|$. It follows from Corollary 2.2.10 that, for every $k \geq 1$, there exists $i \in \{1, \dots, k\}$ such that

$$\tau\, k\Big(\max\left\{\eta\tau\|\lambda_i v_i\|^2, (1-\sigma^2)\|\tilde{z}_i - w_{i-1}\|^2\right\}\Big) \leq \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right) d_0^2,$$

which combined with the assumption $\lambda_i \geq \underline{\lambda} > 0$ and (2.14), and after some simple algebraic manipulations, yields the desired result. □

Remark 2.2.12. We now make some comments regarding the pontwise convergence rate of Algorithm 2.

(i) Theorem 2.2.11 provides a global $\mathcal{O}(1/\sqrt{k})$ *pontwise* convergence rate and ensures, in particular, that for given tolerances $\rho, \epsilon > 0$, Algorithm 2 finds a triple (z, v, ε) satisfying (2.12) after performing at most

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

iterations.

(ii) If $\alpha = 0$ and $\tau = 1$, in which case Algorithm 2 reduces to the HPE method of Solodov and Svaiter, then it follows that Theorem 2.2.11 reduces to [92, Theorem 4.4(a)].

(iii) Analogous global $\mathcal{O}(1/\sqrt{k})$ *pontwise* convergence rates were also obtained in [36, 38] for inertial-type algorithms for variational inequality and convex optimization problems.

Next we study the *ergodic* iteration-complexity of Algorithm 2 under the assumption that $\alpha_k \equiv \alpha$ in (2.13).

Theorem 2.2.13 (global $\mathcal{O}(1/k)$ ergodic convergence rate of *Algorithm* 2)**.** Under the *Assumption* (**A**) on *Algorithm 2* and, additionally, the assumption that $\alpha_k \equiv \alpha$, let $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ be as in (2.42) and let d_0 denote the distance of z_0 to $T^{-1}(0)$. Let also $\eta > 0$ and $q(\cdot)$ be as in (2.19) and (2.33), respectively, and assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$.

Then, for all $k \geq 1$,

$$v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \tag{2.46}$$

$$\|v_k^a\| \leq \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k}\sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \tag{2.47}$$

$$\varepsilon_k^a \leq \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1 + \frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta\tau^2}}\right). \tag{2.48}$$

Proof. Let $z^* \in T^{-1}(0)$ be such that $d_0 = \|z_0 - z^*\|$. Using Algorithm 2's definition and Lemma 2.2.2(a) with $z = \tilde{z}_k^a$ we find, for all $j \geq 1$,

$$\|w_{j-1} - \tilde{z}_k^a\|^2 - \|z_j - \tilde{z}_k^a\|^2 \geq 2\tau\lambda_j\left(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^a, v_j\rangle\right). \tag{2.49}$$

On the other hand, Lemma 2.2.4 yields

$$\|w_{j-1} - \tilde{z}_k^a\|^2 = (1+\alpha_{j-1})\|z_{j-1} - \tilde{z}_k^a\|^2 - \alpha_{j-1}\|z_{j-2} - \tilde{z}_k^a\|^2 + \alpha_{j-1}(1+\alpha_{j-1})\|z_{j-1} - z_{j-2}\|^2,$$

which, in turn, combined with (2.49) gives, for all $j \geq 1$,

$$\|z_j - \tilde{z}_k^a\|^2 - \|z_{j-1} - \tilde{z}_k^a\|^2 + 2\tau\lambda_j\left(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^a, v_j\rangle\right) \leq \alpha_{j-1}\left(\|z_{j-1} - \tilde{z}_k^a\|^2 - \|z_{j-2} - \tilde{z}_k^a\|^2\right) + \delta_j,$$

where the sequence $\{\delta_j\}$ is as in (2.24). Summing the latter inequality over all $j = 1, \ldots, k$ and using (2.42) as well as the assumption $\alpha_k \equiv \alpha$, we obtain

$$\|z_k - \tilde{z}_k^a\|^2 - \|z_0 - \tilde{z}_k^a\|^2 + 2\tau\Lambda_k\varepsilon_k^a \leq \alpha\left(\|z_{k-1} - \tilde{z}_k^a\|^2 - \|z_{-1} - \tilde{z}_k^a\|^2\right) + \sum_{j=1}^{k}\delta_j,$$

which combined with the definition of $\{\delta_j\}$ and (2.34) yields (recall that $z_0 = z_{-1}$)

$$\begin{aligned} 2\tau\Lambda_k\varepsilon_k^a - \frac{2\alpha(1+\alpha)d_0^2}{(1-\alpha)q(\alpha)} &\leq (1-\alpha)\left(\|z_0 - \tilde{z}_k^a\|^2 - \|z_k - \tilde{z}_k^a\|^2\right) \\ &\quad + \alpha\left(\|z_{k-1} - \tilde{z}_k^a\|^2 - \|z_k - \tilde{z}_k^a\|^2\right) \end{aligned}$$

$$\leq 2\max\left\{\|z_0-\tilde{z}_k^a\|\|z_0-z_k\|, \|z_{k-1}-\tilde{z}_k^a\|\|z_{k-1}-z_k\|\right\}, \tag{2.50}$$

where we have also used the inequality $\|a\|^2-\|b\|^2\leq 2\|a\|\|a-b\|$ for all $a,b\in\mathcal{H}$. Now, define

$$(\forall j\geq 1)\quad \hat{z}_j := w_{j-1}-\lambda_j v_j \quad\text{and}\quad \hat{z}_k^a := \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\hat{z}_j. \tag{2.51}$$

From Corollary 2.2.10, the first definition in (2.51), (2.42), (2.15) and the convexity of $\|\cdot\|^2$ we find

$$\|z_\ell - z_j\| \leq \|z_\ell - z^*\| + \|z_j - z^*\| \leq 2d_0\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}} \qquad \forall \ell, j\geq 0, \tag{2.52}$$

$$(1-\tau)^{-2}\sum_{j=1}^{k}\|z_j-\hat{z}_j\|^2 = \sum_{j=1}^{k}\|\lambda_j v_j\|^2 \leq \frac{d_0^2}{\eta\tau^2}\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right) \tag{2.53}$$

and

$$\begin{aligned}\|\tilde{z}_k^a-\hat{z}_k^a\|^2 \leq \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\|\tilde{z}_j-\hat{z}_j\|^2 &\leq \sum_{j=1}^{k}\|\lambda_j v_j+\tilde{z}_j-w_{j-1}\|^2\\ &\leq \sigma^2\sum_{j=1}^{k}\|\tilde{z}_j-w_{j-1}\|^2\\ &\leq \frac{\sigma^2 d_0^2}{(1-\sigma^2)\tau}\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right).\end{aligned} \tag{2.54}$$

From (2.52), (2.53), the convexity of $\|\cdot\|^2$ and the inequality $\|a-b\|^2\leq 2\left(\|a\|^2+\|b\|^2\right)$ (for all $a,b\in\mathcal{H}$), we find

$$\begin{aligned}\|z_\ell-\hat{z}_k^a\|^2 &\leq \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\|z_\ell-\hat{z}_j\|^2\\ &\leq 2\left(\frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\|z_\ell-z_j\|^2+\sum_{j=1}^{k}\|z_j-\hat{z}_j\|^2\right)\\ &\leq 2d_0^2\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right)\left(4+\frac{(1-\tau)^2}{\eta\tau^2}\right) \qquad \forall \ell\geq 0.\end{aligned}$$

Using the above inequality and (2.54) we obtain, for all $\ell\geq 0$,

$$\begin{aligned}\|z_\ell-\tilde{z}_k^a\| &\leq \|z_\ell-\hat{z}_k^a\|+\|\tilde{z}_k^a-\hat{z}_k^a\|\\ &\leq \sqrt{2}d_0\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}}\left(\frac{\sigma}{\sqrt{(1-\sigma^2)\tau}}+\sqrt{4+\frac{(1-\tau)^2}{\eta\tau^2}}\right).\end{aligned} \tag{2.55}$$

Hence, (2.50), (2.52) with $\ell = 0, k-1$ and $j = k$, and (2.55) with $\ell = 0, k-1$ yield

$$
\begin{aligned}
2\tau\Lambda_k\varepsilon_k^a &\leq 4\sqrt{2}d_0^2\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(\frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta\tau^2}}\right) \\
&\quad + \frac{2\alpha(1+\alpha)d_0^2}{(1-\alpha)q(\alpha)} \\
&\leq 4\sqrt{2}d_0^2\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1 + \frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta\tau^2}}\right),
\end{aligned}
$$

which, combined with the assumption $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$, clearly finishes the proof of (2.48).

Now note that using (2.15), (2.13) and the assumption $\alpha_k \equiv \alpha$ we find

$$
\tau\lambda_j v_j = z_{j-1} - z_j + \alpha(z_{j-1} - z_{j-2}) \qquad \forall j \geq 1.
$$

Summing the above identity over $j = 1, \ldots, k$ and using (2.42) and (2.52) with $\ell = 0$ and $j = k-1, k$ we find (recall that $z_0 = z_{-1}$)

$$
\begin{aligned}
\tau\Lambda_k\|v_k^a\| &\leq \|z_0 - z_k\| + \alpha\|z_0 - z_{k-1}\| \\
&\leq 2(1+\alpha)d_0\sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}},
\end{aligned}
$$

which, combined with the assumption $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$, yields (2.47). To finish the proof of the theorem, note that (2.46) is a direct consequence of the inclusion in (2.14) and Theorem 1.2.6(a). □

2.2.3 On the under-relaxed inertial proximal point method

In this subsection we analyze the convergence and iteration-complexity of the under-relaxed inertial proximal point (PP) method (see, e.g., [4, 11]) with constant under-relaxation (Algorithm 3) for solving (2.11). The analysis is performed by viewing Algorithm 3 within the framework of Algorithm 2, for which asymptotic convergence and iteration-complexity were obtained in Theorems 2.2.8, 2.2.11 and 2.2.13.

Algorithm 3. Under-relaxed inertial proximal point method for solving (2.11)

Input: $z_0 = z_{-1} \in \mathcal{H}$ and $0 \le \alpha < 1$ and $0 < \tau \le 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}). \tag{2.56}$$

3: Compute

$$\tilde{z}_k = (\lambda_k T + I)^{-1}(w_{k-1}). \tag{2.57}$$

4: Define

$$z_k := \tau \tilde{z}_k + (1-\tau)(w_{k-1}). \tag{2.58}$$

Next proposition shows that Algorithm 3 is a special instance of Algorithm 2.

Proposition 2.2.14. Algorithm 3 is a special instance of Algorithm 2 with $\sigma = 0$ in the Input, in which case $\varepsilon_k = 0$ and $v_k = (w_{k-1} - \tilde{z}_k)/\lambda_k \in T(\tilde{z}_k)$ for all $k \ge 1$.

Proof. The proof follows from the well-known fact that $\tilde{z} = (\lambda T + I)^{-1} w$ if and only if $v := (w - \tilde{z})/\lambda \in T(\tilde{z})$ and Algorithms 3 and 2's definitions. □

Theorem 2.2.15 (convergence and iteration-complexity of Algorithm 3)**.** Under the *Assumption* **(A)** with $\sigma = 0$ on Algorithm 3, let $\{z_k\}$, $\{v_k\}$, $\{\tilde{z}_k\}$ and $\{\lambda_k\}$ be generated by Algorithm 3 and let the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ be as in (2.42). Let also $q(\cdot)$ be as in (2.33) and let d_0 denote the distance of z_0 to $T^{-1}(0)$. Assume that $\lambda_k \ge \underline{\lambda} > 0$ for all $k \ge 1$. Then, the following statements hold:

(a) The sequence $\{z_k\}$ converges weakly to a solution of the monotone inclusion problem (2.11).

(b) For all $k \ge 1$, there exists $i \in \{1, \ldots, k\}$ such that

$$v_i \in T(\tilde{z}_i), \qquad \|v_i\| \le \frac{d_0}{\underline{\lambda}\tau\sqrt{k}}\sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}. \tag{2.59}$$

(c) If, additionally, $\alpha_k \equiv \alpha$, then, for all $k \ge 1$,

$$v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \tag{2.60}$$

$$\|v_k^a\| \le \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k}\sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \tag{2.61}$$

$$\varepsilon_k^a \le \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1 + \sqrt{4 + \frac{(1-\tau)^2}{\tau(2-\tau)}}\right). \tag{2.62}$$

Proof. The results in (a), (b) and (c) follow directly from Proposition 2.2.14 and Theorems 2.2.8, 2.2.11 and 2.2.13. □

2.3 Inertial under-relaxed forward-backward and Tseng's modified forward-backward methods

Consider the structured monotone inclusion problem (2.2), i.e., the problem of finding $z \in \mathcal{H}$ such that

$$0 \in F(z) + B(z) =: T(z) \tag{2.63}$$

where $F : \text{dom}(F) \subset \mathcal{H} \to \mathcal{H}$ is point-to-point monotone and $B : \mathcal{H} \rightrightarrows \mathcal{H}$ is a (set-valued) maximal monotone operator for which $T^{-1}(0) \neq \emptyset$ (precise assumption on F and B will be stated later).

In this section, we study the convergence and iteration-complexity of inertial (under-relaxed) versions of the forward-backward and Tseng's modified forward-backward methods (2.9) and (2.10), respectively, for solving (2.63), by viewing them within the framework of Algorithm 2, for which asymptotic convergence and iteration-complexity were studied in Section 2.2.

2.3.1 An inertial under-relaxed Tseng's modified forward-backward method

In this subsection, we consider the monotone inclusion problem (2.63) where the following assumptions are assumed to hold:

(C1) $F : \text{dom}(F) \subset \mathcal{H} \to \mathcal{H}$ is monotone and L-Lipschitz continuous on a (nonempty) closed convex set Ω such that $\text{dom}(B) \subset \Omega \subset \text{dom}(F)$, i.e., F is monotone on Ω and there exists $L \geq 0$ such that

$$\|F(z) - F(z')\| \leq L\|z - z'\| \qquad \forall z, z' \in \Omega.$$

(C2) B is a (set-valued) maximal monotone operators on $\mathcal{H}$.

(C3) The solution set of (2.63) is nonempty.

It was proved in [91, Proposition A.1] that under assumptions (C1)–(C3) the operator $T(\cdot)$ defined in (2.63) is maximal monotone, which guarantee that (2.63) is a special instance of (2.11). In particular, it follows that Algorithm 2 can be used to solving the structured monotone inclusion (2.63).

As we mentioned above, in this subsection, we shall study the convergence and iteration-complexity of the following inertial under-relaxed version of the Tseng's modified forward-backward method for solving (2.63).

Algorithm 4. An inertial under-relaxed Tseng's modified forward-backward method for solving (2.63)

Input: $z_0 = z_{-1} \in \mathcal{H}$, $0 \le \alpha < 1$, $0 < \sigma < 1$ and $0 < \tau \le 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}).$$

3: Choose $\lambda_k \in]0, \sigma/L]$, let $w'_{k-1} = P_\Omega(w_{k-1})$ and compute

$$\tilde{z}_k = (\lambda_k B + I)^{-1}(w_{k-1} - \lambda_k F(w'_{k-1})),$$
$$\hat{z}_k = \tilde{z}_k - \lambda_k \left(F(\tilde{z}_k) - F(w'_{k-1})\right).$$

4: Define

$$z_k := (1 - \tau)w_{k-1} + \tau \hat{z}_k.$$

Remark 2.3.1.

(i) Algorithm 4 reduces to the Tseng's modified forward-backward method [128] for solving (2.63) if $\alpha = 0$ and $\tau = 1$, in which case $w_{k-1} = z_{k-1}$ and $z_k = \hat{z}_k$.

(ii) An inertial Tseng's modified forward-backward-type method (based on a different mechanism of iteration) was proposed and studied in [22]. The proposed Tseng's modified forward-backward type method in the latter reference tends to suffer from similar limitations as the inertial HPE-type method proposed in [22], as we discussed in Remark 2.2.7. Moreover, in contrast to the contributions of this chapter which performs the iteration-complexity analysis of Algorithm 4 (see Theorem 2.3.3), [22] focused only on the asymptotic convergence.

Since the proof of the next proposition follows the same outline of [92, Proposition 6.1], we omit it here.

Proposition 2.3.2. Let $\{w_k\}$, $\{w'_k\}$, $\{z_k\}$, $\{\alpha_k\}$, $\{\tilde{z}_k\}$ and $\{\lambda_k\}$ be generated by *Algorithm* 4 and define

$$\varepsilon_k := 0 \;\text{ and }\; v_k := F(\tilde{z}_k) - F(w'_{k-1}) + \frac{1}{\lambda_k}(w_{k-1} - \tilde{z}_k) \qquad \forall k \ge 1. \tag{2.64}$$

Then, the sequences $\{w_k\}$, $\{z_k\}$, $\{\alpha_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$, $\{\varepsilon_k\}$ and $\{\lambda_k\}$ satisfy the conditions (2.13)–(2.15) in *Algorithm* 2. As a consequence, it follows that *Algorithm* 4 is a special instance of *Algorithm* 2 for solving (2.63).

Next we present the convergence and iteration-complexity of Algorithm 4 under the Assumption (**A**) on the Input $(\alpha, \sigma, \tau) \in [0, 1[\times]0, 1[\times]0, 1]$ and on the sequence $\{\alpha_k\}$. We also mention that the observations regarding the parameter τ in Remark 2.2.7(iii) obviously apply to Algorithm 4.

Theorem 2.3.3 (convergence and iteration-complexity of Algorithm 4)**.** Under the *Assumption* (**A**) on $(\alpha, \sigma, \tau) \in [0,1[\times]0,1[\times]0,1]$ and $\{\alpha_k\}$, let $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{\lambda_k\}$ be generated by *Algorithm* 4, let $\{v_k\}$ and $\{\varepsilon_k\}$ be as in (2.64) and let the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ be as in (2.42). Let also $\eta > 0$ and $q(\cdot)$ be as in (2.19) and (2.33), respectively, let d_0 denote the distance of z_0 to $(F+B)^{-1}(0)$ and assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Then, the following statements hold:

(a) The sequence $\{z_k\}$ converges weakly to a solution of the monotone inclusion problem (2.63).

(b) For all $k \geq 1$, there exists $i \in \{1, \ldots, k\}$ such that

$$v_i \in (F+B)(\tilde{z}_i), \qquad \|v_i\| \leq \frac{d_0}{\underline{\lambda}\tau\sqrt{k}}\sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}. \tag{2.65}$$

(c) If, additionally, $\alpha_k \equiv \alpha$, then, for all $k \geq 1$,

$$\begin{aligned} & v_k^a \in (F+B)^{\varepsilon_k^a}(\tilde{z}_k^a), \\ & \|v_k^a\| \leq \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k}\sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \\ & \varepsilon_k^a \leq \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1 + \frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta\tau^2}}\right). \end{aligned} \tag{2.66}$$

Proof. The proof follows directly from Proposition 2.3.2 and Theorems 2.2.8, 2.2.11 and 2.2.13. □

Remark 2.3.4.

(i) Items (b) and (c) ensure, respectively, global pointwise $\mathcal{O}(1/\sqrt{k})$ and ergodic $\mathcal{O}(1/k)$ convergence rates for Algorithm 4. On the other hand, note that the inclusion in (2.66) is potentially weaker than the corresponding one in (2.65).

(ii) If $\lambda_k \equiv \sigma/L$ in Step 3 of Algorithm 4, in which case $\underline{\lambda} = \sigma/L$, then $d_0/\underline{\lambda}$ in (2.65) and (2.66) can be replaced by $d_0 L/\sigma$. In this case, Item (b) gives that for a given tolerance $\rho > 0$, Algorithm 4 finds a pair (z, v) such that (cf. (2.12))

$$v \in (F+B)(z), \quad \|v\| \leq \rho$$

in at most

$$\mathcal{O}\left(\left\lceil \frac{d_0^2 L^2}{\rho^2} \right\rceil\right)$$

iterations, an analogous remark also holding for Item (c).

2.3.2 On the inertial under-relaxed forward-backward method

Similarly to Subsection 2.3.1, in this subsection, we consider the monotone inclusion problem (2.63) but now we assume the following: (C2) and (C3) as in Subsection 2.3.1 and instead of (C1):

(C1′) $F : \mathcal{H} \to \mathcal{H}$ is $(1/L)-$cocoercive, i.e., there exists $L > 0$ such that

$$\langle z - z', F(z) - F(z') \rangle \geq \frac{1}{L} \|F(z) - F(z')\|^2 \qquad \forall z, z' \in \mathcal{H}. \tag{2.67}$$

We observe that it follows from (2.67) that F is, in particular, L–Lipschitz continuous.

Algorithm 5. Inertial under-relaxed forward-backward method for solving (2.63)

Input: $z_0 = z_{-1} \in \mathcal{H}$, $0 \leq \alpha < 1$, $0 < \sigma < 1$ and $0 < \tau \leq 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}).$$

3: Choose $\lambda_k \in]0, 2\sigma^2/L]$ and compute

$$\tilde{z}_k = (\lambda_k B + I)^{-1}(w_{k-1} - \lambda_k F(w_{k-1})).$$

4: Define

$$z_k := (1 - \tau)w_{k-1} + \tau \tilde{z}_k.$$

Remark 2.3.5.

(i) If $\alpha = 0$ and $\tau = 1$, then it follows that Algorithm 5 reduces to the forward-backward [76, 100] method for solving (2.63).

(ii) Inertial versions of the forward-backward method were previously proposed and studied in [96], [77] and [10]. Asymptotic convergence of the forward-backward method proposed in [77] was proved in the latter reference, in particular, under the assumption: $0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < 1$, for all $k \geq 1$, and

$$\alpha = \alpha(\gamma) := 1 + \frac{\sqrt{9 - 4\gamma - 2\varepsilon\gamma} - 3}{\gamma},$$

for some $\varepsilon \in]0, (9 - 4\gamma)/(2\gamma)[$, where $\gamma \in (0, 2)$ and $\lambda_k \equiv \lambda := \gamma/L$ ($\gamma = 2\sigma^2$ in the notation of this chapter). The apparent limitation of this approach is that $\alpha \to 0$ if $\gamma \to 2$, i.e., the inertial effect degenerates for large values of the stepsize (see Fig. 1 in [77]). This contrasts to the approach proposed in this chapter, where the under-relaxation parameter $\tau \in [0, 1[$ is crucial to allowing α sufficiently close to 1, even for large stepsize values, i.e., when $\sigma \approx 1$ (see Assumption **(A)** and part of the discussion in Remark 2.2.7(iii)).

(iii) Algorithm 5 is a special instance (with constant relaxation) of the RIFB algorithm in [10]. We refer the reader to [10] (see, e.g., Theorems 3.8 and 3.15, and Remark 3.13) for a comprehensive discussion of the interplay and benefits of inertia and relaxation.

Next proposition shows that Algorithm 5 is also a special instance of Algorithm 2 for solving (2.63). Since the proof follows the same outline of [125, Proposition 5.3], we omit it here too.

Proposition 2.3.6. Let $\{\tilde{z}_k\}$, $\{z_k\}$, $\{w_k\}$ and $\{\lambda_k\}$ be generated by *Algorithm* 5, let $T = F + B$ be as in (2.63) and define, for all $k \geq 1$,

$$\varepsilon_k := \frac{\|\tilde{z}_k - w_{k-1}\|^2}{4L^{-1}} \quad \text{and} \quad v_k := \frac{w_{k-1} - \tilde{z}_k}{\lambda_k}. \tag{2.68}$$

Then, the following hold for all $k \geq 1$:

$$\begin{aligned} & v_k \in (F^{\varepsilon_k} + B)(\tilde{z}_k) \subset T^{\varepsilon_k}(\tilde{z}_k), \\ & \lambda_k v_k + \tilde{z}_k - w_{k-1} = 0, \quad 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - w_{k-1}\|^2, \quad z_k = w_{k-1} - \tau \lambda_k v_k. \end{aligned} \tag{2.69}$$

As a consequence of (2.69) and Algorithm 5's definition, it follows that Algorithm 5 is a special instance of Algorithm 2 for solving (2.63).

We finish this section by presenting the convergence and iteration-complexity of Algorithm 5, which are a direct consequence of Proposition 2.3.6 and Theorems 2.2.8, 2.2.11 and 2.2.13. We also mention that analogous remarks to those made in the Remark 2.3.4 also apply here.

Theorem 2.3.7 (convergence and iteration-complexity of *Algorithm* 5)**.** Under the *Assumption* (**A**) on $(\alpha, \sigma, \tau) \in [0, 1[\times]0, 1[\times]0, 1]$ and $\{\alpha_k\}$, let $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{\lambda_k\}$ be generated by Algorithm 5, let $\{v_k\}$ and $\{\varepsilon_k\}$ be as in (2.68) and let the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ be as in (2.42). Let also $\eta > 0$ and $q(\cdot)$ be as in (2.19) and (2.33), respectively, let d_0 denote the distance of z_0 to $(F + B)^{-1}(0)$ and assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Then, the following statements hold:

(a) The sequence $\{z_k\}$ converges weakly to a solution of the monotone inclusion problem (2.63).

(b) For all $k \geq 1$, there exists $i \in \{1, \ldots, k\}$ such that

$$\begin{aligned} & v_i \in (F^{\varepsilon_i} + B)(\tilde{z}_i), \\ & \|v_i\| \leq \frac{d_0}{\underline{\lambda} \tau \sqrt{k}} \sqrt{\eta^{-1} \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}, \\ & \varepsilon_i \leq \frac{\sigma d_0^2}{2(1-\sigma^2) \underline{\lambda} \tau\, k} \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right). \end{aligned} \tag{2.70}$$

(c) If, additionally, $\alpha_k \equiv \alpha$, then, for all $k \geq 1$,

$$\begin{aligned} & v_k^a \in (F + B)^{\varepsilon_k^a}(\tilde{z}_k^a), \\ & \|v_k^a\| \leq \frac{2(1+\alpha) d_0}{\underline{\lambda} \tau\, k} \sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \\ & \varepsilon_k^a \leq \frac{2\sqrt{2} d_0^2}{\underline{\lambda} \tau\, k} \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right) \left(1 + \frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta \tau^2}}\right). \end{aligned} \tag{2.71}$$

Chapter 3

A relative-error inertial-relaxed inexact projective splitting algorithm for structured monotone inclusion problems

In this chapter, we present an inertial and inexact variant of projective splitting method for solving structured monotone inclusion problems involving a sum of finitely many maximal monotone operators, which we will refer to as relative-error inertial-relaxed inexact projective splitting algorithm (Algorithm 7). The proposed algorithm benefits from a combination of inertial and relaxation effects, which are both controlled by parameters within a certain range. We propose sufficient conditions on these parameters (as well as we study the interplay between them) to ensure weak convergence of sequences generated by our algorithm. As an application of the proposed algorithm, we derive an inertial algorithm resembling the multi-block ADMM.

We mention that the content of this chapter is partially contained in the manuscript [83].

3.1 Problem statement

We consider the monotone inclusion problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* T_i G_i(z) \tag{3.1}$$

where $n \geq 2$ and the following assumptions hold:

(D1) For each $i = 1, \dots, n$, the operator $T_i : \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ is (set-valued) maximal monotone and $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ is a bounded linear operator.

(D2) The linear operator G_n is equal to the identity map in $\mathcal{H}_0 = \mathcal{H}_n$, i.e., $G_n : z \mapsto z$ for all $z \in \mathcal{H}_0$.

(D3) The solution set of (3.1) is nonempty, i.e., there exists at least one $z \in \mathcal{H}_0$ satisfying the inclusion in (3.1).

Let $\boldsymbol{\mathcal{H}} := \mathcal{H}_0 \times \mathcal{H}_1 \times \ldots \times \mathcal{H}_{n-1}$ be endowed with the inner product and norm defined, respectively, as follows (for some $\gamma > 0$):

$$\langle (z, w), (z', w') \rangle_\gamma = \gamma \langle z, z' \rangle + \sum_{i=1}^{n-1} \langle w_i, w_i' \rangle, \quad \|(z, w)\|_\gamma^2 = \gamma \|z\|^2 + \sum_{i=1}^{n-1} \|w_i\|^2, \tag{3.2}$$

where $z, z' \in \mathcal{H}_0$ and $w := (w_1, \ldots, w_{n-1}), w' := (w_1', \ldots, w_{n-1}') \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_{n-1}$.

Note that using Assumption (D2) above, we obtain that (3.1) can equivalently be written as

$$0 \in \sum_{i=1}^{n-1} G_i^* T_i G_i(z) + T_n(z). \tag{3.3}$$

Consider the extend solution set (or generalized Kuhn-Tucker set) for the problem (3.1) (or (3.3)):

$$\mathcal{S} := \left\{ (z, w_1, \ldots, w_{n-1}) \in \boldsymbol{\mathcal{H}} \;\middle|\; w_i \in T_i(G_i z),\ i = 1, \ldots, n-1,\ -\sum_{i=1}^{n-1} G_i^* w_i \in T_n(z) \right\}. \tag{3.4}$$

Problem (3.1) appears in different fields of applied mathematics and optimization, including machine learning, inverse problems and image processing (see e.g. [25, 44, 45] and references therein), especially in connection with the composed convex optimization problem

$$\min_{z \in \mathcal{H}_0} \sum_{i=1}^{n} f_i(G_i z) \tag{3.5}$$

where, for $i = 1, \ldots, n$, each $f_i : \mathcal{H}_i \to (-\infty, +\infty]$ is proper, convex and lower semicontinuos. Indeed, under mild assumptions on f_i and G_i, the minimization problem (3.5) is equivalent to the monotone inclusion problem (3.1) with $T_i = \partial f_i$ $(i = 1, \ldots, n)$.

As we already pointed out, one of the main goals of this chapter is to develop a projective splitting type-algorithm algorithm for solving (3.1) with both inertial and relaxation effects and, additionally, with inexact subproblems solution within relative-error criteria. We emphasize that, up to our knowledge, this is the first time in the literature that inertial effects are considered in projective splitting algorithms. Our main algorithm is Algorithm 7, for which the convergence is studied in Theorems 3.3.8 and 3.3.9, under flexible assumptions on the inertial and relaxation parameters. Motivated by the above discussion and the fact that (3.1) is equivalent to the problem of finding a point in the closed and convex set $\mathcal{S}$ (see Subsection 1.4.3), we first introduce in Section 3.2 an inertial-relaxed separator-projector method for solving the (feasibility) problem of finding points in closed convex subsets of Hilbert spaces, which unifies the ideas of the classical inertial PPM [15] and the separator-projector algorithm [40].

3.2 An inertial-relaxed separator-projection method

In this section, we propose and study a general separator-projection framework (Algorithm 6) for finding a point in a given closed and convex subset of a real Hilbert space. The main motivation for doing so is the reformulation of the monotone inclusion problem (3.1) (see Section 3.3) as the problem of finding a point in the extended solution set $\mathcal{S}$ in (3.4). Algorithm 6 will be used in Section 3.3 to analyze the convergence of the main algorithm proposed in this chapter (namely Algorithm 7) for solving (3.1).

Now, we present our first framework which combines inertia and relaxation, related to the method of finding a point in a convex and closed subset $\mathcal{S}$ of an arbitrary Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ whose norm is $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. We denote the gradient of an affine function $\varphi : \mathcal{H} \to \mathbb{R}$ by the usual notation $\nabla \varphi$ and, in this case, we also write $\varphi(z) = \langle \nabla \varphi, z \rangle + \varphi(0)$ for all $z \in \mathcal{H}$. The main results in this section are Propositions 3.2.3 and 3.2.4.

Algorithm 6. An inertial-relaxed linear separator-projection method for finding a point in a nonempty closed convex set $\mathcal{S} \subset \mathcal{H}$

(0) Let $p^0 = p^{-1} \in \mathcal{H}$, $\alpha \in [0, 1)$ and $0 < \underline{\beta} < \overline{\beta} < 2$ be given and let $k \leftarrow 0$.

(1) Choose $\alpha_k \in [0, \alpha]$ and define

$$\widehat{p}^k = p^k + \alpha_k (p^k - p^{k-1}). \tag{3.6}$$

(2) Find an affine function φ_k such that $\nabla \varphi_k \neq 0$ and $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$. Choose $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and set

$$p^{k+1} = \widehat{p}^k - \frac{\beta_k \max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla \varphi_k\|^2} \nabla \varphi_k. \tag{3.7}$$

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

Remark 3.2.1. We now make some remarks regarding Algorithm 6.

(i) Denoting by $\widetilde{p}^{k+1}$ the (orthogonal) projection of $\widehat{p}^k$ onto the semispace

$$H_k := \{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\}$$

which is a convex and closed subset of $\mathcal{H}$, by a straightforward calculation one obtains

$$\widetilde{p}^{k+1} = \widehat{p}^k - \frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla \varphi_k\|^2} \nabla \varphi_k \tag{3.8}$$

and using (3.7) we conclude that

$$p^{k+1} = \widehat{p}^k + \beta_k (\widetilde{p}^{k+1} - \widehat{p}^k). \tag{3.9}$$

From now on, let us refer to the step 1 and step 2 as inertial step and projection step, respectively.

(ii) Note that (3.6) and (3.9) illustrate the different effects promoted in Algorithm 6 by inertia and relaxation parameters, which are respectively controlled by α_k and β_k, see Figure 3.1. The relaxation parameter β_k determines the position of the update p^{k+1} in the closed segment between the current iterate (the inertial term) $\widehat{p}^k$ and its reflection $2\widetilde{p}^{k+1} - \widehat{p}^k$ with respect to hyperplane $\widehat{H}_k := \{p \in \mathcal{H} \mid \varphi_k(p) = 0\}$, see Figure 3.1.

(iii) When $\alpha = 0$, in which case $\alpha_k \equiv 0$ and hence $\widehat{p}^k = p^k$ in (3.6), then Algorithm 6 reduces to the generic *separator-projection method* of [40] (see also [68]). Conditions on $\{\alpha_k\}$, $\alpha \in [0,1)$ and $\overline{\beta} \in (0,2)$ to guarantee overall convergence of Algorithm 6 are given in Proposition 3.2.4; see (3.22), (3.23) and Figure 3.2.

(iv) As we mentioned early, Algorithm 6 will be used in the next section for analyzing the convergence of Algorithm 7.

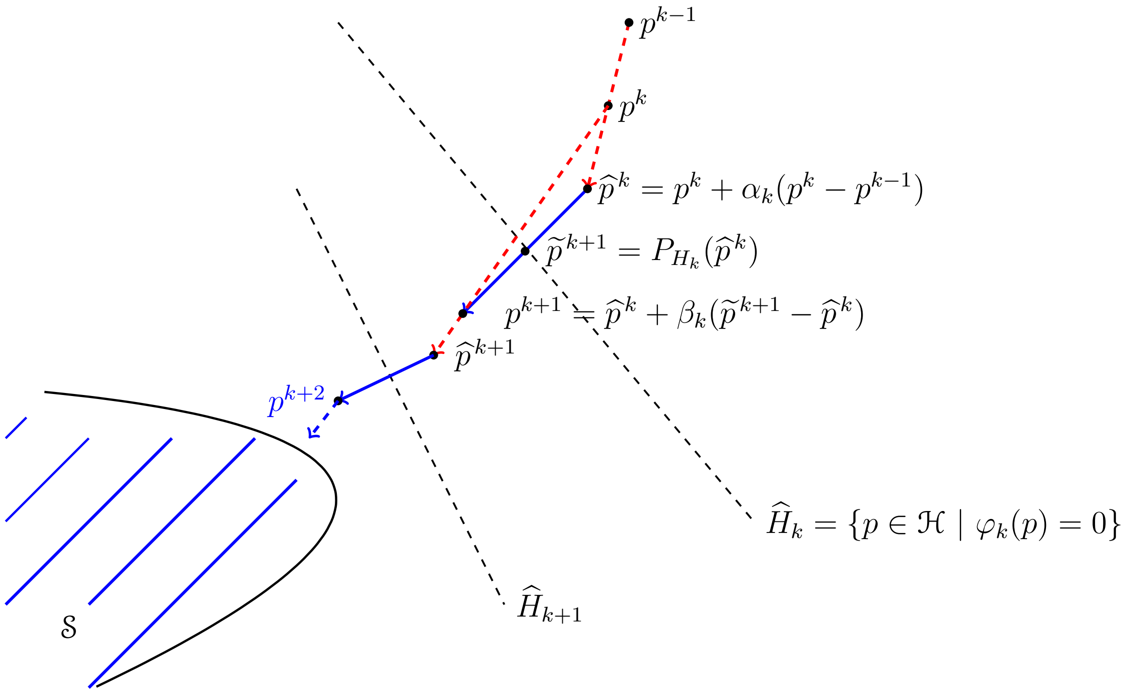

Figure 3.1: Geometric interpretation of steps (3.6) and (3.7) in Algorithm 6. The (overrelaxed) projection step (3.7) is orthogonal to the separating hyperplane $\widehat{H}_k = \{p \in \mathcal{H} \mid \varphi_k(p) = 0\}$, which can differ from the direction between p^{k-1}, p^k and $\widehat{p}^k$ when $\alpha_k > 0$.

It is well known that the standard convergence analysis of Algorithm 6 without inertial steps is based on the Fejér monotonicity of the iterate with respect to $\mathcal{S}$ (see [40]). However, the inertial case needs a special treatment, because in this case $\{p^k\}$ is no longer Fejér monotone [1]; see (3.14) below.

Given $p \in \mathcal{H}$ and a sequence $\{p^k\}$ generated by Algorithm 6, let us consider the sequence $\{h_k\}$ as $h_k = \|p^k - p\|^2$. The difference $h_{k+1} - [h_k + \alpha_k(h_k - h_{k-1})]$ plays an important role in the study of the asymptotic behavior as $k \to \infty$ of the sequence $\{p^k\}$ generated by Algorithm 6.

In what follows we show a technical lemma which will be crucial for the proof of the main results in this section (see Propositions 3.2.3 and 3.2.4).

---

[1] A sequence $\{p^k\}$ in $\mathcal{H}$ is Fejér monotone with respect to set $\mathcal{S}$, if for any $p^* \in \mathcal{S}$,

$$\|p^{k+1} - p^*\| \leq \|p^k - p^*\| \quad \forall k \geq 1.$$

Lemma 3.2.2. Consider the sequences evolved by *Algorithm* 6 and let $\widetilde{p}^{k+1}$ be as in (3.8). For an arbritary $p \in \mathcal{S}$, define

$$h_k = \|p^k - p\|^2 \qquad \forall k \geq -1. \tag{3.10}$$

Then, the following statements hold:

(a) For all $k \geq 0$,

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + s_{k+1} \leq \alpha_k(1+\alpha_k)\|p^k - p^{k-1}\|^2,$$

where

$$s_{k+1} := \beta_k(2-\beta_k)\|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 \qquad \forall k \geq 0. \tag{3.11}$$

(b) For all $k \geq 0$,

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) \leq \gamma_k\|p^k - p^{k-1}\|^2 - (2-\overline{\beta})\overline{\beta}^{-1}(1-\alpha_k)\|p^{k+1} - p^k\|^2, \tag{3.12}$$

where

$$\gamma_k := 2\left(1 - \overline{\beta}^{-1}\right)\alpha_k^2 + 2\overline{\beta}^{-1}\alpha_k \qquad \forall k \geq 0. \tag{3.13}$$

Proof. (a) We shall first prove that

$$\|p^{k+1} - p\|^2 + \beta_k(2-\beta_k)\|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 \leq \|\widehat{p}^k - p\|^2 \qquad \forall p \in \mathcal{S}, \tag{3.14}$$

where $\widetilde{p}^{k+1}$ is as in (3.8), i.e., it is the projection of $\widehat{p}^k$ onto the semispace $H_k = \{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\}$. To this end, note first that, for all $p \in \mathcal{S}$,

$$\begin{aligned}\|\widehat{p}^k - p\|^2 - \|\widetilde{p}^{k+1} - p\|^2 &= \|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 + 2\langle \widehat{p}^k - \widetilde{p}^{k+1}, \widetilde{p}^{k+1} - p\rangle \\ &\geq \|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 \end{aligned} \tag{3.15}$$

where we have used (3.8) and the fact that $\mathcal{S} \subset H_k$ (see Step 2 of Algorithm 6 and the property (1.1) of the projection operator) to obtain the inequality $\langle \widehat{p}^k - \widetilde{p}^{k+1}, \widetilde{p}^{k+1} - p\rangle \geq 0$. Note now that (3.9) is trivially equivalent to

$$p^{k+1} = (1-\beta_k)\widehat{p}^k + \beta_k\widetilde{p}^{k+1},$$

which in turn combined with (A.7) in Lemma A.2.1 yields

$$\|p^{k+1} - p\|^2 = (1-\beta_k)\|\widehat{p}^k - p\|^2 + \beta_k\|\widetilde{p}^{k+1} - p\|^2 - \beta_k(1-\beta_k)\|\widehat{p}^k - \widetilde{p}^{k+1}\|^2$$

or, equivalently,

$$\beta_k\left(\|\widehat{p}^k - p\|^2 - \|\widetilde{p}^{k+1} - p\|^2\right) = \|\widehat{p}^k - p\|^2 - \beta_k(1-\beta_k)\|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 - \|p^{k+1} - p\|^2. \tag{3.16}$$

The desired inequality (3.14) now follows by multiplying the inequality in (3.15) by $\beta_k \geq 0$, by combining the resulting inequality with (3.16) and by using some simple algebraic manipulations.

Now, from (3.6) we have

$$p^k - p = \frac{1}{1+\alpha_k}(\widehat{p}^k - p) + \frac{\alpha_k}{1+\alpha_k}(p^{k-1} - p) \ \text{ and } \ \widehat{p}^k - p^{k-1} = (1+\alpha_k)(p^k - p^{k-1}). \qquad (3.17)$$

Using (A.7) (Lemma A.2.1) and the first identity in (3.17) we obtain

$$\|p^k - p\|^2 = \frac{1}{1+\alpha_k}\|\widehat{p}^k - p\|^2 + \frac{\alpha_k}{1+\alpha_k}\|p^{k-1} - p\|^2 - \frac{\alpha_k}{(1+\alpha_k)^2}\|\widehat{p}^k - p^{k-1}\|^2,$$

which combined with the second identity in (3.17) and some algebraic manipulations gives

$$\|\widehat{p}^k - p\|^2 = (1+\alpha_k)\|p^k - p\|^2 - \alpha_k\|p^{k-1} - p\|^2 + \alpha_k(1+\alpha_k)\|p^k - p^{k-1}\|^2. \qquad (3.18)$$

Hence, (a) follows directly from (3.14), (3.18) and the definitions of h_k and s_{k+1} in (3.10) and (3.11), respectively.

(b) Note that (3.9) is also trivially equivalent to $\widehat{p}^k - \widetilde{p}^{k+1} = \beta_k^{-1}(\widehat{p}^k - p^{k+1})$, which in turn combined with the definition of s_{k+1} in (3.11) and the fact that $\beta_k \leq \overline{\beta}$ for all $k \geq 0$ (see Step 2 of Algorithm 6) yields

$$s_{k+1} = \beta_k(2-\beta_k)\|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 = \left(2\beta_k^{-1} - 1\right)\|\widehat{p}^k - p^{k+1}\|^2 \geq \left(2\overline{\beta}^{-1} - 1\right)\|\widehat{p}^k - p^{k+1}\|^2. \quad (3.19)$$

Using (3.6), the Cauchy-Schwarz inequality, the Young inequality ($2ab \leq a^2 + b^2$ with $a = \|p^{k+1} - p^k\|$ and $b = \|p^k - p^{k-1}\|$) and some algebraic manipulations, we find

$$\begin{aligned}
\|\widehat{p}^k - p^{k+1}\|^2 =& \|p^{k+1} - p^k\|^2 + \alpha_k^2\|p^k - p^{k-1}\|^2 - 2\alpha_k\langle p^{k+1} - p^k, p^k - p^{k-1}\rangle \\
\geq& \|p^{k+1} - p^k\|^2 + \alpha_k^2\|p^k - p^{k-1}\|^2 - 2\alpha_k\|p^{k+1} - p^k\|\|p^k - p^{k-1}\| \\
\geq& \|p^{k+1} - p^k\|^2 + \alpha_k^2\|p^k - p^{k-1}\|^2 - \alpha_k\left(\|p^{k+1} - p^k\|^2 + \|p^k - p^{k-1}\|^2\right) \\
=& (1-\alpha_k)\left(\|p^{k+1} - p^k\|^2 - \alpha_k\|p^k - p^{k-1}\|^2\right). \qquad (3.20)
\end{aligned}$$

From (3.19) and (3.20) we obtain

$$s_{k+1} \geq \left(2\overline{\beta}^{-1} - 1\right)(1-\alpha_k)\left(\|p^{k+1} - p^k\|^2 - \alpha_k\|p^k - p^{k-1}\|^2\right),$$

which in turn combined with the inequality in (a) and (3.13), and after some simple manipulations, gives exactly the desired inequality in (b). □

Next is our first result on the (asymptotic) convergence of Algorithm 6. The key assumption is the summability condition (3.21), for which a sufficient condition, only depending on the parameters α_k and β_k will be given in Proposition 3.2.4 (see conditions (3.22), (3.23) and Figure 3.2).

Proposition 3.2.3 (First result on the convergence of *Algorithm* 6)**.** Let $\{p^k\}$, $\{\widehat{p}^k\}$, $\{\varphi_k\}$ and $\{\alpha_k\}$ be generated by Algorithm 6 and assume that

$$\sum_{k=0}^{\infty} \alpha_k\|p^k - p^{k-1}\|^2 < +\infty. \qquad (3.21)$$

Then, the following holds:

(a) $\lim_{k\to\infty} \|p^k - p\|$ exists for all $p \in \mathcal{S}$, and hence $\{p^k\}$ and $\{\widehat{p}^k\}$ are bounded.

(b) If every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$, then $\{p^k\}$ converges weakly to some element in $\mathcal{S}$.

(c) We have,

$$\frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla \varphi_k\|} \to 0 \qquad k \to +\infty.$$

Proof. (a) Defining $\delta_k = \alpha_k(1+\alpha_k)\|p^k - p^{k-1}\|^2$ and using Lemma 3.2.2(a) we conclude that condition (A.1) in Lemma A.1.1 (Appendix A.1) holds with h_k and s_{k+1} is as in (3.10) and (3.11), respectively. Hence, using the assumption (3.21), Lemma A.1.1(b) and (3.10) we conclude that

$$\lim_{k\to\infty} \|p^k - p\| \quad \text{exists for all} \quad p \in \mathcal{S}.$$

Further, this show the first part of Opial's lemma (Lemma A.2.2), and in particular $\{p^k\}$ and $\{\widehat{p}^k\}$ are bounded (see (3.6)).

(b) The conclusion of this item follows immediately from item (a) (first part) and Opial's lemma (Lemma A.2.2).

(c) Note first that from (3.8) we have

$$\frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla \varphi_k\|} = \|\widetilde{p}^{k+1} - \widehat{p}^k\|.$$

Hence, to conclude the proof of this item, it suffices to prove that $\|\widetilde{p}^{k+1} - \widehat{p}^k\| \to 0$, as $k \to \infty$. To this end, note that (3.21) combined with the definition of δ_k above, the fact that $\alpha_k^2 \leq \alpha_k$ (because $\alpha_k \in [0, 1[$) and Lemma A.1.1(a) gives $\sum_{k=0}^{+\infty} s_{k+1} < +\infty$, where s_{k+1} (for all $k \geq 0$) is as in (3.11), and so $s_{k+1} \to 0$, as $k \to \infty$. The desired result now follows form this fact, (3.11) and the fact that $0 < \underline{\beta} \leq \beta_k \leq \overline{\beta} < 2$ (see Step 2 in Algorithm 6). □

Next we present sufficient conditions on $\{\alpha_k\}$ and $\{\beta_k\}$ to ensure condition (3.21) in Proposition 3.2.3.

Proposition 3.2.4 (Second result on the convergence of *Algorithm* 6)**.** Let $\{p^k\}$ and $\{\alpha_k\}$ be generated by *Algorithm 6*. Assume that $\alpha \in [0, 1)$, $\overline{\beta} \in (0, 2)$ and $\{\alpha_k\}$ satisfy the following *(*for some $\overline{\alpha} > 0$*)*:

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \overline{\alpha} < 1 \qquad \forall k \geq 0 \tag{3.22}$$

and

$$\overline{\beta} = \overline{\beta}(\overline{\alpha}) := \frac{2(\overline{\alpha} - 1)^2}{2(\overline{\alpha} - 1)^2 + 3\overline{\alpha} - 1}. \tag{3.23}$$

Then, the following holds:

(a) We have

$$\sum_{k=0}^{\infty} \|p^k - p^{k-1}\|^2 < \infty. \tag{3.24}$$

(b) Under the assumptions (3.22) and (3.23), if every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$, then $\{p^k\}$ converges weakly to some element in $\mathcal{S}$.

Proof. (a) Define, for all $k \geq 0$,

$$\mu_k = h_k - \alpha_k h_{k-1} + \gamma_k \|p^k - p^{k-1}\|^2 \tag{3.25}$$

where h_k is as in (3.10) (for some $p \in \mathcal{S}$) and γ_k is as in (3.13). Using the assumption (3.22) and Lemma 3.2.2(b), we obtain, for all $k \geq 0$,

$$\begin{aligned}
\mu_{k+1} - \mu_k &\leq h_{k+1} - \alpha_k h_k + \gamma_{k+1}\|p^{k+1} - p^k\|^2 - h_k + \alpha_k h_{k-1} - \gamma_k \|p^k - p^{k-1}\|^2 && \text{[by (3.22)]}\\
&= h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + \gamma_{k+1}\|p^{k+1} - p^k\|^2 - \gamma_k\|p^k - p^{k-1}\|^2 \\
&\leq \left[-\left(2 - \overline{\beta}\right)\overline{\beta}^{-1}(1-\alpha_k) + \gamma_{k+1}\right]\|p^{k+1} - p^k\|^2 && \text{[by Lemma 3.2.2(b)]}\\
&\leq \left[-\left(2 - \overline{\beta}\right)\overline{\beta}^{-1}(1-\alpha_{k+1}) + \gamma_{k+1}\right]\|p^{k+1} - p^k\|^2 && \text{[by (3.22)]}\\
&= -q(\alpha_{k+1})\|p^{k+1} - p^k\|^2 && \text{[by (3.13) and (3.27)]} \quad (3.26)
\end{aligned}$$

where

$$q(\nu) := 2\left(\overline{\beta}^{-1} - 1\right)\nu^2 - \left(4\overline{\beta}^{-1} - 1\right)\nu + 2\overline{\beta}^{-1} - 1, \quad \nu \in \mathbb{R}. \tag{3.27}$$

Next we will show that $q(\alpha_{k+1})$ admits an uniform lower bound. To this end, note first that (3.23) and Lemma A.1.3 yield

$$\overline{\alpha} = \frac{2(2 - \overline{\beta})}{4 - \overline{\beta} + \sqrt{16\overline{\beta} - 7\overline{\beta}^2}},$$

which in turn combined with Lemma A.1.4 below implies that $q(\overline{\alpha}) = 0$ and $q(\cdot)$ is decreasing in $[0, \overline{\alpha}]$. Thus, in view of (3.22), we obtain

$$q(\alpha_{k+1}) \geq q(\alpha) > q(\overline{\alpha}) = 0$$

and so, in view of (3.26), it follows that

$$\mu_{k+1} - \mu_k \leq -q(\alpha_{k+1})\|p^{k+1} - p^k\|^2 \leq -q(\alpha)\|p^{k+1} - p^k\|^2 \leq 0, \tag{3.28}$$

and

$$\|p^{k+1} - p^k\|^2 \leq \frac{1}{q(\alpha)}(\mu_k - \mu_{k+1}) \qquad \forall k \geq 0.$$

Hence, for all $k \geq 0$,

$$\sum_{j=0}^{k} \|p^{j+1} - p^j\|^2 \leq \frac{1}{q(\alpha)}(\mu_0 - \mu_{k+1}) \leq \frac{1}{q(\alpha)}(\mu_0 + \alpha h_k) \tag{3.29}$$

where in the second inequality above we also used the fact that $\mu_{k+1} \geq -\alpha h_k$ (in view of (3.25) and (3.22)). Therefore, to finish the proof of (a) it is enough to find an upper bound on h_k and use (3.29). To this end, note that from (3.28), (3.25), (3.22) and the fact that $\gamma_k \geq 0$ (see (3.13)) we have, for all $k \geq -1$,

$$\begin{aligned}\mu_0 \geq \mu_1 \geq \ldots \geq \mu_{k+1} &= h_{k+1} - \alpha_{k+1} h_k + \gamma_{k+1}\|p^{k+1} - p^k\|^2 \\ &\geq h_{k+1} - \alpha h_k\end{aligned}$$

and so, for all $k \geq -1$,

$$\begin{aligned}h_{k+1} &\leq \alpha^{k+1} h_0 + \left(\sum_{i=0}^{k} \alpha^i\right) \mu_0 \\ &\leq h_0 + \frac{\mu_0}{1-\alpha}\end{aligned} \tag{3.30}$$

where in the second inequality we also used the fact – from (3.25) – that $\mu_0 = (1-\alpha_0)h_0 \geq 0$.

(b) The result follows trivially from (a), the fact that $0 \leq \alpha_k < 1$ for all $k \geq 0$ and Proposition 3.2.3(b). $\square$

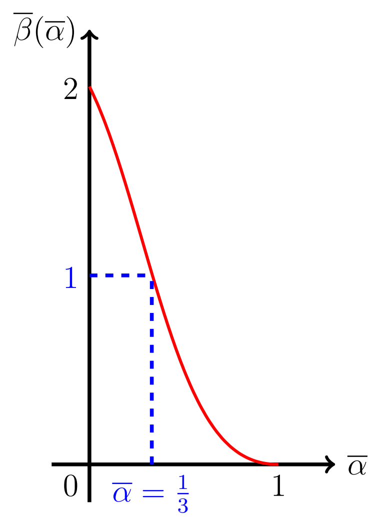

Figure 3.2: Relaxation parameter upper bound $\overline{\beta}(\overline{\alpha})$ defined as in (3.23) as a function of inertial step upper bound $\overline{\alpha} > 0$ of (3.22).

Next proposition will be useful to obtain convergence rates for Algorithm 7 in the next section.

Proposition 3.2.5. Let $\{p^k\}$, $\{\widehat{p}^k\}$ and $\{\alpha_k\}$ be generated by *Algorithm* 6 and assume that (3.22) and (3.23) hold. Then, for any $p \in \mathcal{S}$,

$$\sum_{j=0}^{k} \|\widetilde{p}^{j+1} - \widehat{p}^j\|^2 \leq \underline{\beta}^{-1}\left(2 - \overline{\beta}\right)^{-1}\left(1 + \frac{\alpha(1+2\alpha-\alpha^3)}{(1-\alpha)^2 q(\alpha)}\right)\|p^0 - p\|^2, \tag{3.31}$$

where

$$q(\alpha) := 2\left(\overline{\beta}^{-1} - 1\right)\alpha^2 - \left(4\overline{\beta}^{-1} - 1\right)\alpha + 2\overline{\beta}^{-1} - 1, \quad \alpha \in \mathbb{R}. \tag{3.32}$$

Proof. Let, for all $k \geq 0$,

$$\mu_k = h_k - \alpha_k h_{k-1} + \gamma_k \|p^k - p^{k-1}\|^2.$$

Using the fact that $p^0 = p^{-1}$ and $\alpha_k \in [0, 1)$, one has

$$\mu_0 = h_0 - \alpha_0 h_{-1} = (1 - \alpha_0) h_0 \leq h_0 = \|p^0 - p\|^2.$$

Now, in view of (3.29), (3.30) (inside the proof of Proposition 3.2.4) and the latter inequality, we get

$$\begin{aligned} \sum_{j=0}^{k} \|p^{j+1} - p^j\|^2 &\leq \frac{1}{q(\alpha)} (\mu_0 + \alpha h_k) \leq \frac{1}{q(\alpha)} \left(\mu_0 + \alpha h_0 + \alpha \frac{\mu_0}{1-\alpha} \right) \\ &\leq \frac{1 + \alpha - \alpha^2}{q(\alpha)(1-\alpha)} \|p^0 - p\|^2. \end{aligned} \tag{3.33}$$

On the other hand, from Lemma 3.2.2(a) and Lemma A.1.1 (a) with $\delta_k = \alpha_k(1+\alpha_k)\|p^k - p^{k-1}\|^2$ and s_{k+1} as in (3.11) and using (3.22) and (3.33) we have

$$\begin{aligned} \sum_{j=0}^{k-1} s_{j+1} \leq h_0 - h_k + \frac{1}{1-\alpha} \sum_{j=1}^{k-1} \delta_j &\leq h_0 + \frac{\alpha(1+\alpha)}{1-\alpha} \sum_{j=1}^{k} \|p^j - p^{j-1}\|^2 \\ &\leq \left(1 + \frac{\alpha(1 + 2\alpha - \alpha^3)}{(1-\alpha)^2 q(\alpha)} \right) \|p^0 - p\|^2. \end{aligned} \tag{3.34}$$

In the second "$\leq$" we used the fact that $h_k \geq 0$, for all $k \geq 0$. By using the definition of s_{k+1} in (3.11) and then adding on $j = 0, \ldots, k-1$ and taking into account that $0 < \underline{\beta} \leq \beta_k \leq \overline{\beta} < 2$, one gets

$$\sum_{j=0}^{k-1} s_{j+1} = \sum_{j=0}^{k-1} \beta_j (2 - \beta_j) \|\widetilde{p}^{j+1} - \widehat{p}^j\|^2 \geq \underline{\beta} \left(2 - \overline{\beta}\right) \sum_{j=0}^{k-1} \|\widetilde{p}^{j+1} - \widehat{p}^j\|^2. \tag{3.35}$$

The desired conclusion follows directly by combining (3.34) and (3.35). □

To close this section, we state some further remarks about the analysis of Algorithm 6:

Remark 3.2.6.

(i) Conditions (3.22) and (3.23) on $\{\alpha_k\}$, α and $\overline{\beta}$ guarantee that the summability condition (3.21) in Theorem 6 is satisfied, thus guaranteeing the convergence of Algorithm 6. Similar conditions appear in the convergence analysis of inertial proximal point methods, see e.g., [8, 10, 6]. Since Algorithm 6 is the basis of the projective splitting method developed in the next section, conditions (3.22) and (3.23) will also play an important role in their convergence analysis.

(ii) If we set $\overline{\alpha} = 1/3$ in (3.22), one gets directly from (3.23) that $\overline{\beta} = 1$. On the other hand, we observe that $\overline{\beta} > 1$ whenever $\overline{\alpha} < 1/3$ (see Figure 3.2 above). A standard strategy in the literature of proximal point methods is to set $\overline{\alpha} = 1/3$, see e.g. [4, 36]. We also emphasize that the interplay between inertial and relaxation effects has also been investigated, e.g., in [6, 10, 11, 43].

(iii) Algorithm 6 is a general and abstract template. However, this template by itself does not guarantee the weak convergence of its iterates $\{p^k\}$ to a point of $\mathcal{S}$, because the separator function φ_k might not be chosen to actually separate the extrapolated term $\widehat{p}^k$ from $\mathcal{S}$. To ensure overall convergence, we assumed that every weak sequential cluster point of the sequence $\{p^k\}$ belong to $\mathcal{S}$ (see Propositionss 3.2.3 and 3.2.4). However, in concrete problems this assumption can be verified. For instance, the analysis of the projective splitting methods in [51, 52](without inertia) ensures weak convergence of the iterates under the condition $\varphi_k(p^k) \geq \xi\|\nabla\varphi_k\|^2$ for all $k \geq 0$, where $\xi > 0$ is a constant (see e.g. [52, Proposition 3.2]). In the next section, we will see that this condition is also verified in the inertial case (see Lemma 3.3.6(c)).

3.3 A relative-error inertial-relaxed inexact projective splitting algorithm

In this section, we present a specialization of the general separator-projector methods (Algorithm 6) to problem (3.36) below. We propose and study the asymptotic convergence of a relative-error inertial-relaxed inexact projective splitting algorithm (Algorithm 7). The main convergence results are stated in Theorems 3.3.8, 3.3.9 and 3.3.10.

We recall the monotone inclusion problem (3.1), i.e., the problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* T_i G_i(z) \tag{3.36}$$

where $n \geq 2$ and assumptions (D1)-(D3) in Section 3.1 are assumed to hold.

Since Step 2 of Algorithm 6 demands the construction of an (non constant) affine function φ_k such that $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$, next we discuss the construction of such φ_k satisfying the latter inequality for $\mathcal{S}$ defined as in (3.4).

Let $p := (z, w_1, \dots, w_{n-1})$ be a generic point in $\boldsymbol{\mathcal{H}}$, for $y_i^k \in T_i(x_i^k)$ $(i = 1, \dots, n)$. We define $\varphi_k : \boldsymbol{\mathcal{H}} \to \mathbb{R}$ by

$$\varphi_k(\underbrace{z, w_1, \dots, w_{n-1}}_{p}) = \sum_{i=1}^{n-1} \langle G_i z - x_i^k, y_i^k - w_i \rangle + \langle G_n z - x_n^k, y_n^k + \sum_{i=1}^{n-1} G_i^* w_i \rangle. \tag{3.37}$$

To facilitate the mathematical presentation, we use the following notation in the rest of the chapter:

$$w_n := -\sum_{i=1}^{n-1} G_i^* w_i. \tag{3.38}$$

Hence, the separator function (3.37) may be written more brielfy as:

$$\varphi_k(p) = \sum_{i=1}^{n} \langle G_i z - x_i^k, y_i^k - w_i \rangle. \tag{3.39}$$

Note that φ_k depends on the computation of pairs (x_i^k, y_i^k) in the graph of T_i, for each $i = 1, \ldots, n$, which can be computed by inexact evaluation (with relative-error tolerance) of the resolvent $J_{T_i} = (T_i + I)^{-1}$ of T_i (see Step 2 of Algorithm 7).

The following lemma presents some properties of φ_k which will be useful in the remainder of this chapter.

Lemma 3.3.1. [68, Lemma 4] Let φ_k and $\mathcal{S}$ be as in (3.37) and (3.4), respectively. The following hold:

(a) φ_k is an affine function on $\boldsymbol{\mathcal{H}}$.

(b) $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$.

(c) The gradient of φ_k with respect to the inner product $\langle \cdot, \cdot \rangle_\gamma$ as in (3.2) is

$$\nabla \varphi_k = \left(\frac{1}{\gamma} \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), x_1^k - G_1 x_n^k, \ldots, x_{n-1}^k - G_{n-1} x_n^k \right), \tag{3.40}$$

and

$$\|\nabla \varphi_k\|_\gamma^2 = \gamma^{-1} \left\| \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right\|^2 + \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2. \tag{3.41}$$

(d) If $\nabla \varphi_k = 0$, then $(x_n^k, y_1^k, \ldots, y_{n-1}^k) \in \mathcal{S}$ and in particular x_n^k is a solution of the problem (3.36).

3.3.1 Our main Algorithm

In this subsection we present the main algorithm of this chapter. As we mentioned before, it consists of a relative-error inertial-relaxed inexact projective splitting method for solving (3.36).

Algorithm 7. A relative-error inertial-relaxed inexact projective splitting algorithm

(0) Let $(z^{-1}, w_1^{-1}, \dots, w_{n-1}^{-1}) = (z^0, w_1^0, \dots, w_{n-1}^0) \in \boldsymbol{\mathcal{H}}$, $\alpha, \sigma \in [0,1)$, $0 < \underline{\beta} \le \overline{\beta} < 2$ and $\gamma > 0$ be given and let $k \leftarrow 0$.

(1) Choose $\alpha_k \in [0, \alpha]$ and let

$$\widehat{z}^k = z^k + \alpha_k(z^k - z^{k-1}), \tag{3.42}$$

$$\widehat{w}_i^k = w_i^k + \alpha_k(w_i^k - w_i^{k-1}), \quad i = 1, \dots, n-1, \tag{3.43}$$

$$\widehat{w}_n^k = -\sum_{i=1}^{n-1} G_i^* \widehat{w}_i^k. \tag{3.44}$$

(2) Choose scalars $\rho_i^k > 0$ and compute (x_i^k, y_i^k), for $i = 1, \dots, n$ satisfying

$$y_i^k \in T_i(x_i^k), \quad x_i^k + \rho_i^k y_i^k = G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k + e_i^k \tag{3.45}$$

and

$$\|e_i^k\|^2 \le \sigma^2 \left(\|G_i \widehat{z}^k - x_i^k\|^2 + \|\rho_i^k (y_i^k - \widehat{w}_i^k)\|^2 \right). \tag{3.46}$$

(3) **(3.a)** If $x_i^k = G_i x_n^k$, $i = 1, \dots, n-1$ and $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$, then STOP and set

$$z^{k+1} = x_n^k \quad \text{and} \quad w_i^{k+1} = y_i^k, \quad i = 1, \dots, n-1. \tag{3.47}$$

(3.b) Else, define

$$\varphi_k(\underbrace{z, w_1, \dots, w_{n-1}}_{p}) = \sum_{i=1}^{n-1} \langle G_i z - x_i^k, y_i^k - w_i \rangle + \langle G_n z - x_n^k, y_k^n + \sum_{i=1}^{n-1} G_i^* w_i \rangle, \tag{3.48}$$

$$\theta_k = \frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\gamma^{-1} \|\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k\|^2 + \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2}. \tag{3.49}$$

(4) Choose some relaxation parameter $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and define

$$z^{k+1} = \widehat{z}^k - \gamma^{-1} \beta_k \theta_k \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), \tag{3.50}$$

$$w_i^{k+1} = \widehat{w}_i^k - \beta_k \theta_k \left(x_i^k - G_i x_n^k \right), \quad i = 1, \dots, n-1. \tag{3.51}$$

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

Let us adopt a similar nomenclature to that of [68] for the parameters:

- For each $k \ge 0$ and each $i = 1, \dots, n$, ρ_i^k is a positive scalar parameter.

- For each $k \geq 0$, $\beta_k \in [\underline{\beta}, \overline{\beta}]$ is a relaxation parameter with $0 < \underline{\beta} \leq \overline{\beta} \leq 2$.
- For each $k \geq 0$, $\alpha_k \in [0, \overline{\alpha})$, for some $\overline{\alpha} \in [0, 1)$ is the inertial(extrapolation) parameter.
- The parameter $\gamma > 0$ controls the relative emphasis on the primal and dual variables in the projection step (3.50) and (3.51) (see (3.2) for more details on the parameter γ).
- The parameter of relative-error tolerance $\sigma \in [0, 1)$ and the sequence of errors $\{e_i^k\}$ models the inexact computations of resolvent step in (3.45).

Moreover, we adopt the following assumption on the step-size parameter ρ_i^k in the (approximate) proximal step (3.45):

(D4) Stepsize condition for convergence of Algorithm 7:

$$0 < \underline{\rho} \leq \rho_i^k \leq \overline{\rho} < +\infty, \tag{3.52}$$

where

$$\underline{\rho} := \min_{i=1,\dots,n} \left\{ \inf_{k \geq 0} \rho_i^k \right\} \qquad \text{and} \quad \overline{\rho} := \max_{i=1,\dots,n} \left\{ \sup_{k \geq 0} \rho_i^k \right\}.$$

Before discussing the convergence analysis of Algorithm 7, we make some remarks concerning it.

Remark 3.3.2.

(i) Notice that if $x_i^k = G_i x_n^k$, for $i = 1, \dots, n-1$ and $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$, for some k, that is, $\nabla \varphi_k = 0$, then $(x_n^k, y_1^k, \dots, y_{n-1}^k) \in \mathcal{S}$ by Lemma 3.3.1(d). Therefore, when Algorithm 7 stops at step (3.a), it stops at a point in the extended solution set $\mathcal{S}$. Even more, the test in step (3.a) guarantees that the denominator in (3.49) cannot be zero.

(ii) Similarly to Algorithm 6 of Section 3.2, Algorithm 7 also promotes inertial and relaxation effects, controlled by the parameters α_k and β_k, respectively. The inertial (extrapolation) step is performed in (3.42) and (3.43), while the relaxed projective step is given in (3.50) and (3.51) (in the context of Algorithm 6, see Figure 3.1 of Section 3.2). Conditions on the choice of the upper bounds α and $\overline{\beta}$, as well as on the sequence of extrapolation parameters $\{\alpha_k\}$, to guarantee the convergence of Algorithm 7 will be given in Theorem 3.3.9.

(iii) We also emphasize that if $\alpha_k = 0$ in Algorithm 7, then it reduces to the projective splitting algorithm (or some of its variants), see e.g. [68, Algorithm 2] with $\mathcal{I}_F = \emptyset$, $d(i, k) = k$ and $I_k = \{1, \dots, n\}$ (in the notation of the latter reference). In particular, when $n = 2$, Algorithm 7 results in an inertial and inexact variant of the primal-dual splitting algorithm in [2, Proposition 3.2].

(iv) The computation of (x_i^k, y_i^k) in (3.45) can be performed inexactly within a relative-error tolerance controlled by the parameter $\sigma \in [0, 1)$. In practice, the error condition in (3.45) can be used as a stopping-criterion for some computational procedure (e.g., conjugate gradient algorithm) applied to (inexactly) solving the related inclusion (for $i = 1, \ldots, n$)

$$0 \in \rho_i^k T_i(x) + x - (G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k)$$

until the error-condition in (3.46) is satisfied for the first time. Note also that (x_i^k, y_i^k) is given explicitly by

$$x_i^k = J_{\rho_i^k T_i}\left(G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k\right) \quad \text{and} \quad y_i^k = \frac{G_i \widehat{z}^k - x_i^k}{\rho_i^k} + \widehat{w}_i^k,$$

whenever the resolvent $J_{\rho_i^k T_i} = (\rho_i^k T + I)^{-1}$ is assumed to be easily computed and $\sigma = 0$ in (3.45). In the particular case of the minimization problem (3.5), the computation of (x_i^k, y_i^k) reduces to the (inexact) computation of the proximity operator $\operatorname{prox}_{\rho_i^k f}(x)$ (see (1.3)), i.e., in this case

$$x_i^k \approx \underset{z \in \mathcal{H}_0}{\operatorname{argmin}} \left\{ f_i(z) + \frac{1}{2\rho_i^k} \| z - (G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k) \|^2 \right\}.$$

(vi) By setting $p^k := (z^k, w_1^k, \ldots, w_{n-1}^k)$ for all $k \geq -1$, and defining $\widehat{p}^k$ as in (3.6), after some algebraic manipulations, and using (3.42) and (3.43) one can check that $\widehat{p}^k = (\widehat{z}^k, \widehat{w}_i^k, \ldots, \widehat{w}_{n-1}^k)$. Indeed,

$$\begin{aligned}
\widehat{p}^k =& p^k + \alpha_k (p^k - p^{k-1}) \\
=& (z^k, w_1^k, \ldots, w_{n-1}^k) + \alpha_k \left((z^k, w_1^k, \ldots, w_{n-1}^k) + (z^{k-1}, w_1^{k-1}, \ldots, w_{n-1}^{k-1}) \right) \\
=& \left(z^k + \alpha_k (z^k - z^{k-1}), w_1^k + \alpha_k (w_1^k - w_1^{k-1}), \ldots, w_{n-1}^k + \alpha_k (w_{n-1}^k - w_{n-1}^{k-1}) \right) \\
=& (\widehat{z}^k, \widehat{w}_i^k, \ldots, \widehat{w}_{n-1}^k).
\end{aligned}$$

(vii) Direct substitution of (3.43) into (3.44) gives that, similarly to $\widehat{w}_i^k$ $(i = 1, \ldots, n-1)$, $\widehat{w}_n^k$ also satisfies

$$\widehat{w}_n^k = w_n^k + \alpha_n (w_n^k - w_n^{k-1}),$$

where

$$w_n^k := -\sum_{i=1}^{n-1} G_i^* w_i^k \qquad \forall k \geq 0. \tag{3.53}$$

(viii) The assumption on relative-error condition in (3.46) adopted in this chapter is more flexible in terms of computational implementation when compared to the error criterion considered in the existent projective splitting methods (see e.g., [42, 52, 70, 68]).

Next, we deal with the situation when Algorithm 7 terminates at step (3.a).

Lemma 3.3.3. If *Algorithm 7* stops at step (3.a), then $(z^{k+1}, w_1^{k+1}, \ldots, w_{n-1}^{k+1})$ as in (3.47) belongs to the extended solution set $\mathcal{S}$ defined in (3.4). In particular, z^{k+1} is a solution of (3.36).

Proof. Let φ_k defined as in (3.37). First, note that if Algorithm 7 stop at step (3.a), we have the following

$$x_i^k - G_i x_n^k = 0, \quad i = 1, \dots, n-1 \quad \text{and} \quad \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$$

where $y_i^k \in T_i x_i^k$, for $i = 1, \dots, n$. The last equality and Lemma 3.3.1(c) implies that $\nabla\varphi_k = 0$. Thus, by invoking Lemma 3.3.1(d) and thanks to (3.47) we conclude that $(z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1}) = (x_n^k, y_1^k, \dots, y_{n-1}^k) \in \mathcal{S}$, and as a consequence z^{k+1} solve (3.36). $\square$

Remark 3.3.4. Lemma 3.3.3 asserts that if Algorithm 7 terminate finitely (Algorithm 7 stops at step 3(a)), then the final iterate is a solution of (3.36). From now on, we assume that Algorithm does not stop at step (3.a), i.e., we assume that Algorithm 7 generates infinite sequences.

Next we show that Algorithm 7 (under the assumption that it never stops at Step 3) is a special instance of Algorithm 6 for finding a point in $\mathcal{S}$ as in (3.4) in the Hilbert space $\boldsymbol{\mathcal{H}}$ endowed with the inner product and norm as in (3.2).

Proposition 3.3.5. Assume that *Algorithm* 7 does not stop at step (3.a), let $\{z^k\}, \{w_1^k\}, \dots, \{w_{n-1}^k\}$ be generated by *Algorithm 7*, let $\{\varphi_k\}$ be as in (3.48) and define

$$p^k = (z^k, w_1^k, \dots, w_{n-1}^k) \qquad \forall k \geq -1. \tag{3.54}$$

Then, the following statement hold:

(i) For all $k \geq 0$, $\nabla\varphi_k \neq 0$ and $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$.

(ii) For all $k \geq 0$,

$$\widehat{p}^k = (\widehat{z}^k, \widehat{w}_i^k, \dots, \widehat{w}_{n-1}^k) \ \ \textit{and} \ \ p^{k+1} = \widehat{p}^k - \frac{\beta_k \max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla\varphi_k\|_\gamma^2} \nabla\varphi_k \tag{3.55}$$

where $\widehat{p}^k$ is as in (3.6) and $\mathcal{S}$ is as in (3.4).

As a consequence of the above statement, it follows that *Algorithm 7* is a special instance of *Algorithm 6* for finding a point in the extended solution set $\mathcal{S}$ as in (3.4).

Proof. (i) Note first, from Lemma 3.3.1(a)-(c) and the inclusion in (3.45), that φ_k defined as in (3.48) is an affine function and $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$. Moreover, its gradient is given by (3.40). The fact that $\nabla\varphi_k \neq 0$, for all $k \geq 0$, follows from the assumption that Algorithm 7 does not stop at step (3.a) and Lemma 3.3.1(b).

(ii) The first equality in (3.55) follows from Remark (3.3.2)(iv). On the other hand, using definition of p^k in (3.55), $\widehat{p}^k$ (as we shown previously) and the update (3.50)-(3.51) in Algorithm 7, after some algebraic manipulations, we have

$$\begin{aligned} p^{k+1} - \widehat{p}^k &= (z^{k+1} - \widehat{z}^k, w_1^{k+1} - \widehat{w}_1^k, \dots, w_{n-1}^{k+1} - \widehat{w}_{n-1}^k) \\ &= -\beta_k\theta_k \left(\gamma^{-1} \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), x_1^k - G_1 x_n^k, \dots, x_{n-1}^k - G_{n-1} x_n^k \right) \\ &= -\beta_k\theta_k \nabla\varphi_k, \end{aligned}$$

in the last equality we used definition of $\nabla\varphi_k$ (see (3.40)), where θ_k is given as in (3.49). This establish the second equality in (3.55). To finalize the proof, the last statement of the proposition is a consequence of items (i) and (ii) as well as of Algorithm 6's definition. $\square$

3.3.2 Convergence of Algorithm 7

We now state a technical lemma which will be useful for the convergence analysis of Algorithm 7.

Lemma 3.3.6. Consider the sequences evolved by *Algorithm 7*, let $\widehat{p}^k = (\widehat{z}^k, \widehat{w}_i^k, \ldots, \widehat{w}_{n-1}^k)$ and $\widehat{w}_n^k$ is as in (3.42). Assume that, for $i = 1 \ldots, n$, the assumption (D4) hold, i.e.,

$$0 < \underline{\rho} \le \rho_i^k \le \overline{\rho} < +\infty, \qquad \forall k \ge 0. \tag{3.56}$$

Then the following hold:

(a) For all $k \ge 0$,

$$2\varphi_k(\widehat{p}^k) \ge \left(1 - \sigma^2\right) \min\{\overline{\rho}^{-1}, \overline{\rho}\} \sum_{i=1}^{n} \left(\|G_i \widehat{z}^k - x_i^k\|^2 + \|y_i^k - \widehat{w}_i^k\|^2\right). \tag{3.57}$$

(b) For all $k \ge 0$,

$$\|\nabla \varphi_k\|_\gamma^2 \le E \sum_{i=1}^{n} \left(\|G_i \widehat{z}^k - x_i^k\|^2 + \|y_i^k - \widehat{w}_i^k\|^2\right), \tag{3.58}$$

where

$$E = \max\left\{2 \max\{1, (n-1) \max_{1 \le i \le n-1} \{\|G_i\|^2\}\}, n\gamma^{-1} \left(\max_{1 \le i \le n-1} \|G_i^*\|^2\right)\right\} > 0. \tag{3.59}$$

(c) There exist $\xi > 0$ such that, for all $k \ge 0$,

$$\frac{\varphi_k(\widehat{p}^k)^2}{\xi \|\nabla \varphi_k\|_\gamma^2} \ge \varphi_k(\widehat{p}^k) \ge \xi \|\nabla \varphi_k\|_\gamma^2, \tag{3.60}$$

Proof. (a) From (3.45) and (3.46), using the basic identity $\|a + b\|^2 = \|a\|^2 + \|b\|^2 + 2\langle a, b\rangle$ with $a = x_i^k - G_i \widehat{z}^k$ and $b = y_i^k - \widehat{w}_i^k$, we have

$$\begin{aligned}
\sigma^2 \left(\|G_i \widehat{z}^k - x_i^k\|^2 + \|\rho_i^k (y_i^k - \widehat{w}_i^k)\|^2\right) &\ge \|e_i^k\|^2 \\
&= \|x_i^k - G_i \widehat{z}^k + \rho_i^k (y_i^k - \widehat{w}_i^k)\|^2 \\
&= \|x_i^k - G_i \widehat{z}^k\|^2 + \|\rho_i^k (y_i^k - \widehat{w}_i^k)\|^2 - 2\rho_i^k \langle G_i \widehat{z}^k - x_i^k, y_i^k - \widehat{w}_i^k\rangle,
\end{aligned}$$

which in turn, it is equivalent to

$$2\rho_i^k \langle G_i \widehat{z}^k - x_i^k, y_i^k - \widehat{w}_i^k\rangle \ge (1 - \sigma^2) \left(\|x_i^k - G_i \widehat{z}^k\|^2 + \|\rho_i^k (y_i^k - \widehat{w}_i^k)\|^2\right) > 0.$$

Note now that the desired result follows by dividing the latter inequality by ρ_i^k, summing it on i and taking into account assumption (3.56) and the characterization of φ_k in (3.39).

(b) Recall from (3.41) (Lemma 3.3.1 (c))and the fact $G_n = I$,

$$\|\nabla\varphi_k\|_\gamma^2 = \gamma^{-1}\|\sum_{i=1}^{n} G_i^* y_i^k\|^2 + \sum_{i=1}^{n-1}\|x_i^k - G_i x_n^k\|^2. \tag{3.61}$$

We star by writing the second term on the right side in the above equality as:

$$\begin{aligned}\sum_{i=1}^{n-1}\|x_i^k - G_i x_n^k\|^2 &= \sum_{i=1}^{n-1}\|x_i^k - G_i\widehat{z}^k + G_i(\widehat{z}^k - x_n^k)\|^2 \quad (3.62)\\ &\le 2\sum_{i=1}^{n-1}\left(\|x_i^k - G_i\widehat{z}^k\|^2 + \|G_i(\widehat{z}^k - x_n^k)\|^2\right)\\ &\le 2\sum_{i=1}^{n-1}\|x_i^k - G_i\widehat{z}^k\|^2 + 2(n-1)\max_{1\le i\le n-1}\|G_i\|^2\|\widehat{z}^k - x_n^k\|^2\\ &\le 2\max\left\{1, (n-1)\max_{1\le i\le n-1}\{\|G_i\|^2\}\right\}\sum_{i=1}^{n}\|G_i\widehat{z}^k - x_i^k)\|^2, \quad (3.63)\end{aligned}$$

where in the first inequality we used (A.8)(Lemma A.2.1 (ii)) and second inequality used the fact that for each $i = 1, \ldots, n-1$, G_i is linear and bounded. Similarly, for the first term in (3.61), using (A.8) (Lemma A.2.1), (3.53) and the assumption that $G_n = I$, we obtain

$$\begin{aligned}\left\|\sum_{i=1}^{n} G_i^* y_i^k\right\|^2 &= \left\|\sum_{i=1}^{n} G_i^*\left(y_i^k - w_i^k\right)\right\|^2\\ &\le n\left(\max_{1\le i\le n-1}\|G_i^*\|^2\right)\sum_{i=1}^{n}\|y_i^k - \widehat{w}_i^k\|^2. \quad (3.64)\end{aligned}$$

Setting E as in (3.59) and combining (3.62), (3.64) and (3.61) we get the inequality in (3.58).

(c) Note that from (3.57) and (3.58), we obtain the second inequality in (3.60) with

$$\xi = \frac{(1-\sigma^2)\min\{\overline{\rho}^{-1}, \overline{\rho}\}}{2E} > 0. \tag{3.65}$$

The first inequality in (3.60) is a direct consequence of the second one. This ends the proof of Lemma. □

Proposition 3.3.7. Consider the sequences evolved by *Algorithm 7* and let $\{w_n^k\}$ and $\{p^k\}$ be as in (3.53) and (3.54), respectively. Assume that

$$\sum_{k=0}^{\infty}\alpha_k\|p^k - p^{k-1}\|_\gamma^2 < +\infty \tag{3.66}$$

and assumption (D4) hold, i.e., for $i = 1\ldots, n$

$$0 < \underline{\rho} \le \rho_i^k \le \overline{\rho} < \infty \qquad \forall k \ \ge 0. \tag{3.67}$$

Then, the following hold

(a) We have, $\varphi_k(\widehat{p}^k) \to 0$ and $\|\nabla\varphi_k\|_\gamma \to 0$ as $k \to +\infty$.

(b) We have, $\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \to 0$ and for each $i = 1, \dots, n-1$, $x_i^k - x_n^k \to 0$, as $k \to +\infty$.

(c) For each $i = 1, \dots, n$, we have $\|G_i z^k - x_i^k\| \to 0$ and $\|y_i^k - w_i^k\| \to 0$ as $k \to \infty$.

(d) Every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$, where $\mathcal{S}$ is as in (3.4).

Proof. (a) Using the last statement in Proposition 3.3.5 (that Algorithm 7 is a particular instance of Algorithm 6), Proposition 3.2.3(c) and the fact that $\varphi_k(\widehat{p}^k) \geq 0$ by (3.57), we obtain

$$\frac{\varphi_k(\widehat{p}^k)}{\|\nabla\varphi_k\|_\gamma} \to 0 \quad \text{as} \quad k \to +\infty,$$

which after taking limit in (3.60) (Lemma 3.3.6(c)) gives the desired result in item (a).

(b) This follows directly from the second limit in item (a) combined with (3.41) (and the fact that $G_n = I$).

(c) Note first that, (3.57) (Lemma 3.3.6(a)) and first limit in (a) yields that

$$\|y_i^k - \widehat{w}_i^k\| \to 0 \quad \text{and} \quad \|G_i\widehat{z}^k - x_i^k\| \to 0 \qquad \text{as} \quad k \to \infty \quad (i = 1, \dots, n). \tag{3.68}$$

Using the triangle inequality, the identity (3.42), (3.54) and definition of $\|.\|_\gamma$, we find

$$\begin{aligned}
\|G_i z^k - x_i^k\| &\leq \|z^k - \widehat{z}^k\|\|G_i\| + \|G_i\widehat{z}^k - x_i^k\| \\
&= \alpha_k\|z^k - z^{k-1}\|\|G_i\| + \|G_i\widehat{z}^k - x_i^k\| \\
&\leq \gamma^{-1}\sqrt{\alpha_k}\,\|p^k - p^{k-1}\|_\gamma\|G_i\| + \|G_i\widehat{z}^k - x_i^k\|, \quad i = 1, \dots, n,
\end{aligned} \tag{3.69}$$

where we also used the fact that, since $\alpha_k \in [0,1)$, then $\alpha_k \leq \sqrt{\alpha_k}$. Using a similar reasoning, we also find

$$\|y_i^k - w_i^k\| \leq \sqrt{\alpha_k}\,\|p^k - p^{k-1}\|_\gamma + \|y_i^k - \widehat{w}_i^k\|, \qquad i = 1, \dots, n-1. \tag{3.70}$$

Note also that, using (3.43), (3.105) (3.53), Lemma A.2.1 (ii), the fact that $\alpha_k^2 \leq \alpha_k$ and definition of $\|.\|_\gamma$, we obtain

$$\begin{aligned}
\frac{1}{2}\|y_n^k - w_n^k\|^2 &\leq \|\widehat{w}_n^k - w_n^k\|^2 + \|y_n^k - \widehat{w}_n^k\|^2 \\
&\leq (n-1)\max_{i=1,\dots,n-1}\{\|G_i^*\|^2\}\left(\sum_{i=1}^{n-1}\|\widehat{w}_i^k - w_i^k\|^2\right) + \|y_n^k - \widehat{w}_n^k\|^2 \\
&= (n-1)\max_{i=1,\dots,n-1}\{\|G_i^*\|^2\}\left(\sum_{i=1}^{n-1}\alpha_k^2\|w_i^{k-1} - w_i^k\|^2\right) + \|y_n^k - \widehat{w}_n^k\|^2 \\
&\leq (n-1)\max_{i=1,\dots,n-1}\{\|G_i^*\|^2\}\left(\sum_{i=1}^{n-1}\alpha_k\|w_i^{k-1} - w_i^k\|^2\right) + \|y_n^k - \widehat{w}_n^k\|^2 \\
&\leq (n-1)\max_{i=1,\dots,n-1}\{\|G_i^*\|^2\}\alpha_k\|p^k - p^{k-1}\|_\gamma^2 + \|y_n^k - \widehat{w}_n^k\|^2,
\end{aligned} \tag{3.71}$$

To finish the proof of (c), combine (3.69)–(3.71) with (3.68) and (3.66).

(d) Let $p^\infty := (z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty) \in \boldsymbol{\mathcal{H}}$ be a weak cluster point of $\{p^k\}$ (since that $\{p^k\}$ is bounded by Proposition 3.3.5 and Proposition 3.2.3(a)) and let $\{p^{k_j}\}$ be a subsequence of $\{p^k\}$ such that $p^{k_j} \rightharpoonup p^\infty$, i.e.,

$$z^{k_j} \rightharpoonup z^\infty \quad \text{and} \quad w_i^{k_j} \rightharpoonup w_i^\infty, \quad i = 1, \ldots, n-1. \tag{3.72}$$

Using (c), (3.72) and the fact that $G_n = I$(see Assumption (D2)), we obtain

$$x_n^{k_j} \rightharpoonup z^\infty \quad \text{and} \quad y_i^{k_j} \rightharpoonup w_i^\infty, \quad i = 1, \ldots, n-1. \tag{3.73}$$

For the other hand, item (b) applied to subsequence, yields

$$\sum_{i=1}^{n-1} G_i^* y_i^{k_j} + y_n^{k_j} \to 0 \quad \text{and} \quad x_i^{k_j} - G_i x_n^{k_j} \to 0 \qquad (i = 1, \ldots, n-1). \tag{3.74}$$

Now, we define the operators $A : \mathcal{H}_0 \rightrightarrows \mathcal{H}_0$, $B : \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1} \rightrightarrows \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ and $G : \mathcal{H}_0 \rightrightarrows \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ as

$$\begin{aligned} A :& z \mapsto T_n(z) \\ B :& (w_1, \ldots, w_{n-1}) \mapsto T_1(w_1) \times \cdots \times T_{n-1}(w_{n-1}) \\ G :& z \mapsto (G_1 z, \ldots, G_{n-1} z). \end{aligned} \tag{3.75}$$

Since $\{T_i\}_{i=1}^n$ are maximal monotone operators, A and B are maximal monotone operators [20, Proposition 20.27] and G is linear and bounded operator due to G_i are linear and bounded for all $i = 1, \ldots, n-1$.

Using the above definitions of A and B and the inclusions in (3.45), we have

$$a^j \in A(r^j) \quad \text{and} \quad b^j \in B(s^j), \tag{3.76}$$

where

$$r^j := x_n^{k_j}, \quad a^j := y_n^{k_j}, \quad b^j := (y_1^{k_j}, \ldots, y_{n-1}^{k_j}) \quad \text{and} \quad s^j := (x_1^{k_j}, \ldots, x_{n-1}^{k_j}). \tag{3.77}$$

Moreover, (3.77) and (3.73) yield

$$r^j \to r^\infty \quad \text{and} \quad b^j \to b^\infty \quad \text{as } j \to \infty, \tag{3.78}$$

where,

$$r^\infty := z^\infty \quad \text{and} \quad b^\infty := (w_1^\infty, \ldots, w_{n-1}^\infty) \tag{3.79}$$

Note now that using (3.77), the fact that $G^*(w_1, \ldots, w_{n-1}) = \sum_{i=1}^{n-1} G_i^* w_i$, for all $(w_1, \ldots, w_{n-1}) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$, the fact that $G_n = I$ and the first limit in (3.74), we find

$$a^j + G^* b^j = \sum_{i=1}^{n-1} G_i^* y_i^{k_j} + y_n^{k_j} \to 0 \quad \text{as} \quad j \to \infty. \tag{3.80}$$

Using now the second limit in (3.74) combined with (3.77) and the definition of G in (3.75), we obtain

$$G r^j - s^j = \left(G_1 x_n^{k_j} - x_1^{k_j}, \ldots, G_{n-1} x_n^{k_j} - x_{n-1}^{k_j} \right) \to 0 \quad \text{as} \quad j \to \infty. \tag{3.81}$$

Using Lemma A.2.3 combined with (3.76), (3.78) (3.80) and (3.81) we conclude that

$$-G^*b^\infty \in A(r^\infty) \quad \text{and} \quad b^\infty \in B(Gr^\infty),$$

which, in turn, combined with (3.75) and (3.79) implies that

$$w_i^\infty \in T_i(G_i z^\infty),\ i = 1, \dots, n-1, \quad -\sum_{i=1}^{n-1} G_i^* w_i^\infty \in T_n(z^\infty)$$

from which we conclude that $p^\infty = (z^\infty, w_1^\infty, \dots, w_{n-1}^\infty) \in \mathcal{S}$. Since p^∞ was chosen arbitrarily, then every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$.

□

The following is the first result on the asymptotic convergence of Algorithm 7.

Theorem 3.3.8 (First result on the convergence of *Algorithm* 7)**.** Consider the sequences evolved by *Algorithm 7* and let $\{w_n^k\}$ and $\{p^k\}$ be as in (3.53) and (3.54), respectively. Assume that conditions (3.66) and (3.67) of Proposition 3.3.7 hold, i.e., assume that

$$\sum_{k=0}^{\infty} \alpha_k \|p^k - p^{k-1}\|_\gamma^2 < +\infty \tag{3.82}$$

and, for $i = 1 \dots, n$

$$0 < \underline{\rho} \leq \rho_i^k \leq \overline{\rho} < \infty \qquad \forall\ k \geq 0. \tag{3.83}$$

Then, there exist $(z^\infty, w_1^\infty, \dots, w_{n-1}^\infty) \in \mathcal{S}$ such that $z^k \rightharpoonup z^\infty$ and, $w_i^k \rightharpoonup w_i^\infty$, for $i = 1, \dots, n-1$ as $k \to \infty$. Furthermore, $x_i^k \rightharpoonup G_i z^\infty$ and $y_i^k \rightharpoonup w_i^\infty$, for $i = 1, \dots, n$.

Proof. Since Algorithm 7 is a particular instance of Algorithm 6 (see Proposition 3.3.5), then in view of Proposition 3.3.7(d) and Proposition 3.2.3(b) we conclude that $\{p^k = (z^k, w_1^k, \dots, w_{k-1}^k)\}$ converges weakly to some point in $\mathcal{S}$ as in (3.4), i.e., there exist $(z^\infty, w_1^\infty, \dots, w_{n-1}^\infty) \in \mathcal{S}$, such that, $z^k \rightharpoonup z^\infty$ and $w_i^k \rightharpoonup w_i^\infty$, for $i = 1, \dots, n-1$, which in turn combined with Proposition (3.3.7)(c) implies that $x_i^k \rightharpoonup G_i z^\infty$ and $y_i^k \rightharpoonup w_i^\infty$, for $i = 1, \dots, n$. In particular, z^∞ solve (3.36). □

Next theorem shows the convergence of Algorithm 7 under certain assumptions on α, β_k and the sequence $\{\alpha_k\}$ (see the remarks below).

Theorem 3.3.9 ((Second result on the convergence of Algorithm 7)**.** Consider the sequences evolved by *Algorithm 7* and assume moreover that $\alpha \in [0, 1)$, $\overline{\beta} \in (0, 2)$ and $\{\alpha_k\}$ satisfy (for some $\overline{\alpha} > 0$) the conditions (3.22) and (3.23) of *Proposition 3.2.4*, i.e.,

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \overline{\alpha} < 1 \qquad \forall k \geq 0 \tag{3.84}$$

and

$$\overline{\beta} = \overline{\beta}(\overline{\alpha}) := \frac{2(\overline{\alpha} - 1)^2}{2(\overline{\alpha} - 1)^2 + 3\overline{\alpha} - 1}. \tag{3.85}$$

Assume also that condition (3.83) hold, i.e., for $i = 1, \ldots, n$,

$$0 < \underline{\rho} \leq \rho_i^k \leq \overline{\rho} < \infty \qquad \forall\, k \geq 0. \tag{3.86}$$

Then, the same conclusions of Theorem 3.3.8 hold, i.e., there exist $p^\infty := (z^\infty, w_1^\infty, \ldots, w_{n-1}^\infty) \in \mathcal{S}$ such that $z^k \rightharpoonup z^\infty$ and for $i = 1, \ldots, n-1$, $w_i^k \rightharpoonup w_i^\infty$ (weakly) as $k \to \infty$. Furthermore, $x_i^k \rightharpoonup G_i z^\infty$, $y_i^k \rightharpoonup w_i^\infty$, for $i = 1, \ldots, n$.

Proof. In view of Propositions 3.3.5 and 3.3.7(d) and Proposition 3.2.4(b) one concludes that that $\{p^k\}$ converges weakly to some $p^\infty := (z^\infty, w_1^\infty, \ldots, w_{n-1}^\infty)$ in $\mathcal{S}$ as in (3.4). The rest of the proof follows the same argument used in Theorem 3.3.8's proof. □

In view of Remark 3.3.2 (i) and Lemma 3.3.1(c) if $\nabla\varphi_k = 0$, then the algorithm stops and finds a solution for the monotone inclusion problem (3.36). Therefore, we can consider the following stopping criterion for algorithm 7: Given arbitrary scalars $\epsilon, \delta > 0$, Algorithm 7 stops when it finds points $(x_i, y_i) \in \mathcal{H}_i \times \mathcal{H}_i$, for $i = 1, \ldots, n$, such that

$$\begin{aligned} & y_i \in T_i(x_i), \qquad i = 1, \ldots, n, \\ & \|\sum_{i=1}^n G^* y_i\| \leq \epsilon, \\ & \|x_i - G_i^* x_n\| \leq \delta, \qquad i = 1, \ldots, n-1. \end{aligned} \tag{3.87}$$

We observe that if $\epsilon = \delta = 0$, in view of Lemma 3.3.1(d) and the above criterion we have $x_i = G_i x_n$, for $i = 1, \ldots, n-1$ and $(x_n, y_1, \ldots, y_{n-1}) \in \mathcal{S}$ and in particular $z = x_n$ solves (3.36).

Next we present a result on nonasymptotic global convergence rates/iteration-complexity for Algorithm 7

Theorem 3.3.10 (Nonasymptotic convergence rate of *Algorithm* 7)**.** Consider the sequences evolved by *Algorithm 7* and assume that $\alpha \in [0, 1)$, $\overline{\beta} \in (0, 2)$ and $\{\alpha_k\}$ satisfies (for some $\overline{\alpha} > 0$) the conditions (3.84)-(3.86) of Theorem 3.3.9. Let d_0 be the distance of $p^0 = (z^0, w_1^0, \ldots, w_{n-1}^0)$ to the set $\mathcal{S}$ defined in (3.4), then for all $k \geq 0$, we have

$$y_i^k \in T_i(x_i^k) \qquad \forall\, i = 1, \ldots, n,$$

and there exists $j \in \{0, \ldots, k\}$ such that

$$\begin{aligned} & \|\textstyle\sum_{i=1}^n G_i^* y_i^j\| \leq \frac{d_0}{\xi\sqrt{k}} \sqrt{\gamma \underline{\beta}^{-1} \left(2 - \overline{\beta}\right)^{-1} \left(1 + \frac{\alpha(1+2\alpha-\alpha^3)}{(1-\alpha)^2 q(\alpha)}\right)}, \\ & \|x_i^j - G_i^* x_n^j\| \leq \frac{d_0}{\xi\sqrt{k}} \sqrt{\underline{\beta}^{-1} \left(2 - \overline{\beta}\right)^{-1} \left(1 + \frac{\alpha(1+2\alpha-\alpha^3)}{(1-\alpha)^2 q(\alpha)}\right)} \quad i = 1, \ldots, n-1 \end{aligned} \tag{3.88}$$

where $\xi > 0$ is as in (3.65) and

$$q(\alpha) := 2\left(\overline{\beta}^{-1} - 1\right)\alpha^2 - \left(4\overline{\beta}^{-1} - 1\right)\alpha + 2\overline{\beta}^{-1} - 1, \quad \alpha \in \mathbb{R}. \tag{3.89}$$

Proof. Since Algorithm 7 is a particular instance of Algorithm 6, by Proposition 3.3.5 and $\varphi_k(\widehat{p}^k) \geq 0$ by Lemma 3.3.6(a)), the equality in (3.8) (see Remark 3.2.1(i)) combined with (3.60) in Lemma 3.3.6(c) yields

$$\|\widetilde{p}^{k+1} - \widehat{p}^k\|_\gamma^2 = \frac{\varphi_k(\widehat{p}^k)^2}{\|\nabla\varphi_k\|_\gamma^2} \geq \xi^2 \|\nabla\varphi_k\|_\gamma^2.$$

Hence, summing on k in the above inequality an then using (3.31) (in Proposition 3.2.5) we obtain

$$\xi^2 \sum_{j=0}^{k} \|\nabla\varphi_j\|_\gamma^2 \leq \sum_{j=0}^{k} \|\widetilde{p}^{j+1} - \widehat{p}^j\|_\gamma^2 \leq \underline{\beta}^{-1}\left(2 - \overline{\beta}\right)^{-1} \left(1 + \frac{\alpha(1 + 2\alpha - \alpha^3)}{(1-\alpha)^2 q(\alpha)}\right) \|p^0 - p\|_\gamma^2. \quad (3.90)$$

The desired conclusion following from the last inequality and Lemma 3.3.1(c)-(3.40). □

Remark 3.3.11. (i) Theorem 3.3.10 provides a global $\mathcal{O}(1/\sqrt{k})$ *pointwise* convergence rate and ensures, in particular, that for given tolerances $\epsilon, \delta > 0$, Algorithm 7 finds the points (x_i, y_i) in $\mathcal{H}_i^2$ $(i = 1, \ldots, n)$ satisfying (3.87) after performing at most

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{d_0^2}{\epsilon^2} \right\rceil, \left\lceil \frac{d_0^2}{\delta^2} \right\rceil\right\}\right)$$

iterations.

(ii) A similar global $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate for projective splitting (without inertial case) algorithm also was obtained in [79] with for $n = 2$ and $G_i = I$ (see Theorem 4.2 in the latter reference).

The special case $n = 1$

We consider the special case $n = 1$. In this case, we have by assumption that $G_1 = I$, $w_1^k = 0$ for all $k \geq 0$, and we are solving the problem

$$0 \in T(z). \quad (3.91)$$

Then, the affine function defined in (3.48) becomes

$$\varphi_k(z) = \langle z - x^k, y^k \rangle,$$

where we dropped the unnecessary index, by writing $x_1^k = x^k$ and $y_1^k = y^k$. Then the update carried out by the algorithm is as follows: choose $\alpha_k \in [0, \alpha]$ such that

$$\widehat{z}^k = z^k + \alpha_k(z^k - z^{k-1}). \quad (3.92)$$

Next choose $\rho^k > 0$ and find $(x^k, y^k) \in G(T)$ such that

$$\rho^k y^k + x^k = \widehat{z}^k + e^k \quad (3.93)$$

and

$$\|\rho^k y^k + x^k - \widehat{z}^k\|^2 \leq \sigma^2(\|x^k - \widehat{z}^k\|^2 + \|\rho^k y^k\|^2), \quad (3.94)$$

and update

$$z^{k+1} = \hat{z}^k - \frac{\beta_k \varphi_k}{\|\nabla \varphi_k\|^2} \nabla \varphi_k. \tag{3.95}$$

Summarizing, when $n = 1$, Algorithm 7 reduces to the hybrid projection proximal point method of Solodov and Svaiter [118, Algorithm 1.1] with inertial effects. It is worth mentioning that relative-error condition considered in ([118] eq. (1.7)) is less restrictive than relative-error condition (3.94) considered here. See also [4] for an another development of inertial hybrid projection-proximal point method (see iteration $(\mathcal{A}_0^\rho)$–$(\mathcal{A}_2^\rho)$).

Observe that, $\sigma = 0$ implies $e^k = 0$, see (3.94). This together with (3.93) yields $\varphi_k(\hat{z}^k) = \rho^k \|y^k\|^2$. Furthermore, $\nabla_z \varphi_k = \gamma^{-1} y^k$ and so, $\|\nabla_z \varphi_k\|^2 = \gamma.\gamma^{-2}\|y^k\|^2 = \gamma^{-1}\|y^k\|^2$. Hence, using (3.95), (3.93) and the last equality, we have for all $k \geq 0$,

$$\begin{aligned} z^{k+1} &= \hat{z}^k - \beta_k \rho^k y^k \\ &= (1 - \beta_k)\hat{z}^k + \beta_k x^k \\ &= (1 - \beta_k)\hat{z}^k + \beta_k J_{\rho^k T}(\hat{z}^k). \end{aligned}$$

Thus, when $n = 1$ and $\sigma = 0$, the inertial splitting projective method proposed in this chapter reduces to the relaxed inertial proximal point method analyzed in [11]. Hence, the Algorithm 7 proposed in this chapter extends the inertial and relaxed projection method for solving single operator monotone inclusion to more general structured and composed inclusion problem.

To end this chapter, in the next section we will develop a specific instance of the splitting-projective framework of Section 3.3.

3.4 An inertial algorithm resembling the multi-block ADMM method

Throughout this section, we will assume that for each $i = 1, \ldots, n$, $\mathcal{H}_i \equiv \mathbb{R}^{n_i}$. We now consider the convex optimization problem of the form

$$\min \sum_{i=1}^{n} f_i(u_i) \quad s.t. \quad \sum_{i=1}^{n} M_i u_i = b, \tag{3.96}$$

where $n \geq 2$, $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lower semicontinuos and M_i is a $m \times n_i$ matrix for $i = 1, \ldots, n$, and $b \in \mathbb{R}^m$. For the case $n = 2$ the well-know decomposition method for such problems is the *Alternating Direction Method of Multipliers* (ADMM), which goes back to the work of Glowinski and Marroco [58] and of Gabay and Mercier [57]. ADMM can also be viewed as an instance of the *Douglas-Rachford splitting method* applied to the dual problem of (3.96) as was shown by Gabay in [56](see also [49]).

The dual problem associate to the primal problem (3.96) is obtained by applying the Fenchel duality and it is given by

$$\min_{z \in \mathbb{R}^m} \left(\sum_{i=1}^{n} f_i^*(-M_i^T z) \right) + \langle b, z \rangle, \tag{3.97}$$

where f_i^* denotes the convex conjugate of f_i. By choosing any $b_1, \ldots, b_n \in \mathbb{R}^m$ such that $\sum_{i=1}^n b_i = b$, the dual problem (3.97) may also be written as

$$\min_{z\in\mathbb{R}^m} \sum_{i=1}^{n} \left(f_i^*(-M_i^T z) + \langle b_i, z\rangle \right). \tag{3.98}$$

In the last years, there has also been considerable interest in a class of methods similar to ADMM for case $n > 2$, although the direct generalization of the two block ADMM to $n > 2$ does not converges in general convex case [37]. Some works related to such subject can be found for instance in [47, 48, 63, 127] and references therein.

The main goal of this section is to propose an inertial algorithm resembling the ADMM that can accommodate any number of blocks in a parallel environment to solve a class of linearly constrained optimization problem, which we will refer to as inertial multi-block ADMM-like method (Algorithm 8). Specifically, based in the Eckstein's work [48] we propose an algorithmic scheme for solving (3.96) by applying the Algorithm 7 to its dual problem (3.98), similar to what Gabay [56] developed.

Let us make the following standard assumption:

($\mathcal{B}1$) The Problem (3.96) possesses at least one Karush-Kuhn-Tucker (KKT) point, i.e., there exist $z^\infty \in \mathbb{R}^m$ and $u_i^\infty \in \mathbb{R}^{n_i}$, for $i = 1, \ldots, n$ such that

$$\sum_{i=1}^{n} M_i u_i^\infty = b \qquad \text{and} \quad -M_i^T z^\infty \in \partial f_i(u_i^\infty) \quad (\forall i = 1, \ldots, n). \tag{3.99}$$

($\mathcal{B}2$) $\partial[f_i^*(-M_i)] = -M_i \partial f_i^*(-M_i^T)$, for all $i = 1, \ldots, n$.

Remark 3.4.1. (a) The Assumption ($\mathcal{B}1$) guarantees that the optimal values of (3.96) and (3.98) are equal.

(b) In the finite-dimensional setting, a sufficient condition for Assumption ($\mathcal{B}2$) to hold is

$$\operatorname{range}\operatorname{dom}\partial f_i^* \cap \operatorname{range}(-M_i^T) \neq \emptyset \quad \text{or equivalently} \quad ri \ \operatorname{range}\partial f_i \cap (ker\ M_i)^\perp \neq \emptyset,$$

see, [109, theorem 23.9].

By defining

$$(\forall i = 1, \cdots, n) \qquad h_i := f_i^*(-M_i^T) + \langle \cdot, b_i\rangle, \tag{3.100}$$

the dual problem (3.98) may be expressed as

$$\min_{z\in\mathbb{R}^m} \sum_{i=1}^{n} h_i(z). \tag{3.101}$$

Since f_i^* is necessarily convex and lower semicontinuos function, $f_i^*(-M_i^T)$ is also convex and lower semicontinuos, so $\partial\left[f_i^* \circ (-M_i^T)\right]$ must be a maximal monotne operator. Therefore, in view of the assumption ($\mathcal{B}2$), the operators(necessary monotone) $-M_i\partial f_i^*(-M_i^T)$ are also maximal, and consequently

$$T_i := \partial h_i = -M_i\partial f_i^*(-M_i^T) + b_i, \qquad i = 1, \dots, n \tag{3.102}$$

are maximal monotone operators.

Now, we are ready to solve (3.101) by applying Algorithm 7 with $\mathcal{H}_0 = \mathcal{H}_i = \mathbb{R}^m$, $T_i = \partial h_i$ and $G_i = I$ for $i = 1, \dots, n$. For this purpose, we recall the extended solution (or kuhn-Tucker set)

$$\mathcal{S} = \left\{ (z, w_1, \dots, w_{n-1}) \in (\mathbb{R}^m)^n \mid w_i \in T_i(z), i = 1, \dots, n-1, -\sum_{i=1}^{n-1} w_i \in T_n(z) \right\}. \tag{3.103}$$

It is clear that (3.99) and (3.103) are equivalents, so it is natural to think that the problem (3.96) can be reformulate as an particular case of Algorithm 7.

The next result, proved in [80, Lemma 3.2], shows how we can invert operators $\rho\partial h_i + I$ (for $i = 1, \dots, n$).

Lemma 3.4.2. Let $b \in \mathbb{R}^m$, $f : \mathbb{R}^m \to (-\infty, +\infty]$ a proper lower semicontinuos and convex function and $M : \mathbb{R}^m \to \mathbb{R}^n$ a linear and bounded operator such that the assumption ($\mathcal{B}2$) holds. Consider any $z \in \mathbb{R}^m$ and $\rho > 0$. If $\widehat{u}$ is a solution to the problem

$$\min_{u \in \mathbb{R}^n} \left\{ f(u) + \langle z, Mu - b \rangle + \frac{\rho}{2}\|Mu - b\|^2 \right\}. \tag{3.104}$$

Then

$$b - M\widehat{u} \in \partial h(\tilde{z}),$$

where $h := f^*(-M^T) + \langle \cdot, b \rangle$ and $\tilde{z} = z + \rho(M\widehat{u} - b)$, and as a consequence $\tilde{z} = (\rho\partial h + I)^{-1}(z)$. Furthermore, the set of optimal solutions of (3.104) is nonempty.

In what follows we suppose that assumptions ($\mathcal{B}1$) and ($\mathcal{B}2$) holds.

Algorithm 8. An inertial Multi-Block ADMM-Like Method (IM-ADMM)

(0) Let $(z^{-1}, w_1^{-1}, \dots, w_{n-1}^{-1}) = (z^0, w_1^0, \dots, w_{n-1}^0) \in (\mathbb{R}^m)^n$, $u_i^0 \in \mathbb{R}^{n_i}, i = 1, \dots, n$, $\alpha \in [0, 1)$, $0 < \underline{\beta} \le \overline{\beta} < 2$ and $\gamma > 0$ be given and let $k \leftarrow 0$.

(1) Choose $\alpha_k \in [0, \alpha]$ and let

$$\begin{aligned} \widehat{z}^k &= z^k + \alpha_k(z^k - z^{k-1}), \\ \widehat{w}_i^k &= w_i^k + \alpha_k(w_i^k - w_i^{k-1}), \quad i = 1, \dots, n-1, \\ \widehat{w}_n^k &= -\sum_{i=1}^{n-1} \widehat{w}_i^k. \end{aligned} \tag{3.105}$$

(2) Choose scalars $\rho_i^k > 0$, for $i = 1, \dots, n$ find $u_i^k \in \mathbb{R}^{n_i}$ and $x_i^k, y_i^k \in \mathbb{R}^m$ as

$$\begin{cases} u_i^k = \operatorname{argmin}_{u_i \in \mathbb{R}^{n_i}} \{ f_i(u_i) + \langle M_i u_i, \widehat{z}^k \rangle + \frac{\rho_i^k}{2} \| M_i u_i - b_i + \widehat{w}_i^k \|^2 \}, \\ y_i^k = b_i - M_i u_i^k, \\ x_i^k = \widehat{z}^k + \rho_i^k (\widehat{w}_i^k - y_i^k). \end{cases} \tag{3.106}$$

(3) **(3.a)** If $\| \sum_{i=1}^n y_i^k \| + \sum_{i=1}^{n-1} \| x_i^k - x_n^k \| = 0$, STOP and set

$$z^{k+1} = x_n^k \quad \text{and} \quad w_i^{k+1} = y_i^k, \quad i = 1, \dots, n-1. \tag{3.107}$$

(3.b) Else, define

$$\theta_k := \frac{\max\{0, \sum_{i=1}^n \langle \widehat{z}^k - x_i^k, y_i^k - \widehat{w}_i^k \rangle\}}{\gamma^{-1} \| \sum_{i=1}^n y_i^k \|^2 + \sum_{i=1}^{n-1} \| x_i^k - x_n^k \|^2}. \tag{3.108}$$

(4) Choose some relaxation parameter $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and define

$$\begin{aligned} z^{k+1} &= \tilde{z}^k - \gamma^{-1} \beta_k \theta_k \left(\textstyle\sum_{i=1}^n y_i^k \right), \\ w_i^{k+1} &= \widehat{w}_i^k - \beta_k \theta_k \left(x_i^k - x_n^k \right), \quad i = 1, \dots, n-1. \end{aligned} \tag{3.109}$$

Our first result is to show that Algorithm 8 is in fact a particular instance of Algorithm 7.

Proposition 3.4.3. *Algorithm* 8 *is an especial instance of Algorithm* 7 *with* $\sigma = 0$, *for* $T_i = M_i \partial f_i^*(-M_i^T) + b_i$ *and* $G_i = I$, *with*

$$(\forall i = 1, \dots, n) \qquad x_i^k = \widehat{z}^k + \rho_i^k \left(M_i u_i^k - (b_i - \widehat{w}_i^k) \right) \quad \text{and} \quad y_i^k = b_i - M_i u_i^k \tag{3.110}$$

Proof. Both inertial and projection steps (steps (1) and (4)) of Algorithms 8 and 7 are identical, with the only exception that $G_i = I$ for every $i = 1, \dots, n$. So, it only remains to show that $y_i^k \in T_i(x_i^k)$ and $x_i^k + \rho_i^k y_i^k = \widehat{z}^k + \rho_i^k \widehat{w}_i^k$, with x_i^k and y_i^k are defined as in (3.110). In fact, note that the optimality condition of the minimization problem in (3.106) implies that for each $k \ge 0$,

$$0 \in \partial f_i(u_i^k) + M_i^T \widehat{z}^k + \rho_i^k M_i^T (M_i u_i^k - (b_i - \widehat{w}_i^k))$$

or,

$$0 \in \partial f_i(u_i^k) + M_i^T(\widehat{z}^k + \rho_i^k \widehat{w}_i^k) + \rho_i^k M_i^T(M_i u_i^k - b_i),$$

which is equivalent to,

$$u_i^k \in \operatorname*{argmin}_{u \in \mathbb{R}^m} \left\{ f_i(u) + \langle \widehat{z}^k + \rho_i^k \widehat{w}_i^k, M_i u - b_i \rangle + \frac{\rho_i^k}{2} \|M_i u - b_i\|^2 \right\},$$

i.e., u_i^k is solution to the minimization problem

$$\min_{u \in \mathbb{R}^m} \left\{ f_i(u) + \langle \widehat{z}^k + \rho_i^k \widehat{w}_i^k, M_i u - b_i \rangle + \frac{\rho_i^k}{2} \|M_i u - b_i\|^2 \right\}. \tag{3.111}$$

Hence, from (3.111), (3.104) and applying Lemma 3.4.2 with $\rho = \rho_i^k$, $f = f_i$, $h = h_i$, $M = M_i$, $b = b_i$, $\widehat{u} = u_i^k$ and $z = \widehat{z}^k + \rho_i^k \widehat{w}_i^k$ we conclude that x_i^k and y_i^k defined as in (3.110) satisfies

$$y_i^k = b_i - M_i u_i^k \in \partial h_i(\widehat{z}^k + \rho_i^k \widehat{w}_i^k + \rho_i^k(M_i u_i^k - b_i)) = T_i(x_i^k).$$

Finally, as direct consequence of (3.110) one gets

$$x_i^k + \rho_i^k y_i^k = \widehat{z}^k + \rho_i^k \widehat{w}_i^k.$$

This completes the proof.

□

Remark 3.4.4. We now present some remarks regarding Algorithm 8:

(i) If Algorithm 8 stop at step (3.b), we have

$$b - \sum_{i=1}^{n} M_i u_i^k = \sum_{i=1}^{n} y_i^k = 0 \qquad \text{and} \qquad x_i^k - x_n^k = 0, \quad \forall i = 1, \ldots, n-1. \tag{3.112}$$

Furthermore, from definitions of x_i^k and y_i^k, and optimality condition in (3.106), we have for all $k \geq 0$

$$0 \in M_i^T x_i^k + \partial f_i(u_i^k) \qquad \forall i = 1, \ldots, n. \tag{3.113}$$

Hence, from (3.112) and (3.113) we conclude that, if Algorithm 8 stop at step (3.a), then $(x_n^k, u_1^k, \ldots, u_n^k)$ satisfies the KKT conditions (3.99), and consequently $(u_1^k, \ldots, u_n^k)$ is a solution of the primal problem (3.96) and x_n^k is a solution of dual problem (3.98).

(ii) Because Algorithm 8 is based on a iterative projective splitting framework of Algorithm 7, the penalty parameter ρ_i^k need not be fixed with respect to either the subsystem i or the iteration k, but only bounded, see (3.116) below.

(iii) Since $y_i^k = b_i - M_i u_i^k$ for all $k \geq 0$, and for every $i = 1, \ldots, n$, then the update to the (dual) solution (with respect to problem (3.96)) can be rewritten

$$z^{k+1} = \tilde{z}^k - \gamma^{-1}\beta_k\theta_k \left(\sum_{i=1}^{n} y_i^k \right) = \widehat{z}^k + \gamma^{-1}\theta_k\beta_k \left(\sum_{i=1}^{n} M_i u_i^k - b \right)$$

which is the standard multiplier update of the augmented Lagrangian method and ADMM, but with $\tau_k := \gamma^{-1}\theta_k\beta_k$ playing the role of the stepszise. The role of w_i^k and the associate step directions $x_i^k - x_n^k$ has a less familiar character as was pointed in [48].

(iv) If $\alpha_k = 0$, Algorithm 8 reduces to a variant of the n-block ADMM-like Algorithm in [48], (synchronous and without block-iterate feature, see Algorithms 6 and 7 in such a reference). Also, it is worth mentioning that the convergence analysis for the proposed algorithm is based on the results of convergence in Section 3.3, by viewing them within the framework of Algorithm 7.

Next, we state a convergence result for Algorithm 8, which follows essentially from the analysis in section 3.3.

Theorem 3.4.5 (convergence result)**.** Consider $\{u_1^k\} \dots, \{u_n^k\}$, $\{z^k\}, \{w_1^k\}, \dots, \{w_n^k\}$ sequences generated by *Algorithm*8. Suppose that Assumptions ($\mathcal{B}1$) and ($\mathcal{B}2$) hold and let $\alpha \in [0,1)$, $\overline{\beta} \in (0,2)$, $\{\alpha_k\}$ and $\{\rho_k\}$ satisfying the conditions (3.84)–(3.86) of Theorem 3.3.9, i.e.,

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \overline{\alpha} < 1 \qquad \forall k \geq 0, \tag{3.114}$$

$$\overline{\beta} = \overline{\beta}(\overline{\alpha}) := \frac{2(\overline{\alpha}-1)^2}{2(\overline{\alpha}-1)^2 + 3\overline{\alpha} - 1}, \tag{3.115}$$

and

$$0 < \underline{\rho} \leq \rho_i^k \leq \overline{\rho} < \infty \qquad \forall k \geq 0. \tag{3.116}$$

Then,

(a) The sequences $\{z^k\}, \{x_1^k\}, \dots, \{x_n^k\}$ converge all to the same limit z^*, which is the optimal solution of the dual problem (3.98) of the problem (3.96).

(b) The primal sequences $\{u_i^k\}_{i=1}^n$ generate by *Algorithm* 8 are asymptotically feasible for (3.96), in the sense that

$$\lim_{k\to\infty} \sum_{i=1}^n M_i u_i^k - b = 0.$$

(c) The primal sequences $\{u_i^k\}_{i=1}^n$ generate by *Algorithm* 8 are optimal for (3.96), i.e.,

$$\lim_{k\to\infty} \sum_{i=1}^n f_i(u_i^k) = \zeta^*$$

where ζ^* denotes the optimal solution of Problem (3.96). Moreover, all limit point of $\{u_i^k\}_{i=1}^n$ are optimal solutions.

Proof. (a) According to Proposition 3.4.3, Algorithm 8 is a particular instance of Algorithm 7 for $\boldsymbol{\mathcal{H}} = (\mathbb{R}^m)^n$, $G_i = I$ and $T_i = \partial h_i$, for $i = 1, \dots, n$, where h_i is defined as in (3.100). Moreover, Assumption ($\mathcal{B}1$) implies that the dual problem (3.98) has at least one solution, and hence, there exist a solution to the monotone inclusion (3.1). The above considerations, (3.114)-(3.116) and Theorem 3.3.9 implies that $\{(z^k, w_1^k, \dots, w_{n-1}^k)\}$ converges to some point $(z^*, w_1^*, \dots, w_{n-1}^*)$ in $\mathcal{S}$ defined in (3.103) and that $\{x_1^k\}, \dots, \{x_n^k\}$ all converge also to the same limit z^* (the convergence is strong because weak and strong convergence coincide in the finite dimension setting).

(b) Likewise, Proposition 3.4.3 and Theorem 3.3.9 also ensure that $b_i - M_i u_i^k = y_i^k \to w_i^*$, as $k \to \infty$ for each $i = 1 \ldots, n$, which implies that

$$\sum_{i=1}^{n} M_i u_i^k - b \to -\sum_{i=1}^{n} w_i^* = 0, \quad \text{as} \quad k \to \infty.$$

(c) Since we have assumed that Assumption ($\mathcal{B}2$) holds, then using item (a) and (3.102) we have:

$$w_i^* \in T_i(z^*) = -M_i \partial f_i^*(-M_i^T z^*) + b_i \quad \text{and} \quad \sum_{i=1}^{n} w_i^* = 0.$$

Therefore, there must exist some $u_i^* \in \partial f_i^*(-M_i^T z^*)$ such that $w_i^* = b_i - M_i u_i^*$, for $i = 1, \ldots, n$. Since for each $i = 1, \ldots, n$, f_i is a closed proper and convex function, $u_i^* \in \partial f_i^*(-M_i^T z^*)$ is equivalent to $-M_i^T z^* \in \partial f(u_i^*)$. From which, and the last relation yields

$$0 = \sum_{i=1}^{n} w_i^* = b - \sum_{i=1}^{n} M_i u_i^* \quad \text{and} \quad 0 \in M_i^T z^* + \partial f_i(u_i^*), \quad \text{for} \quad i = 1, \ldots, n. \tag{3.117}$$

It follows that $(z^*, u_1^*, \ldots, u_n^*)$ satisfy the KKT conditions (3.99), $(u_1^*, \ldots, u_n^*)$ is an optimal solution of the primal problem (3.96) and $\sum_{i=1}^{n} f_i(u_i^*) := \zeta^*$.

Now, for each $i = 1, \ldots, n$ and all $k \geq 0$ sufficiently large, the first equality in (3.106) yields

$$f_i(u_i^k) + \langle \widehat{z}^k, M_i u_i^k \rangle + \frac{\rho_i^k}{2} \|M_i u_i^k - b_i + \widehat{w}_i^k\|^2 \leq f_i(u_i^*) + \langle \widehat{z}^k, M_i u_i^* \rangle + \frac{\rho_i^k}{2} \|M_i u_i^* - b_i + \widehat{w}_i^k\|^2. \tag{3.118}$$

Moreover, since for each $i = 1, \ldots, n$, $\lim_{k\to\infty} \widehat{w}_i^k = \lim_{k\to\infty} w_i^k = w_i^*$ and $\lim_{k\to\infty}\{b_i - M_i u_i^k\} = \lim_{k\to\infty} y_i^k = w_i^*$, we have

$$\begin{aligned} M_i u_i^k - b_i + \widehat{w}_i^k &\to -w_i^* + w_i^* = 0, \\ M_i u_i^* - b_i + \widehat{w}_i^k &= -w_i^* + w_i^k \to -w_i^* + w_i^* = 0, \\ M_i u_i^k &= b_i - y_i^k \to b_i - w_i^* = M_i u_i^*. \end{aligned} \tag{3.119}$$

Taking limit in (3.118), using the fact that limit $\lim_{k\to\infty} \widehat{z}^k = \lim_{k\to\infty} z^k = z^*$, (3.119) and (3.116), we conclude that

$$\limsup_{k\to\infty} f_i(u_i^k) + \langle z^*, M_i u_i^* \rangle \leq f_i(u_i^*) + \langle z^*, M_i u_i^* \rangle.$$

Cancelling the identical terms $\langle z^*, M_i u_i^* \rangle$ from both sides and then summing over $i = 1, \ldots, n$, we obtain

$$\limsup_{k\to\infty} \sum_{i=1}^{n} f_i(u_i^k) \leq \sum_{i=1}^{n} \limsup_{k\to\infty} f_i(u_i^k) \leq \sum_{i=1}^{n} f_i(u_i^*) = \zeta. \tag{3.120}$$

On the other hand, for each $k \geq 0$ and each $i = 1, \ldots, n$, using (3.117) and definition of subdifferential of f_i, we have

$$f_i(u_i^*) \leq f_i(u_i^k) - \langle M_i^T z^*, u_i^* - u_i^k \rangle = f_i(u_i^k) - \langle z^*, M_i u_i^* - M_i u_i^k \rangle.$$

Summing over $i = 1, \ldots, n$ in the above equation, we obtain

$$\begin{aligned} \zeta^* = \sum_{i=1}^n f_i(u_i^*) &\leq \sum_{i=1}^n f_i(u_i^k) - \langle z^*, \sum_{i=1}^n M_i u_i^* - \sum_{i=1}^n M_i u_i^k \rangle \\ &= \sum_{i=1}^n f_i(u_i^k) + \langle z^*, b - \sum_{i=1}^n M_i u_i^k \rangle, \end{aligned}$$

where in the second equality we used the fact that $(z^*, u_1^*, \ldots, u_n^*)$ satisfy the KKT condition. Hence, the above inequality together with item (b) yields

$$\zeta^* \leq \liminf_{k\to\infty} \sum_{i=1}^n f_i(u_i^k). \tag{3.121}$$

The first part of (c) follows from (3.120) and (3.121). To end the proof, let $(u_1^\infty, \ldots, u_n^\infty)$ be any limit point of $\{(u_1^k, \ldots, u_n^k)\}$, then there exist a subsequence $\{(u_1^{k_j}, \ldots, u_n^{k_j})\}$ such that $u_i^{k_j} \to u_i^\infty$ as $j \to \infty$, for each $i = 1, \ldots, n$. It follows that

$$\begin{aligned} \sum_{i=1}^n f_i(u_i^\infty) &\leq \liminf_{j\to\infty} \left\{ \sum_{i=1}^n f_i(u_i^{k_j}) \right\} \\ &\leq \limsup_{j\to\infty} \left\{ \sum_{i=1}^n f_i(u_i^{k_j}) \right\} \leq \limsup_{k\to\infty} \left\{ \sum_{i=1}^n f_i(u_i^k) \right\} \leq \zeta^*, \end{aligned}$$

where, the first inequality follows due to that f_i is closed and thus lower semicontinuous, and the last inequality comes from (3.120). On the other hand, since by item (b) $\sum_{i=1}^n M_i u_i^k - b \to 0$, as $k \to \infty$, we have $\sum_{i=1}^n M_i u_i^\infty = b$, which means that $(u_1^\infty, \ldots, u_n^\infty)$ is a feasible point for (3.96) and also that $\sum_{i=1}^n f_i(u_i^\infty) \geq \zeta^*$, due to ζ^* is the optimal value of (3.96). Therefore, $\sum_{i=1}^n f_i(u_i^\infty) = \zeta^*$ and $(u_1^\infty, \ldots, u_n^\infty)$ is a optimal solution of (3.96).

□

In order to develop global convergence bounds for our method we will examine how well its iterates satisfy the KKT conditions (3.99). Observe that the inclusions in (3.113) indicates that the quantities $\| \sum_{i=1}^n M_i u_i^k - b\|$ and $\|x_i^k - x_n^k\|$, for $i = 1, \ldots, n-1$ can be used to measure the accuracy of an iterate $(u_1^k, \ldots, u_n^k, z^k)$ to a saddle point of the Lagrangian function [2]. More

[2]The Lagrangian function $\mathcal{L} : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_n} \times \mathbb{R}^m \to (-\infty, +\infty]$ for problem (3.96) is defined as

$$\mathcal{L}(u_1, \ldots, u_n, z) = \sum_{i=1}^n f_i(u_i) + \left\langle \sum_{i=1}^n M_i u_i - b, z \right\rangle.$$

A point $(u^*, z^*) := (u_1^*, \ldots, u_n^*, z^*)$ such that $\mathcal{L}(u_1^*, \ldots, u_n^*, z^*) < +\infty$ and it satisfies

$$\min_{u_i, \ldots, u_n} \mathcal{L}(u_1, \ldots, u_n, z^*) = \mathcal{L}(u_1^*, \ldots, u_n^*, z^*) = \max_z \mathcal{L}(u_1^*, \ldots, u_n^*, z)$$

is called a saddle point of the Lagrangian function $\mathcal{L}$.

specifically, if we define the primal and dual residuals, associated with $(u_1^k, \dots, u_n^k, z^k)$, by

$$r_p^k = \sum_{i=1}^{n} M_i u_i^k - b,$$
$$r_d^k = (x_1^k - x_n^k, \cdots, x_{n-1}^k - x_n^k).$$

Thus, the inclusion (3.122) and the KKT conditions follows that $(u_1^k, \dots, u_n^k, z^k)$ is a saddle point of L, if $\|r_p^k\| = \|r_d^k\| = 0$. Therefore, the size of the these residuals indicates how far the iterates are from a saddle point, and it can be viewed as an error measurement of the Algortihm 8.

Theorem 3.4.6 (nonasymptotic convergence rate)**.** Consider the sequences evolved by *Algorithm 8* and suppose that Assumptions ($\mathcal{B}1$) and ($\mathcal{B}2$) hold. Assume moreover that $\alpha \in [0,1)$, $\overline{\beta} \in (0,2)$ and $\{\alpha_k\}$ satisfies (for some $\overline{\alpha} > 0$) the conditions (3.84)-(3.86) of *Theorem* 3.3.9. Let d_0 be the distance of $p^0 = (z^0, w_1^0, \dots, w_{n-1}^0)$ to the set $\mathcal{S}$ defined in (3.103), then for all $k \geq 0$, we have

$$0 \in M_i^T x_i^k + \partial f_i(u_i^k) \qquad \forall\, i = 1, \dots, n \tag{3.122}$$

and there exists $j \in \{0, \dots, k\}$ such that

$$\begin{aligned} \|b - \textstyle\sum_{i=1}^n M_i u_i^j\| &\leq \tfrac{d_0}{\xi\sqrt{k}} \sqrt{\gamma \underline{\beta}^{-1} \left(2 - \overline{\beta}\right)^{-1} \left(1 + \tfrac{\alpha(1+2\alpha-\alpha^3)}{(1-\alpha)^2 q(\alpha)}\right)}, \\ \|x_i^j - x_n^j\| &\leq \tfrac{d_0}{\xi\sqrt{k}} \sqrt{\underline{\beta}^{-1} \left(2 - \overline{\beta}\right)^{-1} \left(1 + \tfrac{\alpha(1+2\alpha-\alpha^3)}{(1-\alpha)^2 q(\alpha)}\right)} \quad i = 1, \dots, n-1; \end{aligned} \tag{3.123}$$

where

$$\xi = \frac{(1-\sigma^2)\min\{\overline{\rho}, \overline{\rho}^{-1}\}}{2\max\{2(n-1), \gamma^{-1} n\}} \tag{3.124}$$

and

$$q(\alpha) := 2\left(\overline{\beta}^{-1} - 1\right)\alpha^2 - \left(4\overline{\beta}^{-1} - 1\right)\alpha + 2\overline{\beta}^{-1} - 1, \quad \alpha \in \mathbb{R}. \tag{3.125}$$

Proof. In view of Proposition 3.4.3, Theorem 3.3.10 (with T_i defined in (3.102) and $G_i = I$, for all $i = 1, \dots, n$) and the fact that $\sum_{i=1}^n y_i^k = b - \sum_{i=1}^n M_i u_i^k$, one obtain (3.123), where the scalar $\xi > 0$ in (3.124) is exactly the same that in (3.65) (into the proof of Lemma 3.3.6) with $G_i = I$ forall $i = 1 \dots, n$. The assertion in (3.122) follow immediately from the optimally condition in (3.106) and definitions of x_i^k and y_i^k in (3.110).

□

Remark 3.4.7. (i) Since Algorithm 8 is based on a projective splitting framework (Algorithm 7), the penalty parameter ρ_i^k need not be fixed with respect to either the subsystems i or the iterations k, but only the boundeness (see Assumption ($D4$) in Section 3.3). Unlike the most ADMM-like algorithms, Algorithm 8 allows that the proximal parameters to vary from iteration to iteration.

(ii) Analogous remarks to those made in the Remark 3.3.11(i) also apply here, for given tolerances $\epsilon, \delta > 0$ on the primal and dual residuals, i.e., $\|r_p^k\| \leq \delta$ and $\|r_d^k\| \leq \epsilon$, Algorithm 8 finds a KKT point $(u_1, \dots, u_n, z)$ after performing at most

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{d_0^2}{\epsilon^2} \right\rceil, \left\lceil \frac{d_0^2}{\delta^2} \right\rceil\right\}\right)$$

iterations.

Chapter 4

Final considerations and future perspectives

In this thesis, we presented some contributions to inexact methods for solving monotone inclusion problems combining inertial and relaxation effects, which can help us to better understand acceleration notion in numerical methods for solving general monotone inclusion problems.

The main contributions of this thesis were stated in Chapters 2 and 3 and, as already pointed out, we emphasize that the content of the Chapter 2 resulted in a published article [8], and part of our contributions in the chapter 3 is contained in the manuscript [83].

4.1 Main results

Summarizing the main contributions of this thesis.

- In Chapter 2, we proposed and studied the asymptotic convergence and iteration-complexity of an inertial under-relaxed HPE-type method. As applications, we proposed and studied inertial (under-relaxed) versions of the Tseng's modified forward-backward and forward-backward methods for solving structured monotone inclusion problems with either Lipschitz continuous or cocoercive operators. All the proposed and/or studied algorithms, namely Algorithms 2, 3, 4 and 5 potentially benefit from a specific policy for choosing the upper bound on the sequence of extrapolation parameters, in which case (under) relaxation plays a central role (see Assumption (**A**) and Theorems 2.2.11, 2.2.13, 2.3.3 and 2.3.7). We provided nonasymptotic global convergence rates (iteration-complexity) for inertial HPE-type methods, in particular for the proposed inertial forward-backward and Tseng's modified forward-backward methods.

- In Chapter 3, we have proposed and studied a relative-error inertial-relaxed inexact projective splitting algorithm for solving structured monotone inclusion problems involving the sum of finitely many maximal monotone operators, namely Algorithm 7. The proposed algorithm is a inertial variant of the projective splitting method introduced in [52, 68] for problem 3.1, which can be seen as a feasibility problem of finding a point in a convex and closed subset. So, we proposed a general framework for the feasible problem of finding points in a convex and closed subset with inertial effects for which we established weak convergence (see Propositions 3.2.3 and 3.2.4). Based in these last two results we established weak convergence for Algorithm 7, see Theorems 3.3.8 and 3.3.9. Furthermore, convergence rate and iteration-complexity for Algorithm 7 were also studied; see Theorem 3.3.10. Finally as an application of Algorithm 7 we derived an inertial algorithm resembling the multi block ADMM for linearly constrained optimization problem (Algorithm 8), among other contributions we stabilised a convergence result and convergence rate for this latest algorithm; see Theorems 3.4.5 and 3.4.6.

4.2 Future research

During our study, we encountered many problems whose treatment was not considered in this work and we mention only a few among the ones that we consider relevant, challenging and closely related to our work.

- The algorithms proposed and discussed in Chapters 2 and 3 are two-step iterative schemes, inspired from classical multi-step methods in numerical analysis, it remains as a topic for future research to construct methods that generalize the inertial methods (two step) to multi-step algorithms for inclusion problems. See [43, 102, 105] for some contributions in this direction.

- As we mentioned early, projective splitting methods involving inertial steps is fresh in the literature. It would be interesting to explore more a general procedure than Algorithm 7, allowing forward steps, asynchronous and block iterative implementation combining relaxation and inertial steps. Another interesting challenge in this line is to explore the progressive hedging method for stochastic programming of Rockefellar and Wets [113] within a projective splitting environment (see [53]) and inertial effects.

- Regarding to the *primal-dual splitting algorithms*, we consider a systems of monotone inclusions of the following general form: Let m and K be positives numbers, let $\{\mathcal{H}_i\}_{i=1}^m$ and $\{\mathcal{G}_k\}_{i=1}^K$ be a family of real Hilbert spaces, and for every $i = 1, \ldots, m$ and $k = 1, \ldots, K$ let $A_i : \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ and $B_k : \mathcal{G}_k \rightrightarrows \mathcal{G}_k$ be maximal monotone operators.

  Find $\overline{x}_1 \in \mathcal{H}_1, \ldots, \overline{x}_m \in \mathcal{H}_m$ such that

$$(\forall i = 1, \ldots, m) \qquad z_i \in A_i\overline{x}_i + \sum_{i=1}^{m} L_{ki}^* \left(B_k \left(\sum_{j=1}^{m} L_{kj}\overline{x}_j - r_j \right) \right), \tag{4.1}$$

  and, its dual problem in the sense of Attouch-Théra [18]: find $\overline{w}_1 \in \mathcal{G}_1, \ldots, \overline{w}_K \in \mathcal{G}_K$ such that

$$(\forall k = 1, \ldots, K) \qquad -r_k \in \sum_{i=1}^{m} L_{ki}^* \left(A_i^{-1} \left(z_i - \sum_{l=1}^{K} L_{lj}^*\overline{w}_l \right) \right) + B_k^{-1}\overline{w}_k^*. \tag{4.2}$$

  The associated extended solution (or Kuhn-Tucker) set

$$Z = \{(x_1, \ldots, x_m, w_1, \ldots, w_K) \mid (\forall i \in \{1, \ldots, m\})\ z_i - \sum_{k=1}^{K} L_{ki}^* w_k \in A_i x_i$$
$$\text{and } (\forall k \in \{1, \ldots, K\})\ \sum_{k=1}^{m} L_{ki}x_i - r_k \in B_k^{-1}x_k\}.$$

The problem (4.1) is equivalent to find a point $(\bar{x}, \bar{w}) := (\bar{x}_1, \ldots, \bar{x}_m, \bar{w}_1, \ldots, \bar{w}_K)$ in the convex and closed set Z, where $\bar{x}$ solve the primal problem (4.1) and $\bar{w}$ solve the dual problem (4.2). The importance of Problem (4.1) is due to the fact that it models a wide range of problems arising in game theory, image recovery, evolution equations, machine learning, signal processing, and domain decomposition methods in partial differential equations, as discussed in [2, 41, 42] and references therein.

In [42, Proposition 2], it was shown that the problem (4.1) can be seen as an inclusion problem of the type: $0 \in T_1 z + G^* T_2 G z$, where T_1 and T_2 are maximal monotone operators defined on the product spaces $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \times \ldots \times \mathcal{H}_m$ and $\boldsymbol{\mathcal{G}} = \mathcal{G}_1 \times \ldots \times \mathcal{G}_K$, respectively, and $G : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}}$ is a linear and bounded operator, for which a projective splitting algorithm was developed by the same authors. Following the study done in Chapter 2, a more general method can be developed by adding inertial steps and permitting asynchronous and block-iterative implementation.

An important special case of (4.1) is the optimization problem below, in which the monotone operators $\{A_i\}_{i=1}^m$ and $\{B_k\}_{k=1}^K$ are taken to be subdifferentials of convex functions: Consider the primal minimization problem

$$\min_{x_1 \in \mathcal{H}_i, \ldots, x_m \in \mathcal{H}_m} \sum_{i=1}^{m} \left(f_i(x_i) - \langle x_i, z_i \rangle\right) + \sum_{k=1}^{K} g_k \left(L_{ki} x_i - r_k\right) \tag{4.3}$$

and its dual problem

$$\min_{w_1 \in \mathcal{G}_1, \ldots, w_K \in \mathcal{G}_K} \sum_{i=1}^{m} f_i^* \left(z_i - \sum_{k=1}^{K} L_{ki}^* w_k \right) + \sum_{k=1}^{K} \left(g_k(w_k) + \langle w_k, r_k \rangle\right),$$

where $f_i \in \Gamma_0(\mathcal{H}_i)$ for all $i \in \{1, \ldots, m\}$ and $g_k \in \Gamma_0(\mathcal{G}_k)$ forl all $k \in \{1, \ldots, K\}$ and $L_{ki} : \mathcal{H}_i \to \mathcal{G}_k$ as be linear and bounded operators. It would also be interesting to build an inertial and implementable method for solving (4.3), moreover, to compare its computational performance with other existing primal dual splitting type methods (see e.g., [23, 24]).

Appendix A

Basic results

A.1 Basic results in $\mathbb{R}$

The following lemma was essentially proved by Alvarez and Attouch in [5, Theorem 2.1] but we present here for completeness to the work.

Lemma A.1.1. Let the sequences $\{h_k\}$, $\{s_k\}$, $\{\alpha_k\}$ and $\{\delta_k\}$ in $[0, +\infty[$ and $\alpha \in \mathbb{R}$ be such that $h_0 = h_{-1}$, $0 \leq \alpha_{k-1} \leq \alpha < 1$ and

$$h_k - h_{k-1} + s_k \leq \alpha_{k-1}(h_{k-1} - h_{k-2}) + \delta_k \qquad \forall k \geq 1. \tag{A.1}$$

The following hold:

(a) For all $k \geq 1$,

$$h_k + \sum_{j=1}^{k} s_j \leq h_0 + \frac{1}{1-\alpha} \sum_{j=1}^{k} \delta_j. \tag{A.2}$$

(b) If $\sum_{k=1}^{\infty} \delta_k < +\infty$, then $\lim_{k\to\infty} h_k$ exist, i.e., the sequence $\{h_k\}$ converges to some element in $[0, \infty[$.

Proof. (a) For each $k \geq 0$ set $\theta_k := h_k - h_{k-1}$. Since $s_k \geq 0$ and $0 \leq \alpha_k \leq \alpha$ for all $k \geq 0$, it follows from (A.1) that

$$\theta_k \leq \alpha_{k-1}\theta_{k-1} + \delta_k \leq \alpha[\theta_{k-1}]_+ + \delta_k,$$

where $[t]_+ := \max\{t, 0\}$ for $t \in \mathbb{R}$. Therefore, we have

$$[\theta_k]_+ \leq \alpha[\theta_{k-1}]_+ + \delta_k \leq \alpha^k[\theta_0]_+ + \sum_{i=0}^{k-1} \alpha^i \delta_{k-i}. \tag{A.3}$$

Note that assumption $h_0 = h_{-1}$, implies that $[\theta_0]_+ = \theta_0 = 0$. Hence, it follows from (A.3) that

$$\sum_{j=1}^{k} [\theta_j]_+ \leq \sum_{j=1}^{k} \sum_{i=0}^{j-1} \alpha^i \delta_{j-i} = \sum_{j=1}^{k} \left(\sum_{i=0}^{k-j} \alpha^i \right) \delta_j = \sum_{j=1}^{k} \delta_j \left(\frac{1-\alpha^{k-j+1}}{1-\alpha} \right) \leq \frac{1}{1-\alpha} \sum_{j=1}^{k} \delta_j. \tag{A.4}$$

Moreover, from (A.1) we have

$$s_k \le h_{k-1} - h_k + \alpha[\theta_{k-1}]_+ + \delta_k.$$

Summing the latter inequality over all $j = 1, \dots, k$ and taking into account (A.4) we obtain

$$\sum_{j=1}^{k} s_j \le h_0 - h_k + \alpha \sum_{j=1}^{k-1}[\theta_j]_+ + \sum_{j=1}^{k} \delta_j \le h_0 - h_k + \frac{1}{1-\alpha}\sum_{j=1}^{k} \delta_j,$$

which shows the desired conclusion in (a).

(b) First note that (A.4) and the assumption $\sum_{k=1}^{\infty} \delta_k < +\infty$ implies

$$\sum_{k=1}^{\infty}[\theta_j]_+ \le \frac{1}{1-\alpha}\sum_{j=1}^{\infty} \delta_j < \infty. \tag{A.5}$$

Now, consider the sequence $\{\gamma_k\}$ defined by $\gamma_k = h_k - \sum_{j=1}^{k}[\theta_j]_+$. Since $h_k \ge 0$ for all $k \ge 0$ and $\sum_{j=1}^{\infty}[\theta_j]_+ < \infty$, it follows that $\{\gamma_k\}$ is bounded below. On the other hand,

$$\gamma_{k+1} = h_{k+1} - [\theta_{k+1}]_+ - \sum_{j=1}^{k}[\theta_j]_+ \le h_{k+1} - \theta_{k+1} - \sum_{j=1}^{k}[\theta_j]_+ = h_k - \sum_{j=1}^{k}[\theta_j]_+ = \gamma_k,$$

and so, $\{\gamma_k\}$ is nonincreasing. As a consequence, $\{\gamma_k\}$ converges as $k \to \infty$, it and (A.5) ensure that the following limit

$$\lim_{k\to\infty} h_k = \lim_{k\to\infty}\left(\gamma_k + \sum_{j=1}^{k}[\theta_j]_+\right) = \lim_{k\to\infty}\gamma_k + \sum_{j=1}^{\infty}[\theta_j]_+$$

exist. This completes the proof of the lemma.

□

Lemma A.1.2. For any $\sigma \in [0, 1[$, the inverse function of the scalar map

$$A :=]0, 1+\sigma] \ni t \mapsto \frac{4-2t}{4-t+\sqrt{16t-7t^2}} \in \left[\frac{2(1-\sigma)}{3-\sigma+\sqrt{9+2\sigma-7\sigma^2}}, 1\right[=: B$$

is given by

$$B \ni \beta \mapsto \frac{2(\beta-1)^2}{2(\beta-1)^2+3\beta-1} \in A.$$

Lemma A.1.3. [6, Lemma A.2] The inverse function of the scalar map

$$(0,2) \ni \beta \mapsto \frac{2(2-\beta)}{4-\beta+\sqrt{16\beta-7\beta^2}} \in (0,1)$$

is given by

$$(0,1) \ni \overline{\alpha} \mapsto \frac{2(\overline{\alpha}-1)^2}{2(\overline{\alpha}-1)^2+3\overline{\alpha}-1} \in (0,2).$$

Lemma A.1.4. [6, Lemma A.3] Let $\mathbb{R} \ni \nu \mapsto q(\nu) := a\nu^2 - b\nu + c$ be a real function and assume that $b, c > 0$ and $b^2 - 4ac > 0$. Define

$$\overline{\alpha} := \frac{2c}{b + \sqrt{b^2 - 4ac}} > 0. \tag{A.6}$$

(i) If $a = 0$, then $q(\cdot)$ is a decreasing affine function and $\overline{\alpha} > 0$ as in (A.6) is its unique root (see Figure A.1(a)).

(ii) If $a > 0$ (resp. $a < 0$), then $q(\cdot)$ is a convex (resp. concave) quadratic function and $\overline{\alpha} > 0$ as in (A.6) is its smallest (resp. largest) root (see Figure A.1(b) and Figure A.1(c), resp.).

In both cases (i) and (ii), $\overline{\alpha} > 0$ as in (A.6) is a root of $q(\cdot)$, and $q(\cdot)$ is decreasing in the interval $[0, \overline{\alpha}]$ (see Figure A.1).

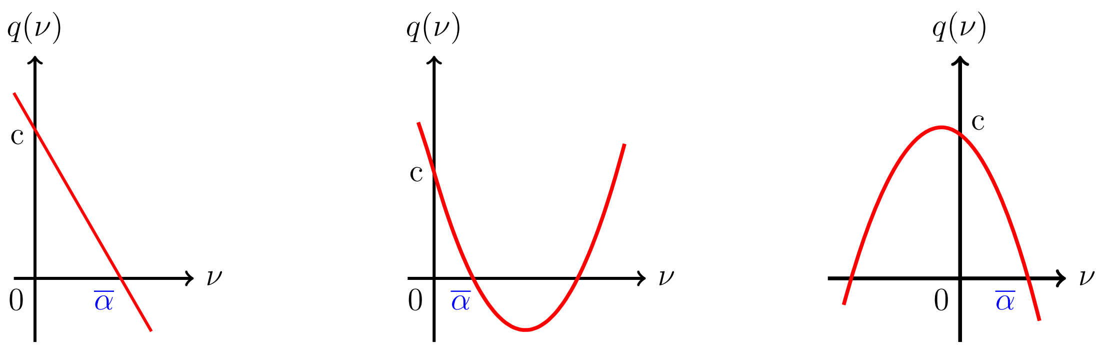

Figure A.1: Possible cases for the real function $q(\cdot)$ in Lemma A.1.4.

A.2 Some auxiliary results

Lemma A.2.1. Let $\mathcal{H}$ a real Hilbert space, the following statement holds

(i) For any $x, y \in \mathcal{H}$ and $t \in \mathbb{R}$, we have

$$\|tx + (1-t)y\|^2 = t\|x\|^2 + (1-t)\|x\|^2 - t(1-t)\|x - y\|^2. \tag{A.7}$$

(ii) For any $x_1, x_2, \ldots, x_n \in \mathcal{H}$, we have

$$\|x_1 + \ldots + x_n\|^2 \leq n\left(\|x_1\|^2 + \ldots + \|x_n\|^2\right). \tag{A.8}$$

Lemma A.2.2 (Opial)**.** Let $\mathcal{H}$ be a real Hilbert space, $\emptyset \neq \mathcal{S} \subset \mathcal{H}$ and let $\{p^k\}$ be any sequence in $\mathcal{H}$ such that

(a) $\lim_{k\to\infty} \|p^k - p\|$ exists for every $p \in \mathcal{S}$;

(b) every weak cluster point of $\{p^k\}$, as $k \to \infty$ belongs to $\mathcal{S}$.

Then $\{p^k\}$ converges weakly as $k \to \infty$ to a point in $\mathcal{S}$.

The lemma below was proved (with a different notation) in [2, Proposition 2.4].

Lemma A.2.3. Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{G} \rightrightarrows \mathcal{G}$ be maximal monotone operators and let $G : \mathcal{H} \to \mathcal{G}$ be a bounded linear operator. Let also $a^k \in A(r^k)$ and $b^k \in B(s^k)$ be such that $r^k \rightharpoonup r^\infty$ and $b^k \rightharpoonup b^\infty$, for some $r^\infty \in \mathcal{H}$ and $b^\infty \in \mathcal{G}$. If, $a^k + G^* b^k \to 0$ and $Gr^k - s^k \to 0$, then $b^\infty \in B(Gr^\infty)$ and $-L^* b^\infty \in A(r^\infty)$.

Bibliography

[1] B. Abbas and H. Attouch. Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator. *Optimization*, 64(10):2223–2252, 2015.

[2] A. Alotaibi, P. L. Combettes, and N. Shahzad. Solving coupled composite monotone inclusions by successive fejér approximations of their Kuhn-Tucker set. *SIAM J. Optim.*, 24(8):2076–2095, 2014.

[3] F. Alvarez. On minimization property of a second order dissipative system in Hilber spaces. *SIAM J. Control Optim.*, 38(4):1102–1119, 2000.

[4] F. Alvarez. Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space. *SIAM J. Optim.*, 14(3):773–782, 2003.

[5] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.*, 9(1-2):3–11, 2001. Wellposedness in optimization and related topics (Gargnano, 1999).

[6] M. Marques Alves, J. Eckstein, M. Geremia, and J.G. Melo. Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. *Comput. Optim. Appl.*, 75(2):389–422, 2020.

[7] M. Marques Alves and S. C. Lima. An inexact Spingarn's partial inverse method with applications to operator splitting and composite optimization. *J. Optim. Theory Appl.*, 175(3):818–847, 2017.

[8] M. Marques Alves and R.T. Marcavillaca. On inexact relative-error hybrid proximal extragradient, forward-backward and Tseng's modified forward-backward methods with inertial effects. *Set-Valued Var. Anal.*, 28:301–325, 2019.

[9] A. S. Antipin. Minimization of convex functions on convex sets by means of differential equations. *Differential Equations*, 30(9):1365–1375, 1994.

[10] H. Attouch and A. Cabot. Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. *Appl. Math. Optim.*, 80(3):547–598, 2019.

[11] H. Attouch and A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Math. Program.*, 2019.

[12] H. Attouch and A. Cabot. Convergence rate of a relaxed inertial proximal algorithm for convex minimization. *Optimization*, 69(6):1281–1312, 2020.

[13] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.*, 168(1-2, Ser. B):123–175, 2018.

[14] H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction I. the continuous dynamical system. *Comm. Contemp. Math.*, 2(1):1–34, 2000.

[15] H. Attouch, M. Marques Alves, and B.F. Svaiter. A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $O(1/n^2)$. *Journal of Convex Analysis*, 23(1):139–180, 2016.

[16] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Math. Program.*, 174:391–432, 2019.

[17] H. Attouch, J. Peypouquet, and P. Redont. Fast convex optimization via inertial dynamics with Hessian driven damping. *J. Differential Equations*, 261(10):5734–5783, 2016.

[18] H. Attouch and M. Thera. A general duality principle for the sum of two operators. *Journal of Convex Analysis*, 3:1–24, 1996.

[19] B. Banert and R. I. Boţ. A forward-backward-forward differential equation and its asymptotic properties. *Journal of Convex Analysis*, 25(2):371–388, 2018.

[20] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces, second edition*. Springer, New York, 2017.

[21] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

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

[23] R. I. Boţ and E. R. Csetnek. An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems. *Numerical Algorithms*, 3(71):519–540, 2016.

[24] R. I. Boţ, E. R. Csetnek, and C. Hendrich. Inertial Douglas-Rachford splitting for monotone inclusion problems. *Appl. Math. Comput.*, 256:472–487, 2015.

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

[26] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North-Holland Mathematics Studies. 5. Amsterdam-London: North-Holland Publishing Comp.; New York: American Elsevier Publishing Comp., 1973.

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

[28] F. E. Browder. Nonlinear elliptic boundary value problems. *Bull. Amer. Math. Soc.*, 69(6):862–874, 1963.

[29] F. E. Browder. The solvability of non-linear functional equations. *Duke Math. J.*, 30(4):557–566, 1963.

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

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

[32] R. S. Burachik and B. F. Svaiter. ϵ-enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal.*, 7(2):117–132, 1999.

[33] J. V. Burke and Maijian Qian. A variable metric proximal point algorithm for monotone operators. *SIAM J. Control Optim.*, 37(2):353–375, 1999.

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

[35] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[36] C. Chen, R. H. Chan, S. Ma, and J. Yang. Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.*, 8(4):2239–2267, 2015.

[37] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1):57–79, 2016.

[38] C. Chen, S. Ma, and J. Yang. A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM J. Optim.*, 25(4):2120–2142, 2015.

[39] X. Chen and M. Fukushima. Proximal quasi-Newton methods for nondifferentiable convex optimization. *Math. Program.*, 85(2):313–334, 1999.

[40] P. L. Combettes. *Fejér monotonicity in convex optimization*, volume 2 of *In: Encyclopedia of optimization*. Springer Science Business Media, 2001.

[41] P. L. Combettes. Systems of structured monotone inclusions: duality, algorithms, and applications. *SIAM J. Optim.*, 4(23):2420–2447, 2013.

[42] P. L. Combettes and J. Eckstein. Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions. *Math. Program.*, 1-2(168):645–672, 2018.

[43] P. L. Combettes and L.E. Glaudin. Quasi-nonexpansive iterations on the affine hull of orbits: From Mann's mean value algorithm to inertial methods. *SIAM J. Optim.*, 27(4):2356–2380, 2017.

[44] P. L. Combettes and J.C. Pesquet. Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212, 2011.

[45] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling Simulation*, 4(4):1168–1200, 2005.

[46] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-Valued Var. Anal.*, 25(4):829–858, 2017.

[47] W. Deng, M. Lai, Z. Peng, and O. Yin. Parallel multi-block admm with o(1/k) convergence. *J. Sci. Comput.*, 71:712–736, 2017.

[48] J. Eckstein. A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. *J. Optim. Theory Appl.*, 1(173):155–182, 2017.

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

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

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

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

[53] J. Eckstein, J. P. Watson, and D. L. Woodruf. Asynchronous projective hedging for stochastic programming. Preprint, http://www.optimization-online.org/DB-FILE/2018/10/6895, 2018.

[54] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas-Rachford splitting and special cases of the ADMM. *Math. Program.*, 170:417–444, 2018.

[55] F. Facchinei and J. S. Pang. *Finite-dimensional variational inequalities and complementarity problems. Vol. I.* Springer Series in Operations Research. Springer-Verlag, New York, 2003.

[56] M. Fortin and R. Glowinski. Chapter ix applications of the method of multipliers to variational inequalities. In *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and Its Applications*, pages 299–331. Elsevier, Dordrecht, 1983.

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

[58] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.

[59] R. Glowinski, S. J. Osher, and W. Yin, editors. *Splitting methods in communication, imaging, science, and engineering*. Scientific Computation. Springer, Cham, 2016.

[60] M.L.N. Goncalves, J.G. Melo, and R.D.C. Monteiro. 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.

[61] J. Haraux and N. A. Jendoubi. Convergence of solutions of second order gradient like systems with analytic nonlinearities. *J. Differential Equations*, 144(2):313–320, 1998.

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

[63] M. Hong and Z. Q. Luo. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162:165–199, 2017.

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

[65] A. N. Iusem and B. F. Svaiter. A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization*, 42(4):309–321, 1999.

[66] F. Iutzeler and Hendrickx J. M. A generic online acceleration scheme for optimization algorithms via relaxation and inertia. *Optimization Methods and Software*, pages 383–405, 2017.

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

[68] P. R. Johnstone and J. Eckstein. Projective splitting with forward steps: Asynchronous and block-iterative operator splitting. Preprint, arXiv:1803.07043v5, 2018.

[69] P. R. Johnstone and J. Eckstein. Convergence rates for projective splitting. *SIAM J. Optim.*, 29(3):1931–1957, 2019.

[70] P. R. Johnstone and J. Eckstein. Single-forward-step splitting: Explointing cocoercivity. Preprint, arXiv:1902.09025v2, 2019.

[71] P. R. Johnstone and J. Eckstein. Projective splitting with forward steps only requires continuity. *Optimization Letters*, 14:229–247, 2020.

[72] R. I. Kachurovskii. Monotone operators and convex functionals. *Usp. Mat. Nauk*, 15(4):213–215, 1960.

[73] R. I. Kachurovskii. Non-linear monotone operators in Banach spaces. *Uspekhi Mat. Nauk*, 23(2):121–168, 1968.

[74] G. M. Korpelevič. An extragradient method for finding saddle points and for other problems. *Èkonom. i Mat. Metody*, 12(4):747–756, 1976.

[75] M.A. Krasnosel'skii. Two remarks on the method of successive aproximations. *Uspekhi Matematicheskikh Nauk*, 10(1):123–127, 1955.

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

[77] D. A. Lorenz and T. Pock. An inertial forward-backward algorithm for monotone inclusions. *J. Math. Imaging Vision*, 51(2):311–325, 2015.

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

[79] M. P. Machado. On the complexity of the projective splitting and Spingarn's methods for the sum of two maximal monotone operators. *J Optim. Theory Appl.*, 178:153–190, 2018.

[80] M. P. Machado. Projective method of multipliers for linearly constrained convex minimization. *Comput. Optim. and Appl.*, 73:237–273, 2019.

[81] P. E. Maingé and N. Merabet. A new inertial-type hybrid projection-proximal algorithm for monotone inclusions. *Appl. Math. Comput.*, 215(9):3149–3162, 2010.

[82] W. R. Mann. Mean value methods in iterations proximal. *Proc. Amer. Math.*, 4(3):506–510, 1953.

[83] M. Marques Alves, M. Geremia, and R.T. Marcavillaca. A relative-error inertial-relaxed inexact projective splitting algorithm. Preprint, https://arxiv.org/abs/2002.07878v1, 2020.

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

[85] B. Mercier and G. Vijayasundaram. *Lectures on topics in finite element solution of elliptic problems*. Tata Institute on Mathematics and Phisics. Springer, New York, 1979.

[86] G. J. Minty. Monotone networks. *roc. R. Soc. Lond. A: Math. Phys. Eng. Sci*, 257(1289):194–212, 1960.

[87] G. J. Minty. On the monotonicity of the gradient of a convex function. *Pac. J. Math.*, 14(1):243–247, 1964.

[88] R. D. C. Monteiro, C. Ortiz, and B. F. Svaiter. A first-order block-decomposition method for solving two-easy-block structured semidefinite programs. *Math. Program. Comput.*, 6(2):103–150, 2014.

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

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

[91] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle point and convex optimization problems. *SIAM J. Optim.*, 21:1688–1720, 2010.

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

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

[94] J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sc. Paris*, 255:2897–2899, 1962.

[95] J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[96] A. Moudafi and M. Oliny. Convergence of a splitting inertial proximal method for monotone operators. *J. Comput. Appl. Math.*, 155(2):447–454, 2003.

[97] Y. E. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$. *Doklady AN SSSR (transalated as Soviet Math. Docl.)*, 269:543–547, 1983.

[98] Y.E. Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.

[99] P. Ochs, Y. Chen, T. Brox, and Pock T. ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sciences*, 7(2):1388–1419, 2014.

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

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

[102] T. Pennanen and B. F. Svaiter. Solving monotone inclusions with linear multi-step methods. *Math. Program.*, Ser. A(96):469–487, 2003.

[103] J. Peypouquet. *Convex Optimization in Normed Spaces: Theory, Methods and Examples.* Springer Monographs in Mathematics. Springer Briefs in Optimization, Valparaıso, Chile, 2015.

[104] B. T. Polyak. *Introduction to Optimization.* Translations Series in Mathematics and Engineering. Optimization Software, 1987.

[105] B.T. Polyak. Some methods of speeding up the convergence of iteration methods. *U.S.S.R. Comput. Math. Phys.*, 4(5):1–17, 1964.

[106] L. Qi and X. Chen. A preconditioning proximal newton method for nondifferentiable convex optimization. *Math. Program.*, 76(3):411–429, 1997.

[107] J. Raguet, H. Fadili and G. Peyré. A generalized forward-backward splitting. *SIAM J. Imaging Sci.*, 6(3):1199–1226, 2013.

[108] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pac. J. Math.*, 17(3):497–510, 1966.

[109] R. T. Rockafellar. *Convex Analysis*, volume 28 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Princeten University Press, Princeton, 1970.

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

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

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

[113] R. T. Rockafellar and R. J. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.*, 1(16):119–147, 1991.

[114] T.R. Rockafellar and R. Rockajellm. Monotone operators associates with sadle functions and minimax problems. *American Math. Soc.*, 1970.

[115] E. Ronald and Jr. Bruck. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *Journal of Functional Analysis*, 18(1):15–26, 1975.

[116] N.Z. Shor. *Minimization Methods for Non-differentiable Functions.* Springer Series in Computational Mathematics. Springer, 1985.

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

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

[119] M. V. Solodov and B. F. Svaiter. A new projection method for variational inequality problems. *SIAM J. Control Optim.*, 37(3):765–776 (electronic), 1999.

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

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

[122] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Texts in Applied Mathematics. Springer-Verlag New York, 1998.

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

[124] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.

[125] B. F. Svaiter. A class of Fejér convergent algorithms, approximate resolvents and the hybrid proximal-extragradient method. *J. Optim. Theory Appl.*, 162(1):133–153, 2014.

[126] B. F. Svaiter. Complexity of the relaxed hybrid proximal-extragradient method under the large-step condition. Preprint A766/2015, IMPA - Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ Brasil 22460-320, 2015.

[127] L. Tianyi, M. Shiqian, and Z. Shuzhong. On the global linear convergence of the ADMM with multiblock variables. *SIAM Journal on Optimization*, 25(3):1478–1497, 2015.

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

[129] C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific. Singapore, 2002.