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

Andrea Jeniree Rujano Rodriguez

A projective splitting algorithm for monotone inclusions with cocoercive operators

Florianópolis
2022

Andrea Jeniree Rujano Rodriguez

A projective splitting algorithm for monotone inclusions with cocoercive operators

Dissertação submetida ao Programa de Pós-Graduação em Matemática Pura e Aplicada da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Matemática Pura e Aplicada.
Orientador: Prof. Maicon Marques Alves, Dr.

Florianópolis
2022

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

Rodriguez, Andrea Jeniree Rujano
A projective splitting algorithm for monotone inclusions with cocoercive operators / Andrea Jeniree Rujano Rodriguez ; orientador, Maicon Marques Alves, 2022.
81 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Programa de Pós-Graduação em Matemática Pura e Aplicada, Florianópolis, 2022.

Inclui referências.

1. Matemática Pura e Aplicada. 2. Operadores monótonos maximais. 3. Problema de inclusão monótona. 4. Algoritmos projetivos de decomposição . 5. Operadores cocoercivos. I. Alves, Maicon Marques . II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Matemática Pura e Aplicada. III. Título.

Andrea Jeniree Rujano Rodriguez

A projective splitting algorithm for monotone inclusions with cocoercive operators

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

Prof. Felipe Lara, Dr.
Universidad de Tarapacá

Prof. Vinícius Albani, Dr.
Universidade Federal de Santa Catarina

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

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Matemática Pura e Aplicada.

Documento assinado digitalmente
Daniel Goncalves
Data: 03/05/2022 10:47:38-0300
CPF: 017.022.459-77
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de
Pós-Graduação

Documento assinado digitalmente
Maicon Marques Alves
Data: 03/05/2022 17:45:19-0300
CPF: 967.631.910-49
Verifique as assinaturas em https://v.ufsc.br

Prof. Maicon Marques Alves, Dr.
Orientador

Florianópolis, 2022.

I dedicate this work to my family.

ACKNOWLEDGEMENTS

Certainly, any accomplishment that we as humans achieve is not entirely due to ourselves. There are many people who one way or another have contributed to the success of what I have set out to do. I am aware of this, and I am grateful for their presence.

I thank God for his goodness, which allows us to move towards our improvement, and for His care every day. I am grateful to my parents Maria Rodriguez and Vladimir Rujano for doing their best to raise a good person, and being motivators for me to be a good student. I am thankful to my husband Frank Osorio, for always supporting me, particularly in this endeavor even if he had to put his own aside, for being attentive to our daughter Kaori during my study moments.

I would like to thank my undergraduate professors at the "Universidad de Los Andes" (ULA) who contributed enormously to my education, a special thanks to the professors Ramón Pino, Olga Porras, and Jesús Guillen for their recommendations to the program.

I am grateful to my master's degree professors Douglas Soares Gonçalves, Fabio Botelho, Juliano de Bem Francisco, Maria Astudillo, and Paulo Mendes de Carvalho. They were excellent at communicating their knowledge and willing to help with my doubts.

Special thanks to my tutor Maicon Marques Alves for his patience during these two years and his sound advice. It has been a privilege to appreciate his expertise, his enthusiasm for mathematics, and the ease with which he conveys them. I would like to thank the members of the dissertation committee for their observations and comments that helped me delve deeper into the subject and for advice to further improve my dissertation.

Many thanks to my colleague Luiz Suzana, who was not only an incredible study partner during this time but also a great person. Also many thanks to Asteroide Santana for his encouragement during the program, and for his advice on how to improve my work. I am grateful to many Brazilians who reached out to me when I arrived here in Brazil, their hospitality will always be remembered.

I am thankful to "Universidade Federal de Santa Catarina" (UFSC) for the support provided for my studies, to the secretary Erica Flores for her quality of service, and to Bruno Wanderley Farias from the International Relations office for his assistance with my migratory processes for study purposes. Finally, I am grateful to CAPES for the financial support without which this accomplishment would not have been possible.

RESUMO

Neste trabalho, apresentamos um algoritmo iterativo para resolver o problema de inclusão na presença de operadores cocoercivos definidos em espaços de Hilbert com métodos projetivos de decomposição. Primeiramente, introduzimos esses métodos no caso da soma de *n* operadores em que o problema é o de encontrar um ponto num conjunto solução estendido e utilizar projeções sobre semi-espaços separadores contendo este conjunto. Estes semi-espaços são construídos com pontos no gráfico dos operadores e uma operação de resolvente é necessária para cada operador. Neste caso, a convergência fraca do algoritmo é obtida através de uma condição de separação suficiente. Em seguida, introduzimos um problema de inclusão envolvendo operadores cocoercivos e composições com operadores lineares limitados. Para este problema, apresentamos um algoritmo projetivo de decomposição que explora cocoercividade de forma que o algoritmo resultante envolve um passo *forward* por cada operador cocoercivo, em contraste com algoritmos prévios na família de métodos projetivos de decomposição, que têm usado apenas passos *backward* ou dois passos *forward*. A prova de convergência do último algoritmo com um passo *forward* requer alguns desvios da estrutura de prova anterior para algoritmos projetivos de decomposição.

Palavras-chave: Operadores monótonos maximais. Problema de inclusão monótona. Algoritmos proximais. Algoritmos projetivos de decomposição. Operadores cocoercivos.

RESUMO EXPANDIDO

Introdução

Existe uma variedade de algoritmos de decomposição para resolver o problema de inclusão

$$0 \in Az + Bz,$$

onde A e B são operadores monótonos maximais. Tais problemas surgem em áreas como equações diferenciais parciais, análise funcional, otimização convexa, entre outras.

Neste trabalho analisamos um esquema recente de algoritmos de decomposição baseado em projeções sobre semi-espaços que contem o conjunto de soluções primal-dual, primeiramente num contexto geral e logo no contexto mais específico envolvendo somas de composições de operadores lineares limitados e operadores cocoercivos.

Objetivos

Esta dissertação de mestrado tem os seguintes objetivos:

1. Estudar e entender os algoritmos de decomposição projetiva.
2. Analisar os conceitos matemáticos relacionados e as referências principais para o problema .
3. Entender a aplicação do método para um caso particular que envolve operadores cocoercivos.

Metodologia

Estudamos um novo *framework* de algoritmo de decomposição no espaço primal-dual de soluções que apresenta grande flexibilidade na escolha dos parâmetros. Para isso fazemos um estudo da bibliografia relevante relacionada, que está disponível na seção de referências, entendendo primeiramente os conceitos matemáticos básicos da teoria de operadores monótonos em espaços de Hilbert. Seguidamente, fazemos revisão do artigo que apresenta o caso general para entender os fundamentos por trás deste esquema.

Finalizamos com uma variação deste esquema para um caso que envolve composição com operadores lineares e operadores cocoercivos onde vemos como é aproveitada a cocoercividade.

Resultados, discussão e considerações finais

Acreditamos que o algoritmo de decomposição projetiva é um poderoso *framework* que faz diferença com métodos já existentes do tipo decomposição, devido a seu mecanismo e a flexibilidade que oferece. Ele tem tido um desenvolvimento para tratar problemas específicos usando suas propriedades particulares como no caso de operadores Lipschitz contínuos que faz dois passos *forward* ao invés de um passo *backward*, ou no caso de operadores cocoercivos que faz apenas um passo *forward*.

Também foram desenvolvidos algoritmos deste esquema que possuem recursos de bloco iterativo e assíncrono que são considerados em outros artigos. Mas para o caso de operadores cocoercivos estes recursos ainda não são implementados sendo um tópico de pesquisa.

Palavras-chave: Operadores monótonos maximais. Problema de inclusão monótona. Algoritmos proximais. Algoritmos projetivos de decomposição. Operadores cocoercivos.

ABSTRACT

We present in this work an iterative algorithm to solve the inclusion problem with the presence of cocoercive operators defined in Hilbert spaces under the projective splitting scheme. First, we introduce the projective splitting scheme in the case of the sum of n operators where the problem is posed as the one of finding a point in an extended solution set and make use of projections over separator half-spaces containing this set. These half-spaces are constructed with points in the graph of the operators and a resolvent operation is needed for each operator. In this case, the weak convergence of the algorithm is obtained via a condition of sufficient separation. Next, we introduce an inclusion problem involving cocoercive operators and compositions with bounded linear operators. For this problem is presented a projective splitting algorithm that exploits cocoercivity in such a manner that the resulting algorithm involves one forward step applied to each cocoercive operator, in contrast with prior algorithms in the projective splitting family, which have used only backward steps or two forward steps. The convergence proof of the algorithm with one forward step requires some detours from the previous proof framework for projective splitting.

Keywords: Maximal monotone operators. Monotone inclusion problem. Proximal algorithms. Projective splitting algorithms. Cocoercive operators.

CONTENTS

1 INTRODUCTION

The theory of monotone and maximal monotone operators plays a central role in nonlinear analysis. It has impacted areas such as partial differential equations, functional analysis, variational inequalities, and convex optimization. Many problems arising in these areas can be studied under the unified and general framework of monotone operator theory. Moreover, many of these problems can be reduced to the one of finding a z such that

$$0 \in Tz, \tag{1}$$

where T is a maximal monotone operator defined on a real Hilbert space $\mathcal{H}$. Problem (1) is known as the Monotone Inclusion Problem (MIP).

An example of such reduction in convex optimization is obtained via the subdifferential ∂f of the convex function $f : \mathcal{H} \to]-\infty,\infty]$, since when certain conditions are met, the problem

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

is equivalent to (1), where T is the subdifferential of f. When f is a proper, convex, and closed function, the subdifferential constitutes an example of a maximal monotone operator; and it is a fundamental tool in the analysis of non-differentiable convex functions. This relation makes a connection between these two seemingly unrelated fields.

The inclusion problem (1) is an important problem in the theory of monotone operators. It turns out that this problem can be related to one of finding a fixed point of certain associated non-expansive operator called the resolvent which is denoted by J_T. Moreover, since the set of zeros of the operator γT with $\gamma > 0$ is equal to the set of fixed points of the operator $J_{\gamma T}$, a zero of T can be approximated iteratively by suitable resolvent iterations. Such algorithms are known as proximal-point algorithms. They consider iterations of the form

$$z^{k+1} = J_{\gamma T} z^k.$$

The proximal-point algorithm in the context of maximal monotone operators can be traced back to (ROCKAFELLAR, 1976).

When considering an extension of this problem to the one of finding the zeros of a sum of two monotone operators, that is

$$0 \in T_1 z + T_2 z, \tag{2}$$

instead of using the resolvent of the operator $T_1 + T_2$, a widely applicable alternative is to devise an operator splitting algorithm. In this case, the approach is to employ the operators T_1 and T_2 in separate steps, so instead of considering the resolvent of the sum, it is considered the resolvent of each operator separately, which it is assumed to be "available". With this approach, several iterative algorithms have been proposed. The three most popular classes of operator splitting algorithms are the Douglas/Peaceman-Rachford class, the forward-backward class, and the double-backward class. Indeed, many algorithms in convex optimization and monotone inclusions are, in fact, applications of one of these underlying techniques to a reduced monotone inclusion in an appropriately defined product space.

These three operator splitting techniques are, in turn, a special case of the Krasnoselskii-Mann (KM) iteration for finding a fixed point of a nonexpansive operator (KRASNOSEL'SKII, 1955; MANN, 1953).

Splitting algorithms have been studied since 1970, where the case $n = 2$ predominates. Even when considering the sum of $n \geq 2$ operators, many of the existing algorithms reduce it to the case of the sum of two operators, by posing the problem in the product space $\mathcal{H}^n$.

A different class of operator splitting algorithms was introduced in (ECKSTEIN; SVAITER, 2007), the *projective splitting* class. This class has a different convergence mechanism based on projection onto separating sets and, in general, does not reduce to the KM iteration. The proposed scheme shows a relation between the solutions to the inclusion problem and the problem of finding a point belonging to a certain "extended solution set."

In each iteration, the operation applied to each operator T_1 and T_2, was a *resolvent operation*, which consists of evaluating the resolvent operator $(I + \rho T_i)^{-1}$ for some $\rho > 0$. This is known as a "backward step." The proximal parameter ρ is allowed to vary from iteration to iteration, and even from operator to operator.

This method was generalized later to the case $n \geq 2$ of (2) in (ECKSTEIN; SVAITER, 2009), that is, it was considered the problem

$$0 \in T_1 z + T_2 z + \cdots + T_n z, \tag{3}$$

where an error criterion in the evaluation of resolvent was included. In those formulations was applied the resolvent to each operator T_i separately, obtaining in this way a splitting algorithm. As explained in (JOHNSTONE; ECKSTEIN, 2020) the root ideas of the projective splitting scheme can be found in the references (IUSEM; SVAITER, 1997), (SOLODOV; SVAITER, 1999b) and (SOLODOV; SVAITER, 1999c).

Projective splitting algorithms work by performing separate calculations on each individual operator to construct a separating hyperplane between the current iterate and the problem's Kuhn–Tucker set S (essentially the set of primal and dual solutions), and then projecting onto this hyperplane. In each iteration k, it is constructed an affine "separator" function φ_k for which $\varphi_k(p) \leq 0$ for every $p \in S$. The next iterate p^{k+1} is then obtained by projecting the current iterate p^k onto the half-space defined by $\varphi_k(p) \leq 0$, possibly with some over-relaxation or under-relaxation. The crucial part of the projective splitting scheme is how φ_k is obtained, its calculation relies on getting points in the graph of each operator. Since some calculations are performed on each operator T_i separately, the procedures are indeed operator splitting algorithms.

Convergence in this algorithm is achieved via a “sufficient separation” condition. Since it may be that the separator φ_k might separate the current iterate p^k from the S in a shallow way or even in a way that does not separate at all.

Projective splitting methods were generalized to cover compositions of maximal monotone operators with bounded linear maps in (ALOTAIBI; COMBETTES; SHAHZAD, 2014), that is, problems of the form

$$0 \in Az + L^* BLz.$$

In prior projective splitting algorithms, the only operation performed on the individual operators T_i is a proximal step or a backward step. Resolvent operations remained the only way to process individual operators. The algorithm in (JOHNSTONE; ECKSTEIN, 2020) was the first to construct projective splitting separators by applying calculations other than resolvent steps to the operators T_i. This paper considers the inclusion problem

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

where a subset of the operators $\{T_1, \dots, T_n\}$ is Lipschitz continuous and $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ is a bounded linear operator for each $i = 1, \dots, n$.

In the context of convex optimization, problem (4) under appropriate constraint qualification is equivalent to the minimization problem

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

where the functions f_i are convex and some of them are also differentiable with Lipschitz-continuous gradients, and the G_i are linear and bounded operators in an appropriate Hilbert space. Minimization problems like this arise in a host of applications such as machine learning, signal and image processing, inverse problems, and computer vision, some examples can be found in (BOYD et al., 2011; COMBETTES; PESQUET, 2009; COMBETTES; WAJS, 2005).

When considering problem (4) under the projective splitting scheme, it was developed a procedure that could instead use two “forward” (explicit or gradient) steps for operators T_i that are Lipschitz continuous equivalent to applying $I - \rho T_i$, which are computationally easier than backward steps. Each step size must be bounded by the inverse of the Lipschitz constant of T_i. The algorithm developed in (JOHNSTONE; ECKSTEIN, 2020) also presented a block-iterative operation, meaning that only a subset of the operators making up the problem need to be considered at each iteration.

As mentioned, the projective splitting scheme exploited the presence of Lipschitz continuous operators to perform *two* forward steps on them. This result raised the question: can projective splitting further exploit the presence of *cocoercive* operators?

The answer to this question is “yes”, and it is developed in (JOHNSTONE; ECKSTEIN, 2021). This paper considers the problem

$$0 \in \sum_{i=1}^{n} G_i^*(A_i + B_i)G_i z, \tag{5}$$

where A_i and B_i are maximal monotone for $i = 1, \dots, n$, each B_i is L_i^{-1}–cocoercive, and $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ is linear and bounded. Problem (5) could be derived from the optimization problem

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

where the functions $f_i : \mathcal{H}_i \to]-\infty, \infty]$ and $h_i : \mathcal{H}_i \to]-\infty, \infty]$ are closed, proper and convex and every h_i is also differentiable with L_i-Lipschitz continuous gradients. Taking A_i as the subdifferential of f_i, B_i as the gradient of h_i, and under some constraint qualification conditions, this optimization problem is equivalent to (5).

Just as in (JOHNSTONE; ECKSTEIN, 2020), the presence of cocoercive operators was exploited to obtain a projective splitting algorithm that performs *one* forward step while processing each cocoercive operator. Hence, the algorithm presented in (JOHNSTONE; ECKSTEIN, 2021) requires one forward step on B_i, and one resolvent for A_i at each iteration.

Cocoercivity is in general a stronger property than Lipschitz continuity. However, in the case $B_i = \nabla h_i$ above, the Baillon-Haddad theorem (BAILLON; HADDAD, 1977) establishes that ∇h_i is L_i-Lipschitz continuous if and only if it is L_i^{-1}–cocoercive, so the two properties are equivalent.

This algorithm has a different mechanism of convergence than that in (ECKSTEIN; SVAITER, 2009). Instead of having a condition of sufficient separation, the convergence is obtained via an “ascent lemma” that relates the values $\varphi_k(p^k)$ and $\varphi_{k-1}(p^{k-1})$ in such a way that overall convergence may still be proved.

STRUCTURE OF THIS WORK

This work is based mainly on the papers (ECKSTEIN; SVAITER, 2009) and (JOHNSTONE; ECKSTEIN, 2021).

Chapter 2 introduces all necessary background, notations, definitions, and results that will be necessary for the work developed in the following chapters. An important result is that contained in Theorem 2.3, together with Lemma 2.2, since they give the tools to prove the convergence of the algorithms presented in this work.

Chapter 3 is based on (ECKSTEIN; SVAITER, 2009), it begins with a generic linear separator-projection method in Algorithm 1 for finding a point in a closed and convex set that produces a Fejér monotone sequence. Then we show how to frame the inclusion problem (3) into this generic method. Proposition 3.4 shows a condition of “sufficient separation” to guarantee weak convergence for the generated sequence. Algorithm 3 gathers the analysis previously presented, together with a relative error criterion. Finally, Theorem 3.1 shows that the resulting algorithm produces a weakly convergent sequence.

Chapter 4 is based on (JOHNSTONE; ECKSTEIN, 2021), it deals with the inclusion problem (5). As in the previous chapter, it begins framing the inclusion problem into the generic linear separator-projection method. Then it is presented how to exploit the presence of cocoercive operators to perform one forward step for those operators. The convergence analysis diverges from that in Chapter 3, it relies on an “ascent” lemma (Lemma 4.13) regarding the separators generated by the algorithm. Finally, Theorem 4.1 condenses the convergence results obtained for Algorithm 4.

2 PRELIMINARIES

This chapter contains a small portion of the theory of maximal monotone operators, projections, and convex optimization. We aim to establish all needed concepts and results on which the development of Chapters 3 and 4 relies. Moreover, this also serves to settle the notation that will be used. The main reference for the content of this chapter is (BAUSCHKE; COMBETTES, 2017), unless otherwise specified.

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot , \cdot \rangle$, the associated norm is denoted as $\| \cdot \|$, where $\|x\| = \sqrt{\langle x, x\rangle}$, and the associated distance d is defined as $d(x,y) = \|x-y\|$. The distance from a point x to a set C is denoted by $d_C(x)$. Throughout this chapter, $\mathcal{K}$ is a real Hilbert space. We denote the Hilbert direct sum of the two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ as $\mathcal{H} \oplus \mathcal{K}$. Given subsets C, D of $\mathcal{H}$ we define

$$C \pm D = \{x \pm y \mid x \in C,\, y \in D\}.$$

In particular for $z \in \mathcal{H}$ we have

$$C \pm z = C \pm \{z\} \text{ and } z \pm C = \{z\} \pm C.$$

For $\lambda \in \mathbb{R}$ we define

$$\lambda C = \{\lambda x \mid x \in C\}.$$

More generally, given a non-empty subset Λ of $\mathbb{R}$ we define

$$\Lambda C = \bigcup_{\lambda \in \Lambda} \lambda C.$$

In particular, we say that a subset C is a *cone* if $C = \mathbb{R}_{++}C$, where $\mathbb{R}_{++} = \{\lambda \in \mathbb{R} \mid \lambda > 0\}$. The *conical hull* of C is the intersection of all the cones in $\mathcal{H}$ containing C, i.e., the smallest cone in $\mathcal{H}$ containing C. It is denoted by cone C, and we have that

$$\operatorname{cone} C = \mathbb{R}_{++}C.$$

The intersection of all the linear subspaces of $\mathcal{H}$ containing C, i.e., the smallest linear subspace of $\mathcal{H}$ containing C, is called the *span* of C and is denoted by $\operatorname{span} C$; its closure is the smallest closed linear subspace of $\mathcal{H}$ containing C and it is denoted by $\overline{\operatorname{span}} C$. The *interior* of a set is denoted by $\operatorname{int} C$. Now we introduce the fundamental notion of the convexity of a set.

Definition 2.1. A subset C of $\mathcal{H}$ is *convex* if for all $\alpha \in [0,1]$ we have $C = \alpha C + (1-\alpha)C$ or, equivalently

$$\alpha x + (1-\alpha)y \in C,\ \forall\, x,y \in C.$$

With these definitions stated, we now define several weaker notions of interiority for convex sets.

Definition 2.2. Let $C \subset \mathcal{H}$ be convex. The *core* of C is

$$\operatorname{core} C = \{x \in C \mid \operatorname{cone}(C - x) = \mathcal{H}\};$$

the *strong relative interior* of C is

$$\operatorname{sri} C = \{x \in C \mid \operatorname{cone}(C - x) = \overline{\operatorname{span}}(C - x)\};$$

the *relative interior* of C is

$$\operatorname{ri} C = \{x \in C \mid \operatorname{cone}(C - x) = \operatorname{span}(C - x)\}$$

We have the inclusions

$$\operatorname{int} C \subset \operatorname{core} C \subset \operatorname{sri} C \subset \operatorname{ri} C \subset C.$$

The following example makes a distinction between the concepts of core and strong relative interior of a set.

Example 2.1. Let C be a proper closed linear subspace of $\mathcal{H}$. Then $\operatorname{core} C = \emptyset$ and $\operatorname{sri} C = C$.

The *orthogonal complement* of a subset $C \subset \mathcal{H}$, denoted by $C^{\perp}$ is

$$C^{\perp} = \{u \in \mathcal{H} \mid \langle x, u \rangle = 0, \ \forall\, x \in C\}. \tag{6}$$

The orthogonal complement is always a closed subspace of $\mathcal{H}$.

Now that we have established some notions relative to subsets oh Hilbert space, we focus our attention on an important class of extended real-valued functions. We begin with some definitions before introducing them. Let $f : \mathcal{H} \to [-\infty, \infty]$, the *domain* of f is

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

The function f is *proper* if $-\infty \notin f(\mathcal{H})$ and $\operatorname{dom} f \neq \emptyset$. The function f is *convex* if

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha) f(y), \ \forall\, \alpha \in [0,1] \text{ and } \forall\, x,y \in \mathcal{H}.$$

A function $f : \mathcal{H} \to [-\infty, \infty]$ is said to be (sequentially) *lower semicontinuous* (or *closed*) if, for every sequence $(x^k)_{k \in \mathbb{N}}$ we have

$$x^k \to x \Rightarrow f(x) \leq \liminf_{k \to \infty} f(x^k).$$

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

Definition 2.3. Let $f : \mathcal{H} \to]-\infty, \infty]$ be a proper function. The *subdifferential* of f is the set-valued operator

$$\partial f : \mathcal{H} \rightrightarrows \mathcal{H} : x \mapsto \{u \in \mathcal{H} \mid \langle y - x, u \rangle + f(x) \leq f(y) \ \forall\, y \in \mathcal{H}\}.$$

We say that f is *subdifferentiable* at $x \in \mathcal{H}$ if $\partial f(x) \neq \emptyset$. In the case of f being differentiable a x we obtain that $\partial f(x) = \{\nabla f(x)\}$.

Definition 2.4. Let $f \in \Gamma_0(\mathcal{H})$, let $\gamma > 0$. Then $\text{Prox}_{\gamma f}$ is the unique point that satisfies

$$\text{Prox}_{\gamma f}(x) = \underset{y \in \mathcal{H}}{\text{argmin}} \left(f(y) + \frac{1}{2\gamma}\|x - y\|^2 \right).$$

The operator $\text{Prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$ is the *proximal operator* of γf. This operator is related to an important concept that will be explored in Section 2.3.

2.1 OPERATORS

Definition 2.5. A *set-valued operator* denoted by $T : \mathcal{H} \rightrightarrows \mathcal{K}$ maps every point $x \in \mathcal{H}$ to a set $Tx \subset \mathcal{K}$. Then T is characterized by its graph

$$\text{gra}\, T = \{(x,u) \in \mathcal{H} \times \mathcal{K} \mid u \in Tx\}.$$

The domain and the range of T are

$$\text{dom}\, T = \{x \in \mathcal{H} \mid Tx \neq \emptyset\}, \text{ and } \text{ran}\, T = \{y \in \mathcal{K} \mid y \in Tx \text{ for some } x \in \mathcal{H}\},$$

respectively. The inverse of T, denoted by T^{-1}, is defined through its graph

$$\text{gra}(T^{-1}) := \{(u,x) \in \mathcal{K} \times \mathcal{H} \mid (x,u) \in \text{gra}\, T\}.$$

The set of zeros of T is

$$\text{zer}\, T = T^{-1}(0) = \{x \in \mathcal{H} \mid 0 \in Tx\}.$$

We say that the operator T is affine if

$$T(\lambda x + (1-\lambda)y) = \lambda Tx + (1-\lambda)Ty \ \forall x,y \in \mathcal{H}, \forall \lambda \in \mathbb{R}.$$

If $A : \mathcal{H} \rightrightarrows \mathcal{K}$ and $B : \mathcal{H} \rightrightarrows \mathcal{K}$ are set-valued operators, then for $\lambda \in \mathbb{R}$ the set-valued operator $A + \lambda B$ has the graph

$$\text{gra}(A + \lambda B) = \{(x, u + \lambda v) \mid u \in Ax,\ v \in Bx\}.$$

If an operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ with a non-empty domain is such that Tx is a singleton for all $x \in \text{dom}\, T$, then we say that T is at most single-valued, and we instead write $T : \text{dom}\, T \to \mathcal{H}$.

An important class of single-valued operators is the set of linear operators between the Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$. The set of linear and bounded (continuous) operators $T : \mathcal{H} \to \mathcal{K}$ is denoted by $\mathcal{B}(\mathcal{H}, \mathcal{K})$. In particular, the identity operator is denoted by I. When $\mathcal{K} = \mathbb{R}$ those operators are called *functionals*. The *kernel* of an operator $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, denoted by $\ker T$ is the closed subspace $\ker T = \{x \in \mathcal{H} \mid Tx = 0\}$.

Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, then the *adjoint* of L is the unique operator $L^* \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ that satisfies

$$\langle Lx, y\rangle = \langle x, L^* y\rangle.$$

The following example computes the adjoint of an operator that will be important in the following chapters.

Example 2.2. Let $\mathcal{H}_i$ be a real Hilbert space for $i = 0, \dots, n$. Let $L_i \in \mathcal{B}(\mathcal{H}_0,\mathcal{H}_i)$ for $i = 1, \dots, n$. Define $T : \mathcal{H}_0 \to \mathcal{H}_1 \times \cdots \times \mathcal{H}_n : x \mapsto (L_1 x, \dots, L_n x)$. Then $T^*(y_1, \dots, y_n) = \sum_{i=1}^n L_i^* y_i$.

Proof. Take $x \in \mathcal{H}_0$, then

$$\begin{aligned}
\langle Tx,(y_1, \dots, y_n)\rangle &= \langle (L_1 x, \dots, L_n x),(y_1, \dots, y_n)\rangle \\
&= \sum_{i=1}^{n} \langle L_i x, y_i\rangle = \sum_{i=1}^{n} \langle x, L_i^* y_i\rangle \\
&= \langle x, \sum_{i=1}^{n} L_i^* y_i)\rangle \\
&= \langle x, T^*(y_1, \dots, y_n)\rangle.
\end{aligned}$$

□

The following example computes the orthogonal complement of a linear subspace that we will appear later.

Example 2.3. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Set $\mathcal{G} = \mathcal{H} \oplus \mathcal{K}$ and $V = \{(x,y) \in \mathcal{G} \mid Lx = y\}$. Then $V^\perp = \{(z,w) \in \mathcal{G} \mid z = -L^* w\}$.

Proof. Let $(z,w) \in \mathcal{G}$. According to the definition of the orthogonal complement in (6), $(z,w) \in V^\perp$ if for all $(x,y) \in V$ we have

$$\begin{aligned}
\langle (x,y), (z,w)\rangle &= 0 \\
\langle x,z\rangle + \langle Lx,w\rangle &= 0 \\
\langle x,z\rangle + \langle x, L^* w\rangle &= 0 \\
\langle x, z + L^* w\rangle &= 0.
\end{aligned}$$

Hence, $(z,w) \in V^\perp$ must satisfy $z + L^* w = 0$. □

The following propositions contain useful and interesting properties about a bounded linear operator and its adjoint.

Proposition 2.1. *Let $\mathcal{K}$ a Hilbert space, let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and let* $\ker T = \{x \in \mathcal{H} \mid Tx = 0\}$. *Then the following hold:*

1. $T^{**} = T$.
2. $(\ker T)^{\perp} = \overline{ranT^*}$.
3. $(\operatorname{ran} T)^{\perp} = \ker T^*$.
4. $kerT^*T = kerT$ *and* $\overline{\operatorname{ran}} TT^* = ranT$.

Proposition 2.2. *Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then* $\operatorname{ran} T$ *is closed* $\Leftrightarrow$ $\operatorname{ran} T^*$ *is closed* $\Leftrightarrow$ $\operatorname{ran} TT^*$ *is closed* $\Leftrightarrow$ $\operatorname{ran} T^*T$ *is closed.*

2.2 WEAK CONVERGENCE AND FEJÉR MONOTONE SEQUENCES

Given that we have defined a distance d in the Hilbert space $\mathcal{H}$, we say that a sequence $(x^k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ converges *strongly* to point x if $\|x^k - x\| \to 0$; in symbols, $x^k \to x$. In addition to strong convergence, the weak convergence of sequences can be introduced. Before formally defining it, we recall some concepts.

Let $u \in \mathcal{H} \setminus \{0\}$ and $\eta \in \mathbb{R}$. A *closed hyperplane* in $\mathcal{H}$ is a set of the form

$$\{x \in \mathcal{H} \mid \langle x, u\rangle = \eta\}$$

Moreover, a *closed half-space* with *outer normal* u is a set of the form

$$\{x \in \mathcal{H} \mid \langle x, u\rangle \leq \eta\},$$

and an *open half-space* with *outer normal* u is a set of the form

$$\{x \in \mathcal{H} \mid \langle x, u\rangle < \eta\}.$$

A sequence $(x^k)_{k\in\mathbb{N}}$ converges *weakly* to $x \in \mathcal{H}$ if for every $u \in \mathcal{H}$ we have

$$\langle x^k, u\rangle \to \langle x, u\rangle.$$

We denote this weak convergence by $x^k \rightharpoonup x$. Geometrically this means that the distance between the sequence $(x^k)_{k\in\mathbb{N}}$ and any closed hyperplane containing x converges to zero. We say that an operator $T : D \subset \mathcal{H} \to \mathcal{K}$ is *weakly continuous* if for every sequence $(x^k)_{k\in\mathbb{N}}$ such that $x^k \rightharpoonup x \in D$ we have $Tx^k \rightharpoonup Tx$. The following lemmas establish some conditions regarding the weak convergence of sequences.

Lemma 2.1. *Let $(x^k)_{k\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$. Then $(x^k)_{k\in\mathbb{N}}$ possesses a weakly convergent subsequence.*

Lemma 2.2 (Opial)**.** *Let $(x^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let C be a nonempty subset of $\mathcal{H}$. Suppose that, for every $x \in C$, $(\|x^k - x\|)_{k\in\mathbb{N}}$ converges and that every weak sequential cluster point of $(x^k)_{k\in\mathbb{N}}$ belongs to C. Then $(x^k)_{k\in\mathbb{N}}$ converges weakly to a point in C.*

Now we define Fejér monotone sequences. We will see that the algorithms presented in the following chapters generate Fejér sequences. Their convergence relies heavily upon Lemma 2.2, as we shall see.

Definition 2.6. Let S be a nonempty closed and convex set in a real Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$. A sequence $(x^k)_{k\geq 0}$ of points in $\mathcal{H}$ is said to be *Fejér monotone* with respect to S if

$$\|x^{k+1} - x\| \leq \|x^k - x\|, \forall\, x \in S, \forall\, k \geq 0.$$

In words, each point in the sequence is not further from any point in S than its predecessor. Basic properties of these sequences are stated in the following proposition.

Proposition 2.3. *Let $(x^k)_{k\geq 0}$ a Fejér monotone sequence with respect to S, then the following hold:*

1. *The sequence $(x^k)_{k\geq 0}$ is bounded.*
2. *For all $x \in S$ the sequence $(\|x^k - x\|)_{k\geq 0}$ converges.*
3. *The sequence $(d_S(x^k))_{k\geq 0}$ is nonincreasing.*

In general, Fejér monotone sequences do not converge, not even weakly. However, by Proposition 2.3(1) and Lemma 2.1, the set of weak limit points of a Fejér monotone sequence is non-empty. Additionally, from Proposition 2.3(2) and Lemma 2.2, we conclude that it is sufficient for the weak convergence of the Fejér monotone sequence $(x^k)_{k\geq 0}$ that their weak limit points belong to S.

2.3 MAXIMAL MONOTONE OPERATORS AND THE RESOLVENT

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

$$\langle u - v, x - y\rangle \geq 0\ \forall\, u \in Tx \text{ and } \forall\, v \in Ty.$$

Example 2.4. The following are examples of monotone operators.

1. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{K} \rightrightarrows \mathcal{K}$ be monotone operators, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ and let $\gamma \geq 0$. Then the operators A^{-1}, γA and $A + L^* BL$ are monotone.
2. Let $f : \mathcal{H} \to]-\infty, \infty]$ proper, then ∂f is monotone.

Definition 2.7. The monotone operator T is called *maximal monotone* (or *maximally monotone*) if its graph is not contained properly in the graph of any other monotone operator $S : \mathcal{H} \rightrightarrows \mathcal{H}$.

For every monotone operator there exists a maximally monotone extension containing its graph.

Example 2.5. The following are examples of maximal monotone operators.

1. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{K} \rightrightarrows \mathcal{K}$ be maximal monotone operators. Then A^{-1} and the operator $T : \mathcal{H} \oplus \mathcal{K} \rightrightarrows \mathcal{H} \oplus \mathcal{K} : (x,y) \mapsto Ax \times By$ are maximal monotone.
2. Let $f \in \Gamma_0(\mathcal{H})$, then ∂f is maximal monotone.
3. Let $A : \mathcal{H} \to \mathcal{H}$ be monotone and continuous. Then A is maximal monotone.
4. Let $A \in \mathcal{B}(\mathcal{H})$ be such that $A^* = -A$. Then A is maximal monotone.

One important property of maximal monotone operators is the following:

Proposition 2.4. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ maximal monotone, and let $x \in \mathcal{H}$. Then Tx is closed and convex.*

A notable consequence of this proposition is that for a maximal monotone operator T we have that the set

$$\operatorname{zer} T = T^{-1}(0)$$

is closed and convex, since as seen in Example 2.5(1) the operator T^{-1} is maximal monotone as well.

Notice that as seen in Example 2.4(1) by taking $L = I$, the sum of two monotone operators is monotone. However, it remains the question of whether this sum is additionally maximal. It turns out that some algebraic conditions are needed to ensure this.

Proposition 2.5. *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{K} \rightrightarrows \mathcal{K}$ maximal monotone, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and suppose that*

$$\operatorname{cone}(\operatorname{dom} B - L(\operatorname{dom} A)) = \overline{\operatorname{span}}(\operatorname{dom} B - L(\operatorname{dom} A)). \tag{7}$$

Then $A + L^*BL$ is maximal monotone. In the case of $L = I$, condition in (7) reduces to

$$\operatorname{cone}(\operatorname{dom} B - \operatorname{dom} A) = \overline{\operatorname{span}}(\operatorname{dom} B - \operatorname{dom} A), \tag{8}$$

from where $A + B$ is maximal monotone.

The following proposition, which a proof is found in (BRICEÑO-ARIAS; COMBETTES, 2011) shows an example of a maximal monotone operator constructed via the operations in Example 2.5.

Proposition 2.6. *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{K} \rightrightarrows \mathcal{K}$ be maximal monotone operators. Let $L : \mathcal{H} \to \mathcal{K}$ a linear bounded operator. Define on $\mathcal{G} = \mathcal{H} \oplus \mathcal{K}$ the operators*

$$\boldsymbol{M} : \mathcal{G} \rightrightarrows \mathcal{G} : (x,v) \mapsto Ax \times B^{-1}v \text{ and } \boldsymbol{S} : \mathcal{G} \to \mathcal{G} : (x,v) \mapsto (L^*v, -Lx).$$

Then the following hold:

1. ***M** is maximal monotone;*
2. $\boldsymbol{S} \in \mathcal{B}(\mathcal{G})$ *and* $\boldsymbol{S}^* = -\boldsymbol{S}$*;*
3. $\boldsymbol{M} + \boldsymbol{S}$ *is maximal monotone.*

Proof. 1. Follows from the definition of $\boldsymbol{M}$ and Example 2.5(1).

2. Clearly $\boldsymbol{S}$ is linear and

$$\begin{aligned}\langle \boldsymbol{S}(x,v),(y,w)\rangle &= \langle (L^*v,-Lx),(y,w)\rangle = \langle L^*v,y\rangle + \langle -Lx,w\rangle \\ &= \langle v,Ly\rangle + \langle x,-L^*w\rangle = \langle (x,v),(-L^*w,Ly)\rangle \\ &= \langle (x,v),-\boldsymbol{S}(y,w)\rangle,\end{aligned}$$

that is, $\boldsymbol{S}^* = -\boldsymbol{S}$. Additionally, we have

$$\begin{aligned}\|\boldsymbol{S}(x,v)\|^2 &= \|(L^*v,-Lx)\|^2 = \|L^*v\|^2 + \|Lx\|^2 \\ &\leq \|L\|^2(\|v\|^2 + \|x\|^2) \\ &= \|L\|^2\|(x,v)\|^2.\end{aligned}$$

Thus, $\boldsymbol{S} \in \mathcal{B}(\mathcal{G})$.

3. From item 4 of Example 2.5 it follows that $\boldsymbol{S}$ is maximal monotone. In addition, since $\operatorname{dom} \boldsymbol{S} = \mathcal{G}$ we conclude from Proposition 2.5 that the sum $\boldsymbol{M} + \boldsymbol{S}$ is maximal monotone.

□

The importance of this proposition is that since $\boldsymbol{M} + \boldsymbol{S}$ is maximal monotone we can conclude following Proposition 2.4 that $\operatorname{zer}(\boldsymbol{M} + \boldsymbol{S})$ is closed and convex. We will express a certain important set as the zeros of an operator of the form $\boldsymbol{M} + \boldsymbol{S}$.

In (MINTY, 1962) we see an important characterization of a maximal monotone operator.

Theorem 2.1 (Minty's Theorem)**.** *Let* $T : \mathcal{H} \rightrightarrows \mathcal{H}$ *be monotone. Then* T *is maximally monotone if and only if* $\operatorname{ran}(I + T) = \mathcal{H}$.

An implication of this Theorem is that for a maximal monotone operator T and for $z \in \mathcal{H}$ there is a unique $(x,y) \in \operatorname{gra}(T)$ such that

$$x + y = z \Rightarrow z \in (I + T)x \Rightarrow x \in (I + T)^{-1}z.$$

Therefore, we can define the resolvent, proximal mapping, or proximal operator of T, denoted by J_T as

$$J_T = (I + T)^{-1}.$$

Notice that if $\rho > 0$, then the resolvent of ρT satisfies

$$x = J_{\rho T}(t) \Leftrightarrow x + \rho u = t \text{ and } u \in Tx, \tag{9}$$

with the x and u satisfying this relation being unique. In view of Minty's Theorem, the proximal mapping (of a maximal monotone operator) is a function whose domain is the whole underlying Hilbert space.

As we saw in Example 2.5(2) the subdifferential of a function $f \in \Gamma_0(\mathcal{H})$ is maximal monotone, and accordingly its resolvent is well defined. It is interesting that in addition to this we have the following relation

$$J_{\partial f} = (I + \partial f)^{-1} = \mathrm{Prox}_f\,.$$

2.4 PROJECTIONS

This section defines projections over closed and convex subsets. More specifically, it deals with projections over affine subspaces. We begin with the concept of best approximation. Let C be a nonempty subset of $\mathcal{H}$, let $x \in \mathcal{H}$, and let $p \in C$. Then p is a *best approximation* to x from C (or a projection of x onto C) if $\|x - p\| = d_C(x)$. In other words,

$$\|x - p\| \leq \|x - q\| \; \forall\, q \in C.$$

As it is, the best approximation to x from C could not even exist, or it could be more than two approximations. The following theorem gives conditions over C to ensure a well-defined projection. Additionally, it provides a characterization of the projection mapping.

Theorem 2.2 (Projection Theorem)**.** *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then for every $x \in \mathcal{H}$ and every p in $\mathcal{H}$,*

$$p = P_C x \Leftrightarrow p \in C \text{ and } \langle y - p, x - p\rangle \leq 0, \forall\, y \in C. \tag{10}$$

Therefore, closed and convex sets allow us to define a projection operator with no ambiguity. In this case, the *projector* onto C is the operator, denoted by P_C, that maps every point in $\mathcal{H}$ to its unique projection onto C. The projector P_C is an example of a maximal monotone operator.

The following example details the projection over a hyperplane.

Example 2.6. Suppose that $u \in \mathcal{H} \setminus \{0\}$, let $\eta \in \mathbb{R}$, and set $C = \{x \in \mathcal{H} \mid \langle x, u\rangle = \eta\}$. Then for $x \in \mathcal{H}$

$$P_C(x) = x + \frac{\eta - \langle x, u\rangle}{\|u\|^2} u. \tag{11}$$

Proof. To prove it, we apply (10) in Theorem 2.2. Set

$$p = x + \frac{\eta - \langle x,u\rangle}{\|u\|^2} u.$$

Then

$$\begin{aligned}\langle p,u\rangle &= \left\langle x + \frac{\eta - \langle x,u\rangle}{\|u\|^2} u, u\right\rangle \\ &= \langle x,u\rangle + \left(\frac{\eta - \langle x,u\rangle}{\|u\|^2}\right)\langle u,u\rangle \\ &= \eta,\end{aligned}$$

that is, $p \in C$. Next, take arbitrary $y,z \in C$ and compute

$$\begin{aligned}\langle y-z, x-p\rangle &= \left\langle y-z, x-x-\left(\frac{\eta - \langle x,u\rangle}{\|u\|^2}\right) u\right\rangle \\ &= \left(\frac{\langle x,u\rangle - \eta}{\|u\|^2}\right)(\langle y,u\rangle - \langle z,u\rangle) \\ &= \left(\frac{\langle x,u\rangle - \eta}{\|u\|^2}\right)(\eta - \eta) \\ &= 0.\end{aligned}$$

Therefore, $P_C x = p$. □

In what follows, we will focus on projections over linear subspaces. We begin with some general properties.

Proposition 2.7. *Let V be a closed linear subspace of $\mathcal{H}$ and let $x \in \mathcal{H}$. Then the following hold:*

1. *Let $p \in \mathcal{H}$, then $p = P_V x$ if and only if $(p, x-p) \in V \times V^{\perp}$.*
2. *$P_V \in \mathcal{B}(\mathcal{H})$, $\|P_V\| = 1$ if $V \neq \{0\}$, and $\|P_V\| = 0$ if $V = \{0\}$.*
3. *$P_{V^{\perp}} = I - P_V$.*

Now we study the projection over the range of an operator $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$. We say that x is a *least square solution* to the problem $Tz = y$ if

$$\|Tx - y\| = \min_{u\in\mathcal{H}} \|Tu - y\|.$$

Notice that the right-hand side is the best approximation problem over the set ran T, which is convex. Furthermore, it is a linear subspace. Provided that ran T is closed we have according to Theorem 2.2 that the projection over ran T is well-defined and for $x \in \mathcal{H}$

$$\begin{aligned}\|Tx - y\| &\leq \|Tu - y\| \,\forall\, u \in \mathcal{H} \\ &\leq \|r - y\| \,\forall\, r \in \operatorname{ran} T,\end{aligned}$$

that is,

$$Tx = P_{\text{ran } T} y. \tag{12}$$

Additionally, since ran T is a linear subspace we deduce from Proposition 2.7(1) that $y - Tx \in (\text{ran } T)^{\perp}$, hence

$$\begin{aligned} \langle v, y - Tx \rangle &= 0 \quad \forall\, v \in \text{ran } T \\ \langle Tu, y - Tx \rangle &= 0 \quad \forall\, u \in \mathcal{H} \\ \langle u, T^*(y - Tx) \rangle &= 0 \end{aligned}$$

which yields that $T^*y = T^*Tx$. Hence, we had proven that the set of least square solutions is

$$\{x \in \mathcal{H} \mid T^*y = T^*Tx\}.$$

If $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ is such that T^*T is invertible, we have as unique solution of the least square problem: $x = (T^*T)^{-1}T^*y$. Furthermore, we conclude by (12) that

$$P_{\text{ran } T} y = Tx = T(T^*T)^{-1}T^*y. \tag{13}$$

With this in mind, we will obtain the projection over two important subspaces. The first one is a direct consequence of (13).

Proposition 2.8. *Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{H}^n)$ defined by $L : z \mapsto (z, \ldots, z)$, then*

$$P_{\text{ran } L}(w_1, \ldots, w_n) = \frac{1}{n}\left(\sum_{i=1}^{n} w_i, \ldots, \sum_{i=1}^{n} w_i\right)$$

Proof. Notice that the operator L is an instance of Example 2.2, with $G_i = I$ for $i = 1 \ldots, n$. Hence $L^*(w_1, \ldots, w_n) = \sum_{i=1}^{n} w_i$, from which we easily obtain $L^*L = nI$. From (13), it follows that

$$\begin{aligned} P_{\text{ran } L}(w_1, \ldots, w_n) &= L(L^*L)^{-1}L^*(w_1, \ldots, w_n) \\ &= L\left(\frac{1}{n}\sum_{i=1}^{n} w_i\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} w_i, \ldots, \sum_{i=1}^{n} w_i\right). \end{aligned}$$

□

The second subspace is a little more elaborated.

Proposition 2.9. *Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Set $\mathcal{G} = \mathcal{H} \oplus \mathcal{K}$ and $V = \{(x,y) \in \mathcal{G} \mid Lx = y\}$. Then, for every $(x,y) \in \mathcal{G}$, the following hold:*

$$P_V(x,y) = (x - L^*(I + LL^*)^{-1}(Lx - y), y + (I + LL^*)^{-1}(Lx - y)). \tag{14}$$

Proof. Define $T : \mathcal{G} \to \mathcal{K} : (z, w) \mapsto Lz - w$. Hence, $V = \ker T$ and its orthogonal complement is given by

$$V^\perp = \{(z,w) \in \mathcal{G} \mid z = -L^* w\}. \tag{15}$$

Notice that

$$\langle T(z,w),v\rangle = \langle Lz - w,v\rangle = \langle (z,w),(L^* v, - v)\rangle,$$

that is, $T^*(v) = (L^* v, - v)$ so we obtain that

$$TT^* v = T(L^* v, - v) = LL^* v + v \Rightarrow TT^* = I + LL^*.$$

This implies that TT^* is invertible and $\operatorname{ran} TT^* = \mathcal{K}$, which is closed and therefore by Proposition 2.2, $\operatorname{ran} T^*$ is closed as well.

To relate the projection over V to the projection over a range of a linear bounded operator we first notice that since $V = \ker T$ is a linear subspace we have from Proposition 2.7(3) that $P_V = I - P_{V^\perp}$ and by Proposition 2.1 $V^\perp = \overline{\operatorname{ran}} T^*$.

Since $\operatorname{ran} T^*$ is closed we can write $V^\perp = \operatorname{ran} T^*$, and from TT^* being invertible we can use (13) and the fact that $T^{**} = T$ to compute $P_{\operatorname{ran} T^*}$ as follows

$$\begin{aligned} P_{\operatorname{ran} T^*}(x,y) &= T^*(TT^*)^{-1} T(x,y) \\ &= T^*\Big((I + LL^*)^{-1}(Lx - y)\Big) \\ &= \big(L^*(I + LL^*)^{-1}(Lx - y), -(I + LL^*)^{-1}(Lx - y)\big). \end{aligned}$$

Therefore,

$$\begin{aligned} P_{\ker T}(x,y) &= I - P_{\operatorname{ran} T^*}(x,y) \\ &= (x,y) - \big(L^*(I + LL^*)^{-1}(Lx - y), -(I + LL^*)^{-1}(Lx - y)\big) \\ &= \big(x - L^*(I + LL^*)^{-1}(Lx - y), y + (I + LL^*)^{-1}(Lx - y)\big). \end{aligned}$$

□

We derive from (14) one more expression for the projector P_V and one for $P_{V^\perp}$.

Proposition 2.10. *Let* $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ *and define* $V = \{(z, w) \in \mathcal{H} \oplus \mathcal{K} \mid Lz = w\}$, *then*

1. $P_V(x,y) = ((I + L^*L)^{-1}(x + L^*y), L(I + L^*L)^{-1}(x + L^*y))$
2. $P_{V^\perp}(x,y) = (L^*(I + LL^*)^{-1}(Lx - y), -(I + LL^*)^{-1}(Lx - y))$.

Proof. The proof is based on showing that $x - L^*(I + LL^*)^{-1}(Lx - y) - (I + L^*L)^{-1}(x + L^*y)$ is on the kernel of $I + L^*L$ for all $(x,y) \in \mathcal{H} \oplus \mathcal{K}$, and since $I + LL^*$ is invertible this yields the desired equality $x - L^*(I + LL^*)^{-1}(Lx - y) = (I + L^*L)^{-1}(x + L^*y)$. The same argument is applied to the second entry. For the first entry we have

$$\begin{aligned}
&(I+L^*L)\Big(x-L^*(I+LL^*)^{-1}(Lx-y)-(I+L^*L)^{-1}(x+L^*y)\Big)\\
=\;& x+L^*Lx-(L^*(I+LL^*)^{-1}(Lx-y)+L^*LL^*(I+LL^*)^{-1}(Lx-y))-x-L^*y\\
=\;& L^*(Lx-y)-L^*(I+LL^*)(I+LL^*)^{-1}(Lx-y)\\
=\;& L^*(Lx-y-(Lx-y))\\
=\;& 0.
\end{aligned}$$

Since $I+LL^*$ is invertible, this means that

$$x-L^*(I+LL^*)^{-1}(Lx-y)=(I+L^*L)^{-1}(x+L^*y),\ \forall\,(x,y)\in\mathcal{H}\oplus\mathcal{K}.$$

With the same procedure for the second entry we have

$$\begin{aligned}
&(I+LL^*)\Big(y+(I+LL^*)^{-1}(Lx-y)-L(I+L^*L)^{-1}(x+L^*y)\Big)\\
=\;& y+LL^*y+(Lx-y)-L(I+L^*L)^{-1}(x+L^*y)+LL^*L(I+L^*L)^{-1}(x+L^*y)\\
=\;& L(x+L^*y)-L(I+L^*L)^{-1}(x+L^*y)-LL^*L(I+L^*L)^{-1}(x+L^*y)\\
=\;& (L(I+LL^*)-L-LL^*L)(I+L^*L)^{-1}(x+L^*y)\\
=\;& L(I+LL^*-I-LL^*)(I+L^*L)(x+L^*y)\\
=\;& 0.
\end{aligned}$$

Thus,

$$y+(I+LL^*)^{-1}(Lx-y)=L(I+LL^*)^{-1}(x+L^*y)\ \forall\,(x,y)\in\mathcal{H}\oplus\mathcal{K}.$$

Therefore

$$\begin{aligned}
P_V(x,y)&=\big(x-L^*(I+LL^*)^{-1}(Lx-y),y+(I+LL^*)^{-1}(Lx-y)\big)\\
&=\big((I+L^*L)^{-1}(x+L^*y),L(I+L^*L)^{-1}(x+L^*y)\big)
\end{aligned}$$

The second item follows readily from the identity $P_{V^\perp}=I-P_V$.

□

The following theorem introduced in (BAUSCHKE, 2009) will be handy later in proving the weak convergence of the generated sequences of the projective splitting algorithms.

Theorem 2.3. *Let $T:\mathcal{H}\rightrightarrows\mathcal{H}$ be maximal monotone, and let V be a closed linear subspace of $\mathcal{H}$. Let $(x^k,u^k)_{k\in\mathbb{N}}$ a sequence in* gra T *such that $(x^k,u^k)\rightharpoonup(x,u)\in\mathcal{H}\times\mathcal{H}$. Suppose that*

$$x^k-P_Vx^k\to 0 \text{ and } P_Vu^k\to 0,$$

where P_V denotes the projector onto V. Then

$$(x,u)\in(V\times V^\perp)\cap\operatorname{gra}T,\ \text{and}\ \langle x^k,u^k\rangle\to\langle x,u\rangle=0.$$

The proof is made using a firmly non-expansive operator related to the resolvent of T and the projection mapping P_C. A different proof based on the Spingarn's partial-inverse can be found in (ALVES, 2020).

2.5 COCOERCIVE OPERATORS

In this section, we define an important class of operators, namely, firmly non-expansive operators. From these, we obtain the class of cocoercive operators, they will be important in Chapter 4.

Definition 2.8. Let $D \subset \mathcal{H}$ nonempty, and let $T : D \to \mathcal{H}$. Then T is firmly non-expansive if $\forall\ x,y \in \mathcal{H}$

$$\|Tx - Ty\|^2 + \|(I-T)x - (I-T)y\|^2 \leq \|x-y\|^2. \tag{16}$$

It follows from (16) that every firmly non-expansive operator is also *nonexpansive*, that is,

$$\|Tx - Ty\| \leq \|x-y\|, \forall\ x,y \in \mathcal{H}, \tag{17}$$

in other words, T is Lipschitz continuous with constant 1. It is easy to see that (16) is equivalent to

$$\langle x-y, Tx-Ty\rangle \geq \|Tx-Ty\|^2\ \forall\ x,y \in \mathcal{H}. \tag{18}$$

A particular example of a firmly non-expansive operator is

Definition 2.9. Let $D \subset \mathcal{H}$ nonempty, and let $T : D \to \mathcal{H}$, and let $\beta > 0$. Then T is *β–cocoercive* (or β–inverse strongly monotone), if βT is firmly non-expansive, that is,

$$\langle x-y, Tx-Ty\rangle \geq \beta\|Tx-Ty\|^2, \forall\ x,y \in \mathcal{H}. \tag{19}$$

Notice that (19) follows from (18) applied to βT. It turns out that β–cocoercive operators have a desirable property as states the following

Proposition 2.11. *Let $T : \mathcal{H} \to \mathcal{H}$ be β–cocoercive, then T is maximal monotone.*

Proof. From (19) we deduce that the operator T is monotone. Additionally, since every firmly non-expansive operator is also non-expansive and therefore continuous, we conclude that a β–cocoercive operator is continuous and hence maximal by Example 2.5(3). □

The following proposition shows us that the sum of compositions of linear operators and cocoercive operators is a cocoercive operator.

Proposition 2.12. *Let $\mathcal{K}_i$ be a real Hilbert space for $i = 1,\ldots,n$. Suppose that $L_i \in \mathcal{B}(\mathcal{H},\mathcal{K}_i)\setminus\{0\}$ for $i = 1,\ldots,n$, let $\beta_i \in \mathbb{R}_{++}$, and let $T_i : \mathcal{K}_i \to \mathcal{K}_i$ be β_i–cocoercive. Set*

$$T = \sum_{i=1}^{n} L_i^* T_i L_i \ \text{ and } \ \beta = \frac{1}{\sum_{i=1}^{n} \frac{\|L_i\|}{\beta_i}}.$$

Then T is β–cocoercive.

We finish this section with two examples of firmly non-expansive operators.

Example 2.7. The following are examples of firmly non-expansive operators

1. The resolvent of a maximal monotone operator.
2. The projector P_C where C is nonempty closed convex subset of $\mathcal{H}$.

2.6 THE MINIMIZATION AND INCLUSION PROBLEMS

In this section, we will show how the problem of minimizing a convex function is related to the problem of finding zeros of a maximal monotone operator.

Proposition 2.13. *Let* $f : \mathcal{H} \to]-\infty, \infty]$ *proper. Then*

$$\operatorname{Argmin} f = \operatorname{zer} \partial f. \tag{20}$$

Now consider a problem of the form

$$\min_{z \in \mathcal{H}} f(z) + g(Lz),$$

where $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{K})$ and $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. To apply (20) we have to compute the subdifferential $\partial(f + g \circ L)$. When certain conditions are met, we can compute this subdifferential in terms of the subdifferentials of the convex functions f and g. These conditions are called constraint qualification conditions. The following theorem shows one of such conditions.

Theorem 2.4. *Let* $f \in \Gamma_0(\mathcal{H})$, *let* $g \in \Gamma_0(\mathcal{K})$, *let* $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. *Suppose that*

$$0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f)), \tag{21}$$

then $\partial(f + g \circ L) = \partial f + L^* \circ \partial g \circ L$.

The condition in Theorem 2.4 could be a consequence of the following:

1. $\operatorname{dom} g - L(\operatorname{dom} f)$ is a closed linear subspace.
2. $0 \in \operatorname{core}(\operatorname{dom} g - L(\operatorname{dom} f))$.
3. $0 \in \operatorname{int}(\operatorname{dom} g - L(\operatorname{dom} f))$.
4. $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \emptyset$ or $L(\operatorname{dom} f) \cap \operatorname{int} \operatorname{dom} g \neq \emptyset$.

Notice that when $f = 0$ in Theorem 2.4, condition (21) reduces to

$$0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{ran} L),$$

yielding $\partial(g \circ L) = L^* \circ \partial g \circ L$.

Corollary 2.5. *Let* f *and* g *functions in* $\Gamma_0(\mathcal{H})$ *such that one of the following holds:*

1. $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$.

2. $\operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g \neq \emptyset$.
3. $\operatorname{dom} g = \mathcal{H}$.
4. $\mathcal{H}$ *is finite-dimensional and* $\operatorname{ri} \operatorname{dom} f \cap \operatorname{ri} \operatorname{dom} g \neq \emptyset$.

Then $\partial(f+g) = \partial f + \partial g$.

Due to the relation in (20) when considering the problem

$$\min_{z \in \mathcal{H}} \sum_{i=1}^{n} f_i(z), \;\; f_i \in \Gamma_0(\mathcal{H}),$$

we ask under what conditions $\partial\left(\sum_{i=1}^{n} f_i\right) = \sum_{i=1}^{n} \partial f_i$. Some conditions are established in the following corollary of Theorem 2.4.

Corollary 2.6. *Let n be an integer such that* $n \geq 2$, *set* $I = \{1, \ldots, n\}$, *and let* $(f_i)_{i \in I}$ *be functions in* $\Gamma_0(\mathcal{H})$ *such that one of the following holds:*

1. *We have*
$$0 \in \bigcap_{i=2}^{n} \operatorname{sri}\left(\operatorname{dom} f_i - \bigcap_{j=1}^{i-1} \operatorname{dom} f_j\right).$$
2. *For every* $i \in \{2, \ldots, n\}$ $\operatorname{dom} f_i - \cap_{j=1}^{i-1} \operatorname{dom} f_j$ *is a closed linear subspace.*
3. *The sets* $(\operatorname{dom} f_i)_{i \in I}$ *are linear subspaces and, for every* $i \in \{2, \ldots, n\}$, $\operatorname{dom} f_i + \cap_{j=1}^{i-1} \operatorname{dom} f_j$ *is closed.*
4. $\operatorname{dom} f_n \cap \bigcap_{i=1}^{n-1} \operatorname{int} \operatorname{dom} f_i \neq \emptyset$.
5. $\mathcal{H}$ *is finite-dimensional and* $\bigcap_{i \in I} \operatorname{ri} \operatorname{dom} f_i \neq \emptyset$.

Then $\partial\left(\sum_{i=1}^{n} f_i\right) = \sum_{i=1}^{n} \partial f_i$.

The following example shows an application of these results.

Example 2.8. Consider the minimization problem

$$\min_{z \in \mathbb{R}^{p+1}} \sum_{i=1}^{2} (f_i \circ G_i + h_i \circ G_i)(z). \tag{22}$$

where the functions $f_1, h_1 : \mathbb{R}^d \to \mathbb{R}$, $f_2, h_2 : \mathbb{R}^{p+1} \to \mathbb{R}$ are defined as follows

$$\begin{aligned} f_1(t) &= \|t\|_1, & h_1(t) &= 0, \\ f_2(\gamma, \beta) &= \|\gamma\|_1, & h_2(\gamma, \beta) &= \frac{1}{2}\|\beta e + H\gamma\|_2^2, \end{aligned}$$

$e \in \mathbb{R}^d$ has all elements equal to 1, and the linear operators G_1 and G_2 are

$$G_1 = [H \,|\, 0], \text{ and } G_2 = I,$$

where $G_1 \in \mathbb{R}^{d\times(p+1)}$ with $H \in \mathbb{R}^{d\times p}$ and G_2 is the $(p+1)\times(p+1)$ identity matrix. Hence, problem (22) reduces to

$$\min_{z\in\mathbb{R}^{p+1}} [f_1 \circ G_1 + (f_2 + h_2)](z).$$

More specifically

$$\min_{\substack{\gamma\in\mathbb{R}^p\\ \beta\in\mathbb{R}}} \|H\gamma\|_1 + \|\gamma\|_1 + \frac{1}{2}\|\beta e + H\gamma\|_2^2.$$

Notice that the function h_2 is Lipschitz differentiable. Condition in Theorem 2.4 is

$$0 \in \text{sri}(\text{dom}\, f_1 - G_1(\text{dom}(f_2 + h_2)))$$

which follows from dom $f_1 = \mathbb{R}^d$. Thus we obtain

$$\partial(f_1 \circ G_1 + (f_2 + g_2)) = G_1^* \circ \partial f_1 \circ G_1 + \partial(f_2 + h_2).$$

Additionally, f_2 and h_2 satisfies one of the hypotheses in Corollary 2.5, specifically that dom $h_2 = \mathbb{R}^{p+1}$. Thus

$$\partial(f_2 + g_2) = \partial f_2 + \partial h_2.$$

Altogether, we have according to (20) that the minimization problem is equivalent to the inclusion problem

$$0 \in [G_1^* \circ \partial f \circ G_1 + \partial f_2 + \partial h_2](z).$$

This example is a simplified version of the problem in (JOHNSTONE; ECKSTEIN, 2021, Sect. 6.3).

Now that we have established a connection between the minimization problem and the inclusion problem, we explore a little more about the later. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{K} \rightrightarrows \mathcal{K}$ be maximal monotone operators, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Consider the problem

$$\text{find } z \in \mathcal{H} \text{ such that } 0 \in Az + L^*BLz, \tag{23}$$

together with the dual inclusion

$$\text{find } w \in \mathcal{K} \text{ such that } 0 \in -LA^{-1}(-L^*w) + B^{-1}w. \tag{24}$$

Consider now the operators $\boldsymbol{M}$ and $\boldsymbol{S}$ defined in Proposition 2.6 applied to the operators A, B and L. For the problem

$$\text{find } z \in \mathcal{H} \text{ such that } 0 \in \boldsymbol{M}z + \boldsymbol{S}z,$$

denote by $\mathcal{Z}$ its set of solutions

$$\begin{aligned} \mathcal{Z} &= \text{zer}(\boldsymbol{M} + \boldsymbol{S}) \\ &= \{(z, w) \in \mathcal{H} \oplus \mathcal{K} \mid -L^*w \in Az \text{ and } Lz \in B^{-1}w\}. \end{aligned}$$

We call it the set of *Kuhn - Tucker points* associated with problem (23)-(24). We will see the Kuhn - Tucker set for two inclusion problems in Chapters 3 and 4. Notice that in the case of $L = I$ we have the primal inclusion

$$0 \in Az + Bz \tag{25}$$

and the dual inclusion

$$0 \in -A^{-1}(-w) + B^{-1}w.$$

Having established a simple form of an inclusion problem, we present two well-known iterative algorithms to find a solution. These employ the operators in separate steps; this is what a splitting scheme is. We begin with the Douglas-Rachford splitting algorithm, which traces back to (DOUGLAS; RACHFORD, 1956).

Theorem 2.7 (Douglas–Rachford algorithm)**.** *Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone operators such that* $\operatorname{zer}(A+B) \neq \emptyset$*, let $(\rho_k)_{k\in\mathbb{N}}$ be a sequence in* $[0,2]$ *such that $\sum_{k\in\mathbb{N}} \rho_k(2-\rho_k) = +\infty$ and let $\gamma \in \mathbb{R}_{++}$. Let $y^0 \in \mathcal{H}$, and set for $k = 0, 1, \ldots$*

$$\begin{aligned} x^k &= J_{\gamma B}\, y^k, \\ z^k &= J_{\gamma A}(2x^k - y^k), \\ y^{k+1} &= y^k + \rho_k(z^k - x^k). \end{aligned}$$

Then there exists $y \in \mathcal{H}$ such that $y^k \rightharpoonup y$, and $x := J_{\gamma B} y$ is a solution of the primal problem (25).

The following theorem presents the Forward-Backward algorithm (FB) which considers the case when one of the operators is cocoercive.

Theorem 2.8 (Forward-Backward algorithm)**.** *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone, let $\beta \in \mathbb{R}_{++}$, let $B : \mathcal{H} \to \mathcal{H}$ be β–cocoercive, let $\gamma \in\,]0, 2\beta[$, and set $\delta = 2 - \gamma/(2\beta)$. Futhermore, let $(\rho_k)_{k\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{k\in\mathbb{N}} \rho_k(\delta - \rho_k) = +\infty$, and let $x^0 \in \mathcal{H}$. Suppose that* $\operatorname{zer}(A+B) \neq \emptyset$ *and set for $k = 0, 1, \ldots$*

$$\begin{aligned} y^k &= x^k - \gamma B x^k, \\ x^{k+1} &= x^k + \rho_k(J_{\gamma A}\, y^k - x^k). \end{aligned}$$

Then the following hold:

1. *$(x^k)_{k\in\mathbb{N}}$ converges weakly to a point in* $\operatorname{zer}(A+B)$.
2. *Let $x \in$* $\operatorname{zer}(A+B)$*. Then $(Bx^k)_{k\in\mathbb{N}}$ converges strongly to the unique dual solution Bx.*

The forward–backward splitting algorithm dates back to (LIONS; MERCIER, 1979).

3 GENERAL SPLITTING ALGORITHM

In this chapter we consider the general problem of finding a zero of a sum of n maximal monotone operators, or MIP. The approach is as in (ECKSTEIN; SVAITER, 2009). The organization of the chapter is as follows. Section 3.1 introduces a general projection-separator algorithm to find a point in a set, that generates a Fejér monotone sequence. Section 3.2 shows how to frame the inclusion problem so a projection-separator algorithm can be applied, and it is presented a convergence condition in Proposition 3.4. Section 3.3 specifies some features of this general framework applied to the MIP, and presents an algorithm with those features. Finally, Section 3.4 shows that the algorithm presented in Section 3.3 converges weakly to a solution point in Theorem 3.1.

3.1 A GENERAL FRAMEWORK FOR A PROJECTION ALGORITHM

Let S be a closed convex set in a real Hilbert space $\mathcal{H}$. The following framework consider at each iteration k a half-space H_k containing the set S, then it performs a projection P_k onto this half-space.

Algorithm 1: Abstract projection algorithm

Data: Start with an arbitrary $p^0 \in \mathcal{H}$
for $k = 0,1,2,...$ **do**
  Find a half-space H_k such that $S \subset H_k$;
  Compute the projection $P_k p^k$ of p^k onto to the half-space H_k;
  Choose $\rho_k \in [0,2]$ and set $p^{k+1} = p^k + \rho_k(P_k p^k - p^k)$
end

This framework described in (COMBETTES, 2001) produces a Fejér monotone sequence with the properties listed in Proposition 2.3. Now we are going to describe analytically the elements of the general framework in Algorithm 1. To define such half-space at iteration k, we make use of an affine function φ_k, and we denote by P_k the projection onto the half-space

$$H_k := \{x \in \mathcal{H} \mid \varphi_k(x) \leq 0\} = \{x \in \mathcal{H} \mid \langle \nabla\varphi_k, x\rangle + \eta \leq 0\}. \tag{26}$$

As follows from Example 2.6 we have that

$$P_k(p) = \begin{cases} p & p \in H_k \\ p - \dfrac{\varphi_k(p)}{\|\nabla\varphi_k\|^2}\nabla\varphi_k & p \notin H_k. \end{cases}$$

Since $\varphi_k(p) > 0$ for $p \notin H_k$, the definition for the projection is summarized in the expression

$$P_k(p) = p - \frac{\max\{0, \varphi_k(p)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k.$$

We now restate Algorithm 1 in terms of the affine function φ_k as found in (ECKSTEIN; SVAITER, 2009), where S is closed convex subset of a Hilbert space U.

Algorithm 2: Abstract projection algorithm

Data: Start with an arbitrary $p^0 \in U$

for $k = 0, 1, 2, \ldots$ **do**

Determine a non-constant differentiable affine function $\varphi_k : U \to \mathbb{R}$ such that $\varphi_k(p) \leq 0$ for all $p \in S$.

Let $\overline{p}^k$ the projection of p^k onto the half-space (26) given by

$$\overline{p}^k = p^k - \frac{\max\{0, \varphi_k(p^k)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k. \tag{27}$$

Choose some relaxation parameter $\rho_k \in (0, 2)$, and set

$$p^{k+1} = p^k + \rho_k(\overline{p}^k - p^k). \tag{28}$$

end

The last two steps may simply be condensed to

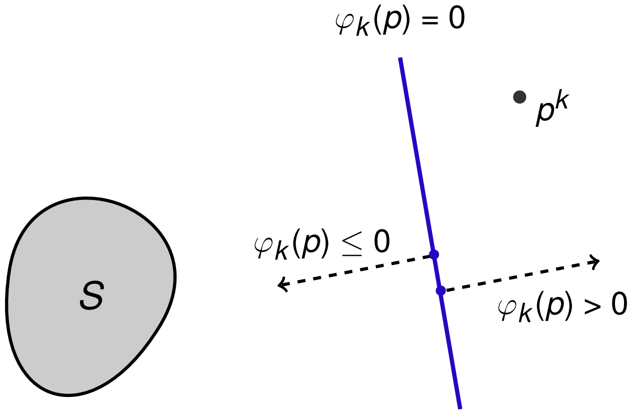
$$p^{k+1} = p^k - \rho_k \frac{\max\{0, \varphi_k(p^k)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k. \tag{29}$$

Note that in the projection computation, it may happen that $\varphi(p^k) \leq 0$, giving that $p^{k+1} = p^k$. This might happen if the hyperplane does not separate the current iterate p^k from the set S. Later, we will present a condition to ensure this separation.

Figure 1 presents a rough depiction of the current algorithm iterate p^k and the separator φ_k in the case that $\varphi_k(p^k) > 0$. The hyperplane is the boundary of the half-space H_k, and it always holds that $\varphi_k(p^*) \leq 0$ for every $p^* \in S$. When $\varphi_k(p^k) > 0$ (as shown), the hyperplane separates the current point p^k from the solution set S.

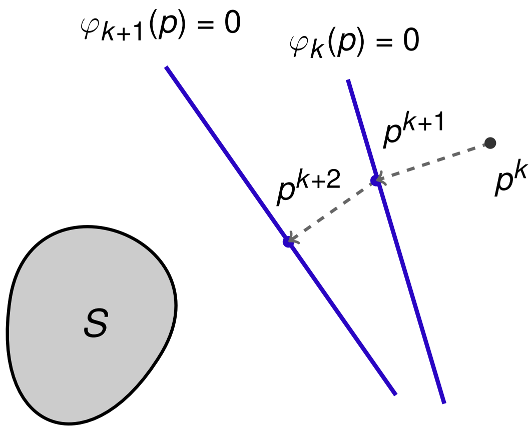
Figure 2 presents a rough depiction of two iterations of this process in the absence of over-relaxation or under-relaxation. Each iteration k constructs a separator φ_k as shown in Figure 1 and then obtains the next iteration by projecting onto the half-space $H_k = \{p \in U \mid \varphi_k(p) \leq 0\}$, within which the solution set S is known to lie.

Figure 1 – Properties of the hyperplane $H_k = \{p \in U \mid \varphi_k(p) = 0\}$ obtained from the affine function φ_k.



Source: (JOHNSTONE; ECKSTEIN, 2021), modified by the author.

Figure 2 – The basic operation of the method.



Source: (JOHNSTONE; ECKSTEIN, 2021), modified by the author.

Notice that in (27) we can write $P_k p^k = \overline{p}^k$, and that for $p^* \in S$ we have $P_k p^* = p^*$ for all $k \geq 1$. Now, we can use the fact that the projector is a firmly non-expansive operator to obtain

$$\begin{aligned} \|P_k p^* - P_k p^k\|^2 + \|(I - P_k)p^* - (I - P)p^k\|^2 &\leq \|p^* - p^k\|^2 \\ \|p^* - \overline{p}^k\|^2 + \|\overline{p}^k - p^k\|^2 &\leq \|p^* - p^k\|^2 \\ \|p^* - \overline{p}^k\|^2 &\leq \|p^* - p^k\|^2 - \|\overline{p}^k - p^k\|^2. \end{aligned} \tag{30}$$

Additionally, since P_k is projection we obtain using (10) in Theorem 2.2 that

$$\begin{aligned}\langle p^* - p^k, p^k - \overline{p}^k \rangle &= \langle p^* - p^k + \overline{p}^k - \overline{p}^k, p^k - \overline{p}^k \rangle \\ &= \langle p^* - P_k p^k, p^k - P_k p^k \rangle - \|p^k - \overline{p}^k\|^2 \\ &\leq -\|p^k - \overline{p}^k\|^2.\end{aligned}$$

Using this last inequality and (29), we have for $p^* \in S$ that

$$\begin{aligned}\|p^* - p^{k+1}\|^2 &= \|p^* - p^k - \rho_k(\overline{p}^k - p^k)\|^2 \\ &= \|p^* - p^k\|^2 + 2\rho_k \langle p^* - p^k, p^k - \overline{p}^k \rangle + \rho_k^2 \|\overline{p}^k - p^k\|^2 \\ &\leq \|p^* - p^k\|^2 - 2\rho_k \|p^k - \overline{p}^k\|^2 + \rho_k^2 \|\overline{p}^k - p^k\| \\ &= \|p^* - p^k\|^2 - \rho_k(2 - \rho_k)\|\overline{p}^k - p^k\|^2. \end{aligned} \tag{31}$$

The properties of a sequence generated by Algorithm 2 are in the following

Proposition 3.1. *Any infinite sequence $(p^k)_{k\geq 0}$ generated by Algorithm 2 behaves as follows*

1. *For any $p^* \in S$, the sequence $(\|p^k - p^*\|)_{k\geq 0}$ is nonincreasing, that is, $(p^k)_{k\geq 0}$ is Fejér monotone with respect to S.*
2. *If $p^{k_0} \in S$ for some $k_0 \geq 0$, then $p^k = p^{k_0}$ for all $k \geq k_0$.*
3. *If $(p^k)_{k\geq 0}$ has a strong accumulation point in S, then the whole sequence converges to that point.*
4. *If S is nonempty, then $(p^k)_{k\geq 0}$ is bounded. Moreover, if there exist $\underline{\rho}, \overline{\rho}$ such that $0 < \underline{\rho} \leq \rho_k \leq \overline{\rho} < 2$ for all k, then*
$$\sum_{k=0}^{\infty} \|p^k - \overline{p}^k\|^2 < \infty, \quad \sum_{k=0}^{\infty} \|p^k - p^{k+1}\|^2 < \infty. \tag{32}$$
5. *The sequence $(p^k)_{k\geq 0}$ has at most one weak accumulation point in S.*

Proof. 1. It follows from (31) that the sequence $(\|p^k - p^*\|)_{k\geq 0}$ is nonincreasing for any $p^* \in S$, thus $(p^k)_{k\geq 0}$ is Fejér monotone with respect to S.

2. If $p^{k_0} \in S$ for some $k_0 \geq 0$, then $\varphi_k(p^{k_0}) \leq 0$ for all $k \geq k_0$ and from (29) we conclude that $p^k = p^{k_0}$ for all $k \geq k_0$.
3. Follows from the fact of the sequence $(\|p^k - p^*\|)_{k\geq 0}$ being nonincreasing.
4. If S is nonempty we can consider for $p^* \in S$ the sequence $(\|p^k - p^*\|)_{k\geq 0}$. Thus from item 1 we obtain the boundedness of $(p^k)_{k\geq 0}$. To obtain the convergence of the first series, notice that from (31) follows the convergence of the series
$$\sum_{k=0}^{\infty} \rho_k(2 - \rho_k)\|p^k - \overline{p}^k\|^2.$$

Now, since there exist $\underline{\rho}, \overline{\rho}$ such that $0 < \underline{\rho} \leq \rho_k \leq \overline{\rho} < 2$ for all k, we have that there exist some $\beta > 0$ such that

$$[\rho_k(2-\rho_k)]^{-1} < \beta^{-1}.$$

Hence

$$\sum_{k=0}^{\infty} \|p^k - \overline{p}^k\|^2 \leq \frac{1}{\beta} \sum_{k=0}^{\infty} \rho_k(2-\rho_k)\|p^k - \overline{p}^k\|^2 < \infty.$$

To obtain the convergence of the second series we use (28) since we can write

$$\sum_{i=1}^{\infty} \|p^k - p^{k+1}\|^2 = \sum_{i=1}^{\infty} \rho_k^2 \|\overline{p}^k - p^k\|^2,$$

thus using the boundedness of ρ_k^2 and the convergence of the first series yields the result.

5. Suppose the sequence $(p^k)_{k\geq 0}$ has two weak accumulation points p and p', then there exist subsequences such that

$$p^{k_n} \to p \text{ and } p^{k_m} \to p'.$$

Since p and p' are in S we have by item 1 that the sequences $(\|p^k - p\|)_{k\in\mathbb{N}}$ and $(\|p^k - p'\|)_{k\in\mathbb{N}}$ converge. From

$$2\langle p^k, p-p'\rangle = \|p^k - p'\|^2 - \|p^k - p\|^2 - \|p\|^2 - \|p'\|^2$$

follow that $(\langle p^k, p-p'\rangle)_{k\geq 0}$ converges, say, to y. From the weakly convergence of the two subsequences follows that

$$\langle p^{k_n}, p-p'\rangle \to \langle p, p-p'\rangle = y \text{ and } \langle p^{k_m}, p-p'\rangle \to \langle p', p-p'\rangle = y.$$

Therefore, $\|p-p'\|^2 = 0$, that is, $p = p'$.

□

3.2 APPLICATION TO THE INCLUSION PROBLEM

Now we consider the general framework introduced in the last section applied to the following inclusion problem. Let $n \geq 2$ and let $T_i : \mathcal{H} \rightrightarrows \mathcal{H}$, be set-valued maximal monotone operator for $i = 1, \dots, n$. The problem is to find $z \in \mathcal{H}$ such that

$$0 \in T_1 z + \cdots + T_n z. \tag{33}$$

Since problem (33) deals with n maximal monotone operators, we will introduce a splitting scheme to deal only with each operator separately or their resolvents, instead of combinations such as $T_i + T_j$. Hence, it is assumed that each resolvent is available.

With the objective to use the general framework of Section 3.1 we need to define a closed convex set and a sequence of half-spaces containing it. Hence, we will construct a set of solutions S and for each iteration an affine function φ_k that generates the separating half-space.

3.2.1 Extended solution set

We start noticing that if an element $z \in \mathcal{H}$ is a solution to the problem $0 \in T_1 z + \cdots + T_n z$, then there exist $w_1, w_2, \ldots, w_n \in \mathcal{H}$ such that $w_i \in T_i z$ and that $w_1 + w_2 + \cdots + w_n = 0$. With this in mind we define the set

$$W := \{(w_1, \ldots, w_n) \in \mathcal{H}^n \mid w_1 + w_2 + \cdots + w_n = 0\}. \tag{34}$$

Consider the linear operator L and its adjoint L^* defined by

$$L : \mathcal{H} \to \mathcal{H}^n : z \mapsto (z, \ldots, z), \quad L^*(w_1, \ldots, w_n) = \sum_{i=1}^{n} w_i, \tag{35}$$

according to Example 2.2. We can express the set W as

$$W = \{(w_1, \ldots, w_n) \in \mathcal{H}^n \mid L^*(w_1, \ldots, w_n) = 0\}. \tag{36}$$

We will make use of this representation later.

Proposition 3.2. *The set W defined above is a subspace of $\mathcal{H}^n$ and it is closed.*

We now move the problem (33) to the Hilbert space $\mathcal{H} \times \mathcal{H}^n = \mathcal{H}^{n+1}$ under the canonical inner product

$$\langle (v, w_1, \ldots, w_n), (x, y_1, \ldots, y_n) \rangle = \langle v, x \rangle + \sum_{i=1}^{n} \langle w_i, y_i \rangle, \tag{37}$$

and define the set

$$V := \mathcal{H} \times W = \{(v, w_1, \ldots, w_n) \in \mathcal{H}^{n+1} \mid w_1 + \cdots + w_n = 0\}. \tag{38}$$

Clearly, the set V is a closed linear subspace of $\mathcal{H}^{n+1}$. To make an association with the problem we are considering we defined the *extended solution set* to be

$$\mathcal{S}_e(T_1, \ldots, T_n) := \{(z, w_1, \ldots, w_n) \in V \mid w_i \in T_i z, i = 1, \ldots, n\}, \tag{39}$$

which we will denote simply as $\mathcal{S}_e$. Notice that the way we defined the set $\mathcal{S}_e$ is deeply connected with the solutions to problem (33). This is stated in the following

Lemma 3.1. *Finding a point in $\mathcal{S}_e$ is equivalent to solving (33) in the sense that*

$$0 \in T_1 z + \ldots + T_n z \Longleftrightarrow \exists\, w_1, \ldots, w_n \in \mathcal{H} : (z, w_1, \ldots, w_n) \in \mathcal{S}_e.$$

The following proposition contains a necessary property of the set $\mathcal{S}_e$. The proof is based on the techniques found in (JOHNSTONE; ECKSTEIN, 2021) instead of the original in (ECKSTEIN; SVAITER, 2009).

Proposition 3.3. *The extended solution set $\mathcal{S}_e$ is closed and convex in $\mathcal{H}^{n+1}$.*

Proof. The main idea here is to express the set S_e as the zero set of a maximal monotone operator and then use Proposition 2.4. Consider $A : \mathcal{H} \to \mathcal{H} : x \mapsto 0$, the zero operator which is maximal monotone and $B : \mathcal{H} \to \mathcal{H}^n : z \mapsto T_1 z \times \cdots \times T_n z$ which is also maximal monotone by Example 2.5 (1). Define $L \in \mathcal{B}(\mathcal{H}, \mathcal{H}^n)$ as in (35). Denoting $(w_1, \ldots, w_n) \in \mathcal{H}^n$ by $\boldsymbol{w}$ we define the operators as in Example 2.6 by

$$\boldsymbol{M}(z, \boldsymbol{w}) = Az \times B^{-1}\boldsymbol{w}, \text{ and } \boldsymbol{S}(z, \boldsymbol{w}) = (L^*\boldsymbol{w}, -Lz).$$

Let's show that $\operatorname{zer}(\boldsymbol{M} + \boldsymbol{S}) = \mathcal{S}_e$. Indeed, let $(z, \boldsymbol{w}) \in \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S})$, then

$$\begin{aligned} 0 \in \boldsymbol{M}(z, \boldsymbol{w}) + \boldsymbol{S}(z, \boldsymbol{w}) &\Leftrightarrow 0 \in (Az + L^*\boldsymbol{w}) \times (B^{-1}\boldsymbol{w} - Lz) \\ &\Leftrightarrow L^*\boldsymbol{w} = 0 \text{ and } Lz \in B^{-1}\boldsymbol{w} \\ &\Leftrightarrow \boldsymbol{w} \in W \text{ and } \boldsymbol{w} \in B(Lz) \qquad \text{[using (36)]} \\ &\Leftrightarrow \boldsymbol{w} \in W \text{ and } w_i \in T_i z, \forall i = 1, \ldots, n \\ &\Leftrightarrow (z, w_1, \ldots, w_n) \in \mathcal{S}_e. \end{aligned}$$

Since $\operatorname{zer}(\boldsymbol{M} + \boldsymbol{S})$ is closed and convex, it follows the conclusion. □

3.2.2 Definition of the separators

We have the first ingredient of the standard projection algorithm, next we are going to construct the hyperplanes. To do that, we make use of the monotonocity of the operators T_i. Consider a point $(x_i, y_i) \in \operatorname{gra} T_i$ for $i = 1, \ldots, n$. From the definition of the set $\mathcal{S}_e$, we have that if a point $(z, w_1, w_2, \ldots, w_n) \in \mathcal{S}_e$ then specifically $(z, w_i) \in \operatorname{gra} T_i$ and by monotonicity

$$\langle z - x_i,\ w_i - y_i \rangle \geq 0.$$

Equivalently, $\langle z - x_i, y_i - w_i \rangle \leq 0$, and therefore

$$\sum_{i=1}^{n} \langle z - x_i, y_i - w_i \rangle \leq 0. \tag{40}$$

With this in mind we define the following function on $\mathcal{H}^{n+1}$

Definition 3.1. Let $V \subset \mathcal{H}^{n+1}$ be the subspace defined in (38), given $(x_i, y_i) \in \operatorname{gra} T_i$ for $i = 1, \ldots, n$, define $\varphi : V \to \mathbb{R}$ as

$$\varphi(z, w_1, \ldots, w_n) := \sum_{i=1}^{n} \langle z - x_i, y_i - w_i \rangle. \tag{41}$$

The properties of this function is a step closer to the application of the standard projection algorithm. In view of Lemma 3.1 and Proposition 3.3, we attempt to solve (33) by finding a point in $S_e(T_1, \ldots, T_n)$. The following lemma details the properties of these separators.

Lemma 3.2. *Let $\varphi : V \to \mathbb{R}$ as in Definition 3.1. Then, for any $(z, w_1, \ldots, w_n) \in \mathcal{S}_e$, one has $\varphi(z, w_1, \ldots, w_n) \leq 0$, that is,*

$$\mathcal{S}_e \subseteq \{(z, w_1, \ldots, w_n) \in V \mid \varphi(z, w_1, \ldots, w_n) \leq 0\}. \tag{42}$$

Additionally, φ is affine on V, with

$$\nabla\varphi = \left(\sum_{i=1}^{n} y_i, x_1 - \overline{x}, x_2 - \overline{x}, \ldots, x_n - \overline{x}\right), \text{ where } \overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{43}$$

and

$$\begin{aligned}\nabla\varphi = 0 &\iff (x_1, y_1, \ldots, y_n) \in \mathcal{S}_e,\ x_1 = x_2 = \ldots = x_n \\ &\iff \varphi(z, w_1, \ldots, w_n) = 0\ \forall\,(z, w_1, \ldots, w_n) \in V.\end{aligned}$$

Proof. The inclusion in (42) follows from the discussion that leads to the definition of the separator in (40). To prove that φ is affine on V, define the operator L as in (35), setting $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$, and $\boldsymbol{w} = (w_1, \ldots, w_n)$ we can write (41) using the canonical inner product of the product space $\mathcal{H}^n$ as

$$\varphi(z, \boldsymbol{w}) = \langle Lz - \boldsymbol{x}, \boldsymbol{y} - \boldsymbol{w}\rangle. \tag{44}$$

Recall from (36) that for $(z, \boldsymbol{w}) \in V$ we have that $L^*\boldsymbol{w} = 0$. Define $\overline{x} = \frac{1}{n}L^*\boldsymbol{x}$, and using the expression in (44) we obtain

$$\begin{aligned}\varphi(z, \boldsymbol{w}) &= \langle z, L^*\boldsymbol{y}\rangle - \langle z, L^*\boldsymbol{w}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{w}\rangle \qquad (45)\\ &= \langle z, L^*\boldsymbol{y}\rangle + \langle \boldsymbol{x} - L\overline{x}, \boldsymbol{w}\rangle + \langle L\overline{x}, w\rangle - \langle x, y\rangle \\ &= \langle z, L^*\boldsymbol{y}\rangle + \langle \boldsymbol{x} - L\overline{x}, \boldsymbol{w}\rangle + \langle \overline{x}, L^*\boldsymbol{w}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle \\ &= \langle (z, \boldsymbol{w}), (L^*\boldsymbol{y}, \boldsymbol{x} - L\overline{x})\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle. \qquad (46)\end{aligned}$$

We have to prove that the vector $(L^*y, x - L\overline{x})$ is in V, or more specifically that $x - L\overline{x} \in W$ or that $L^*(x - L\overline{x}) = 0$ according to (36). This follows from the fact that $L^*L = nI$ and $L^*x = n\overline{x}$. Hence, from (46) this gives us that φ is affine on V, and yields that $\nabla\varphi = (L^*y, x - L\overline{x})$.

Lastly, from (43) we have that $\nabla\varphi = 0$ if and only if

$$\sum_{i=1}^{n} y_i = 0 \text{ and } x_i = \overline{x}\ \forall\ i = 1, \ldots, n.$$

Since $(x_i, y_i) \in \operatorname{gra} T_i$ we obtain that $(\overline{x}, y_1, \ldots, y_n) \in \mathcal{S}_e$, which proves the first equivalence. For the second equivalence, we notice that if $\nabla\varphi = 0$ the expression of φ in (46) reduces to

$$\begin{aligned}\varphi(z, \boldsymbol{w}) &= -\langle x, y\rangle = \sum_{i=1}^{n}\langle x_i, y_i\rangle \\ &= \left\langle \overline{x}, \sum_{i=1}^{n} y_i \right\rangle = 0.\end{aligned}$$

□

Note that φ is not an affine function in the space $\mathcal{H}^{n+1}$ but only on its subspace V, where the cross term $\langle z, L^*w\rangle$ in (45) must be zero. We will implement the algorithm within the subspace V.

3.2.3 First Convergence Analysis

If we want that the affine function φ_k creates a separation between the set $\mathcal{S}_e$ and the iterate $p^k = (z^k, w_1^k, \ldots, w_n^k) \notin \mathcal{S}_e$ then we require that $\varphi_k(p^k) > 0$. Since the function is defined in terms of the points $(x_i^k, y_i^k) \in \text{gra}\, T_i$ then we must have a way to choose such points to guarantee this separation. As we commented, in Algorithm 1 the affine function φ_k might not separate p^k from $\mathcal{S}_e$, or even when it makes some separation, this might be shallow. The condition $\varphi_k(p^k) \geq \xi\|\nabla\varphi_k\|^2$ for all $k \geq 0$, with $\xi > 0$ a fixed constant, guarantees convergence.

We now perform a preliminary analysis of the convergence properties of Algorithm 2. The following proposition contains a separation condition that along with Hypotheses 5 and 6 ensure weak convergence. In the original paper (ECKSTEIN; SVAITER, 2009) we found Hypothesis 4 as part of the reasoning. However, the note in (BAUSCHKE, 2009) shows how Theorem 2.3 can be used to prove the weak convergence of the sequence $(p^k)_{k\geq 0}$ without the Hypothesis 4 below. The same idea is repeated in a similar result found in Lemma 4.4. In view of that, the proof presented here make use of Theorem 2.3.

Proposition 3.4. *Suppose that the following conditions are met in Algorithm 2:*

1. $S_e(T_1,...,T_n) \neq \emptyset$
2. $0 < \rho \leq \rho_k \leq \overline{\rho} < 2$ *for all* k.
3. *There exists some scalar* $\xi > 0$ *such that, for all* $k \geq 0$,

$$\varphi_k(p^k) = \varphi_k(z^k, w_1^k, \ldots, w_n^k) \geq \xi\|\nabla\varphi_k\|^2 = \xi\left(\|\sum_{i=1}^n y_i^k\|^2 + \sum_{i=1}^n \|x_i^k - \overline{x}^k\|^2\right) \tag{47}$$

Then $\nabla\varphi_k \to 0$, *that is* $x_i^k - x_j^k \to 0$ *for all* $i,j = 1, \ldots, n$ *and* $\sum_{i=1}^n y_i^k \to 0$. *Furthermore,* $\varphi_k(p^k) \to 0$. *If it is also true that*

4. *Either* $\mathcal{H}$ *has finite dimension or the operator* $T_1 + \cdots + T_n$ *is maximal,*
5. $z^k - \overline{x}^k \to 0$
6. $w_i^k - y_i^k \to 0$, *for* $i = 1, \ldots, n$,

then $(p^k)_{k\geq 0}$ *converges weakly to some* $p^\infty = (z^\infty, w_1^\infty, \ldots, w_n^\infty) \in S_e(T_1, \cdots, T_n)$, *which implies that* z^∞ *solves (33). Furthermore,* $x_i^k \rightharpoonup z^\infty$ *and* $y_i^k \rightharpoonup w_i^\infty$ *for* $i = 1, \ldots, n$.

Proof. For the first part of the proof we assume Hypotheses 1-3 true. Hypothesis 3 implies that $\varphi_k(p^k) \geq 0$, and according to (27) we have

$$\|p^k - \overline{p}^k\| = \frac{\varphi_k(p^k)}{\|\nabla \varphi_k\|}, \tag{48}$$

For all k having $\nabla\varphi_k \neq 0$. Substituting $\varphi_k(p^k) \geq \xi\|\nabla\varphi_k\|^2$ into this equation, we obtain

$$\|p^k - \overline{p}^k\| \geq \xi\|\nabla\varphi_k\|. \tag{49}$$

Hypotheses 1 and 2 are those in Proposition 3.1(4), hence (32) implies that

$$\|p^k - \overline{p}^k\| \to 0.$$

Thus (49) implies $\nabla\varphi_k \to 0$. From the expression for $\nabla\varphi_k$ in (43), we immediately have

$$\sum_{i=1}^{n} y_i^k \to 0 \text{ and } x_i^k - \overline{x}^k \to 0 \text{ for } i = 1, \ldots, n. \tag{50}$$

And thus $x_i^k - x_j^k \to 0$ for all $i,j = 1, \ldots, n$. Recall that by Lemma 3.2, $\varphi_k(p^k) = 0$ whenever $\nabla\varphi_k = 0$, hence if $\nabla\varphi_k \neq 0$ we can multiply (48) by $\|\nabla\varphi_k\|$ and obtain

$$\varphi_k(p^k) = \|p^k - \overline{p}^k\|\|\nabla\varphi_k\|. \tag{51}$$

Therefore, we have established the equality in (51) for all $k \geq 0$, since $\nabla\varphi_k \to 0$ and $\|p^k - \overline{p}^k\| \to 0$ we conclude that $\varphi_k(p^k) \to 0$.

Now, we focus on the proof of the second part without Hypothesis 4. The strategy is to use Lemma 2.2 and Theorem 2.3. Recall that the sequence $(p^k)_{k\geq 0}$ by Proposition 3.1(1) is Fejér monotone with respect to $\mathcal{S}_e$, that is, it satisfies the first hypothesis of Lemma 2.2. For the second hypothesis, consider any weak cluster point $p^\infty = (z^\infty, w_1^\infty, \ldots, w_n^\infty)$ of the bounded sequence $(p^k)_{k\geq 0}$, to prove that this point belongs to $\mathcal{S}_e$ we will use Theorem 2.3.

Since p^∞ is a weak cluster point of the sequence $(p^k)_{k\geq 0}$, there exists a subsequence $(p^{k_m})_{m\geq 0}$ such that $z^{k_m} \rightharpoonup z^\infty$ and $w_i^{k_m} \rightharpoonup w_i^\infty$ for $i = 1, \ldots, n$.

The sequences in Theorem 2.3 will be $(x_1^{k_m}, \ldots, x_n^{k_m})$ and $(y_1^{k_m}, \ldots, y_n^{k_m})$, recall that they satisfy $y_i^{k_m} \in T_i x_i^{k_m}$, for $i = 1, \ldots, n$. Next, we establish the weak convergence of these sequences. From Hypothesis 5 and $x_i^k - \overline{x}^k \to 0$, we immediately obtain

$$z^k - x_i^k \to 0, \text{ for } i = 1, \ldots, n,$$

thus combining this and $z^{k_m} \rightharpoonup z^\infty$ follows that

$$x_i^{k_m} \rightharpoonup z^\infty, \text{ for } i = 1, \ldots, n. \tag{52}$$

From Hypothesis 6 and $w_i^{k_m} \rightharpoonup w_i^\infty$ we also obtain

$$y_i^{k_m} \rightharpoonup w_i^\infty, \text{ for } i = 1, \ldots, n. \tag{53}$$

Now, for the second part of Theorem 2.3 we consider projections over a closed subspace of $\mathcal{H}^n$. This closed subspace is defined by

$$C = \{(v_1, v_2, \ldots, v_n) \in \mathcal{H}^n \mid v_1 = v_2 = \cdots = v_n\}.$$

Whose orthogonal complement is given by

$$C^{\perp} = \left\{ (v_1, v_2, \ldots, v_n) \in \mathcal{H}^n \mid \sum_{i=1}^{n} v_i = 0 \right\}.$$

The projection over the set C was treated in Proposition 2.8. Notice that the two convergences in (50) can be represented in terms of this set as

$$P_C(y_1^{k_m}, \ldots, y_n^{k_m}) \to 0 \text{ and } (x_1^{k_m}, \ldots, x_n^{k_m}) - P_C(x_1^{k_m}, \ldots, x_n^{k_m}) \to 0,$$

respectively. This, (52) and (53) are the hypotheses in Theorem 2.3 with $T = T_1 \times \cdots \times T_n$, so we can conclude that

$$\big((z^{\infty}, \ldots, z^{\infty}), (w_1^{\infty}, \ldots, w_n^{\infty})\big) \in (C \times C^{\perp}) \cap \operatorname{gra} T.$$

This implies that $\sum w_i^{\infty} = 0$ and $w_i^{\infty} \in T_i z^{\infty}$ for $i = 1, \ldots, n$, in other words, $p^{\infty} \in \mathcal{S}_e$. In conclusion, we have that any weak cluster point of the sequence $(p^k)_{k \geq 0}$ is in $\mathcal{S}_e$, hence by Lemma 2.2 the whole sequence $(p^k)_{k \geq 0}$ converges weakly to the point $(z^{\infty}, w_1^{\infty}, \ldots, w_n^{\infty}) \in \mathcal{S}_e$.

□

Remark 3.1. Proposition 3.4 allows to understand that under condition (47) we have that the sequence of points $(x_i^k, y_i^k) \in \operatorname{gra} T_i$ actually approach to a solution of problem (33), since we are getting $x_i^k = x_j^k$ for all $i, j = 1, \ldots, n$ and $\sum_{i=1}^{n} y_i^k \to 0$. Therefore, if at some iteration of Algorithm 2 we have that $x_1^k = \cdots = x_n^k$ and $\sum_{i=1}^{n} y_i^k = 0$ we set $w_i^{k+1} = y_i^k$ for $i = 1, \ldots, n$ and $z^{k+1} = x_1^k$ and we have encountered a solution.

3.3 PROJECTIVE SPLITTING ALGORITHM

Having established a convergence condition for Algorithm 2, now we show how the crucial part, that is, how the construction of the separator can be made. Since the definition of the separator φ depends on the chosen points $(x_i, y_i) \in \operatorname{gra} T_i$, it is natural to ask how to choose them. In each iteration of the algorithm we require separation from the current point $p = (z, w_1, \ldots, w_n) \in V$ and the set $\mathcal{S}_e$, we deduce from (42) that we require that

$$\varphi(z, w_1, \ldots, w_n) > 0.$$

Suppose that $p \in V \setminus \mathcal{S}_e$, notice that if in (41) we have that

$$z - x_i = \lambda_i (y_i - w_i) \text{ with } \lambda_i > 0, \tag{54}$$

this yields

$$\varphi(z, w_1, \ldots, w_n) = \sum \lambda_i ||z - x_i||^2 > 0,$$

unless $z = x_1 = \cdots = x_n$ in which case we have from (54) that $y_i = w_i$ for all $i = 1, \ldots, n$, and since $p \in V$ this implies that $(z, w_1, \ldots, w_n) \in \mathcal{S}_e$, contrary to the assumption. The condition in (54) can be stated as

$$z + \lambda w_i = x_i + \lambda_i y_i \Rightarrow z + \lambda_i w_i \in (I + \lambda_i T)(x_i). \tag{55}$$

By the maximal monotonicity of the T_i, there exists according to Theorem 2.1 a unique $(x_i, y_i) \in \text{gra } T_i$ satisfying the inclusion in (55). Finding the $(x_i, y_i) \in \text{gra } T_i$ is equivalent to evaluating the resolvent $(I + \lambda_i T_i)^{-1}$, which is, by assumption, tractable for each individual T_i.

3.3.1 Generalizations of the way of choosing the points

As we saw earlier, a way to choose the points $(x_i^k, y_i^k) \in \text{gra } T_i$ is one satisfying

$$x_i^k + \lambda_i^k y_i^k = z^k + \lambda_i^k w_i^k.$$

As commented in (ECKSTEIN; SVAITER, 2009) we can generalized this scheme by performing the proximal calculations for the T_i sequentially at each iteration starting with $i = 1$ and finishing with $i = n$, using "recent" information generated in calculating (x_j^k, y_j^k) where $j < i$ when calculating (x_i^k, y_i^k). Specifically, when calculating (x_i^k, y_i^k), we consider replacing z^k with an affine combination of the vectors z^k and x_j^k, $j < i$. Starting with operator T_1 we find the unique $(x_1^k, y_1^k) \in \text{gra } T_1$ such that

$$x_1^k + \lambda_1^k y_1^k = z^k + \lambda_1^k w_1^k.$$

Next, for operator T_2 we take some $\alpha_{21}^k \in \mathbb{R}$ and find the unique $(x_2^k, y_2^k) \in \text{gra } T_2$

$$x_2^k + \lambda_2^k y_2^k = (1 - \alpha_{21}^k) z^k + \alpha_{21}^k x_1^k + \lambda_2^k w_2^k.$$

To continue, we choose some $\alpha_{31}^k, \alpha_{32}^k \in \mathbb{R}$ and find the unique $(x_3^k, y_3^k) \in \text{gra } T_3$ such that

$$x_3^k + \lambda_3^k y_3^k = (1 - \alpha_{31}^k - \alpha_{32}^k) z^k + \alpha_{31}^k x_1^k + \alpha_{32}^k x_2^k + \lambda_3^k w_3^k,$$

and so forth. In general, we choose $(x_i^k, y_i^k) \in \text{gra } T_i$ to satisfy the conditions

$$x_i^k + \lambda_i^k y_i^k = \left(1 - \sum_{j=1}^{i-1} \alpha_{ij}^k\right) z^k + \sum_{j=1}^{i-1} \alpha_{ij}^k x_j^k + \lambda_i^k w_i^k, \quad y_i^k \in T_i x_i^k. \tag{56}$$

In addition to this flexibility afforded by the choice of the α_{ij}^k and λ_i^k, it is considered two more generalizations

1. Errors $e_i^k \in \mathcal{H}$ are allowed in (56) as long as they satisfy a condition defined later in (61).
2. The order of processing the operators may vary from iteration to iteration. At iteration k, this order is specified by an permutation $\pi_k(\cdot)$ of $\{1,\ldots,n\}$.

This flexible scheme is summarize in the equation

$$x_{\pi_k(i)}^k + \lambda_i^k y_{\pi_k(i)}^k = \left(1 - \sum_{j=1}^{i-1} \alpha_{ij}^k\right) z^k + \sum_{j=1}^{i-1} \alpha_{ij}^k x_{\pi_k(j)}^k + \lambda_i^k w_{\pi_k(i)}^k + e_i^k. \tag{57}$$

3.3.2 Presenting the Algorithm

Having introduced a flexible scheme to choose a point in gra T_i we would like to state an instance of Algorithm 2 applied to problem (33) and analyze the convergence of it using Proposition 3.4, specifically the condition in item 3. To introduce such condition we will employ some standard matrix analysis.

Given an $n \times n$ real matrix $\boldsymbol{L}$, we define $\|\boldsymbol{L}\|$ to be its operator 2-norm and $\kappa(\boldsymbol{L})$ to be the smallest eigenvalue of its symmetric part, that is,

$$\|\boldsymbol{L}\| = \max_{\|x\|=1} \|\boldsymbol{L}x\|, \quad \operatorname{sym} \boldsymbol{L} = \frac{1}{2}(\boldsymbol{L} + \boldsymbol{L}^\top), \quad \kappa(\boldsymbol{L}) = \min \operatorname{eig} \operatorname{sym} \boldsymbol{L}.$$

It is straightforward to show that $\kappa(\boldsymbol{L}) \leq \|\boldsymbol{L}\|$ and that, for any $x \in \mathbb{R}^n$ $\langle x, \boldsymbol{L}x\rangle \geq \kappa(\boldsymbol{L})\|x\|^2$. In analogy to the usual linear map $\mathbb{R}^n \to \mathbb{R}^n$ associated with $\boldsymbol{L}$ we can define a linear mapping $\mathcal{H}^n \to \mathcal{H}^n$ corresponding to $\boldsymbol{L}$ via

$$\boldsymbol{L}u = \boldsymbol{L}(u_1,\ldots,u_n) = (v_1,\ldots,v_n), \text{ where } v_i = \sum_{j=1}^{n} \ell_{ij} u_j \in \mathcal{H}, \tag{58}$$

with ℓ_{ij} denoting the elements of $\boldsymbol{L}$. In turns out, this mapping retains key spectral properties that $\boldsymbol{L}$ exhibits over $\mathbb{R}^n$.

Lemma 3.3. *Let $\boldsymbol{L}$ be any $n \times n$ real matrix. For all $u = (u_1,\ldots,u_n) \in \mathcal{H}^n$,*

$$\|\boldsymbol{L}u\| \leq \|\boldsymbol{L}\|\|u\| \tag{59}$$

$$\langle u, \boldsymbol{L}u\rangle \geq \kappa(\boldsymbol{L})\|u\|^2. \tag{60}$$

Where $\boldsymbol{L}u$ is defined as in (58), $\langle\cdot,\cdot\rangle$, denotes the canonical inner product for $\mathcal{H}^n$ induced by the inner product for $\mathcal{H}$, and $\|\cdot\|$ applied to elements of $\mathcal{H}^n$ denotes the norm induced by this inner product.

The proof of this lemma can be found in the Appendix. Related to (57) the following $n \times n$ matrices are constructed

$$\boldsymbol{\Lambda}_k = \operatorname{diag}(\lambda_1^k,\ldots,\lambda_n^k),$$

that is, the diagonal matrix with entries λ_i^k. The matrix $\boldsymbol{A}_k = (a_{ij}^{(k)})_{i,j=1,\dots,n}$ where

$$a_{ij}^{(k)} = \begin{cases} 1, & \text{if } i = j, \\ -\alpha_{ij}^k, & \text{if } i > j, \\ 0, & \text{if } i < j \end{cases}$$

In (ECKSTEIN; SVAITER, 2009) the error condition is stated in terms of these matrices as follows

$$\sum_{i=1}^{n} (\lambda_i^k)^{-1} \|e_i^k\|^2 \leq \sigma^2 \kappa(\boldsymbol{\Lambda}_k^{-1}\boldsymbol{A}_k)^2 \sum_{i=1}^{n} \|x_i^k - z^k\|^2, \quad \sigma \in [0,1). \tag{61}$$

Additionally, if the matrices $\boldsymbol{\Lambda}_k^{-1}\boldsymbol{A}_k$ are such that there exist $\beta, \zeta > 0$ such that

$$\kappa(\boldsymbol{\Lambda}_k^{-1}\boldsymbol{A}_k) \geq \zeta \text{ and } \|\boldsymbol{\Lambda}_k^{-1}\boldsymbol{A}_k\| \leq \beta \quad \forall\ k \geq 0, \tag{62}$$

then when choosing the points $(x_i^k, y_i^k) \in \operatorname{gra} T_i$ via (57) the hypotheses of Proposition 3.4 are met. Algorithm 3 gathers all these conditions.

Algorithm 3: Projective splitting algorithm

Data: Choose scalars $\beta, \zeta > 0$, $0 < \underline{\rho} < \overline{\rho} < 2$, and $\sigma \in [0,1)$. Start with an arbitrary $(z_0, w_1, w_2, \dots, w_n) \in V$, that is, $w_1 + \cdots + w_n = 0$.

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

Choose scalars $\lambda_i^k > 0$, $i = 1,\dots,n$ and α_{ij}^k with $1 \leq j < i \leq n$ such that $\kappa(\boldsymbol{\Lambda}_k^{-1}\boldsymbol{A}_k) \geq \zeta$ and $\|\boldsymbol{\Lambda}_k^{-1}\boldsymbol{A}_k\| \leq \beta$, where $\boldsymbol{\Lambda}_k^{-1}$ and $\boldsymbol{A}_k$ are defined as above.

Let $\pi_k(\cdot)$ be any permutation of $\{1,...,n\}$. For $i = 1,\dots,n$, find $(x_i^k, y_i^k) \in \operatorname{gra} T_i$ satisfying (57) and (61).

If $x_1^k = x_2^k = \cdots = x_n^k$ and $\sum_{i=1}^n y_i^k = 0$, let $w_i^{k+1} = y_i^k$ for $i = 1,\dots,n$ and $z^{k+1} = x_1^k$. Otherwise, continue.

Choose some $\rho_k \in [\underline{\rho}, \overline{\rho}]$ and set

$$\overline{x}^k := \frac{1}{n}\sum_{i=1}^{n} x_i^k \tag{63}$$

$$\theta_k := \frac{\sum_{i=1}^n \langle z^k - x_i^k, y_i^k - w_i^k \rangle}{\|\sum_{i=1}^n y_i^k\|^2 + \sum_{i=1}^n \|x_i^k - \overline{x}^k\|^2} \tag{64}$$

$$z^{k+1} = z^k - \rho_k \theta_k \sum_{i=1}^{n} y_i^k \tag{65}$$

$$w_i^{k+1} = w_i^k - \rho_k \theta_k (x_i^k - \overline{x}^k) \tag{66}$$

end

Equations (63)-(66) came from the update formula in (29), the definition of φ_k, and the form of $\nabla\varphi_k$ as obtained in (43). Notice that step 4 guarantees that the denominator in (64) cannot be zero. The computation of θ_k in (64) is basically $\varphi(p^k)/\|\nabla\varphi_k\|^2$, in fact, steps 2 and 3 ensure that $\varphi_k(p^k) > 0$ as we will see in Remark 3.2. Finally, note also that $p^k \in V$ and the update (66) ensures $w_1^{k+1} + \cdots + w_n^{k+1} = 0$, so by induction all iterates $p^k = (z^k, w_1^k, \cdots, w_n^k)$ produced by Algorithm 3 lies in V.

Here we consider a simpler case than that in (ECKSTEIN; SVAITER, 2009), where $\alpha_{ij}^k = 0$ for all $k \leq 0$ and $i,j = 1,\ldots,n$. With this change we have $\boldsymbol{A}_k = \boldsymbol{I}_n$ for all $k \geq 0$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Hence (62) transform into

$$\kappa(\boldsymbol{\Lambda}_k^{-1}) \geq \zeta \text{ and } \|\boldsymbol{\Lambda}_k^{-1}\| \leq \beta \quad \forall\, k \geq 0. \tag{67}$$

Therefore, the expression in (57) becomes

$$x_{\pi_k(i)}^k + \lambda_i^k y_{\pi_k(i)}^k = z^k + \lambda_i^k w_{\pi_k(i)}^k + e_i^k. \tag{68}$$

And the error condition is now

$$\sum_{i=1}^n (\lambda_i^k)^{-1}\|e_i^k\|^2 \leq \sigma^2\kappa(\boldsymbol{\Lambda}_k^{-1})^2 \sum_{i=1}^n \|x_i^k - z^k\|^2, \quad \sigma \in [0,1). \tag{69}$$

The error condition (61) is an n–operator generalization of the relative error tolerance proposed in (SOLODOV; SVAITER, 1999b, 1999a, 2001) for modified proximal-extragradient projection methods.

3.4 MAIN CONVERGENCE PROOF

In this section we will prove the convergence of Algorithm 3.

First, we prove a general result about the gradient of φ_k. In order to do that, we define auxiliary sequences $(p^k)_{k\geq 0} \subset \mathcal{H}^{n+1}$, $(u^k)_{k\geq 0} \subset \mathcal{H}^n$, and $(v^k)_{k\geq 0} \subset \mathcal{H}^n$ via

$$p^k := (z^k, w_1^k, \ldots, w_n^k), \qquad u_i^k := x_i^k - z^k, \qquad v_i^k := w_i^k - y_i^k \tag{70}$$

for all $i = 1, \ldots, n$ and $k \geq 0$, and also define as in (41) the function

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

From (70), we immediately have

$$\begin{aligned}\varphi_k(p^k) = \varphi_k\left(z^k, w_1^k, \ldots, w_n^k\right) &= \sum_{i=1}^n \langle z^k - x_i^k, y_i^k - w_i^k\rangle \\ &= \langle u^k, v^k\rangle = \sum_{i=1}^n \langle u_i^k, v_i^k\rangle.\end{aligned} \tag{71}$$

With this in mind we prove the following

Lemma 3.4. *The gradient* $\nabla\varphi_k$ *satisfies*

$$\|\nabla\varphi_k\|^2 \le n\|v^k\|^2 + \|u^k\|^2. \tag{72}$$

Proof. To do so, first note that since $\sum_{i=1}^{n} w_i^k = 0$,

$$\sum_{i=1}^{n} v_i^k = \sum_{i=1}^{n}\left(w_i^k - y_i^k\right) = -\sum_{i=1}^{n} y_i^k \Rightarrow \left\|\sum_{i=1}^{n} v_i^k\right\|^2 = \left\|\sum_{i=1}^{n} y_i^k\right\|^2. \tag{73}$$

Next, define

$$\overline{u}^k := \frac{1}{n}\sum_{i=1}^{n} u_i^k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^k - z^k\right) = \overline{x}^k - z^k,$$

and observe that for all $i = 1, \dots, n$ and $k \ge 0$,

$$u_i^k - \overline{u}^k = x_i^k - z^k - \left(\overline{x}^k - z^k\right) = x_i^k - \overline{x}^k. \tag{74}$$

Substituting (73) and (74) into the expression for $\|\nabla\varphi_k\|^2$ arising from Lemma 3.2, we obtain

$$\begin{aligned} \|\nabla\varphi_k\|^2 &= \left\|\sum_{i=1}^{n} y_i^k\right\|^2 + \sum_{i=1}^{n}\|x_k^k - \overline{x}^k\|^2 \\ &= \left\|\sum_{i=1}^{n} v_i^k\right\|^2 + \sum_{i=1}^{n}\|u_i^k - \overline{u}^k\|^2 \\ &= \frac{1}{n}\|\boldsymbol{E}v^k\|^2 + \|\boldsymbol{M}u^k\|^2, \end{aligned}$$

where we define $\boldsymbol{E}$ to be the $n \times n$ matrix of all ones and $\boldsymbol{M} = \boldsymbol{I} - (1/n)\boldsymbol{E}$. Applying (59), it then follows that

$$\|\nabla\varphi_k\|^2 \le \frac{1}{n}\|\boldsymbol{E}v^k\|^2 + \|\boldsymbol{M}u^k\|^2 \le \frac{1}{n}\|\boldsymbol{E}\|^2\|v^k\|^2 + \|\boldsymbol{M}\|^2\|u^k\|^2. \tag{75}$$

Over $\mathbb{R}^n$, the matrix $\boldsymbol{M}$ represents orthogonal projection onto the nontrivial subspace $T = \{(t_1, \dots, t_n) \in \mathbb{R}^n \mid t_1 + \cdots + t_n = 0\}$, so we conclude $\|\boldsymbol{M}\| = 1$. It also follows that $\boldsymbol{I} - \boldsymbol{M}$ represents orthogonal projection onto the nontrivial subspace $T^\perp$, so

$$\|\boldsymbol{I} - \boldsymbol{M}\| = 1 \Rightarrow \|\boldsymbol{E}\| = \|n(\boldsymbol{I} - \boldsymbol{M})\| = n\|\boldsymbol{I} - \boldsymbol{M}\| = n.$$

Therefore, (75) reduces to

$$\|\nabla\varphi_k\|^2 \le \left(\frac{1}{n}\right)n^2\|v^k\|^2 + \|u^k\|^2 = n\|v^k\|^2 + \|u^k\|^2.$$

□

Finally, we can prove the convergence of Algorithm 3 in the following theorem. The goal here is to prove that the conditions (69) and (67) imply the sufficient separation condition (47).

This theorem as found in (ECKSTEIN; SVAITER, 2009) has the hypothesis of either $\mathcal{H}$ having finite dimension or the operator $T_1 + \ldots + T_n$ being maximal, here we drop this hypothesis since, as was commented before Proposition 3.4, it is not longer needed.

Theorem 3.1. *Suppose that (33) has a solution. Then, in Algorithm 3, the sequences* $(z^k)_{k\geq 0}$, $(x_1^k)_{k\geq 0}, \ldots, (x_n^k)_{k\geq 0} \subset \mathcal{H}$ *all weakly converge to some* z^∞ *solving (33). For each* $i = 1, \ldots, n$ *, we also have* $w_i^k, y_i^k \rightharpoonup y_i^\infty$ *, where* $y_i^\infty \in T_i z^\infty$ *and also* $y_1^\infty + \ldots + y_n^\infty = 0$.

Proof. Define $e^k = \left(e_1^k, \ldots, e_n^k\right) \in \mathcal{H}^n$ for all $k \geq 0$, and observe that by taking square roots and substitution of the definitions of e^k and u^k, (69) simplifies via the notation (58) and the definition of $\boldsymbol{\Lambda}_k$ to

$$\|\boldsymbol{\Lambda}_k^{-1} e^k\| \leq \sigma\kappa(\boldsymbol{\Lambda}_k^{-1})\,\|u^k\|. \tag{76}$$

Take any $i \in 1, \ldots, n$. Subtracting z^k from both sides of (68) and regrouping yields

$$\begin{aligned} &\left(x_{\pi_k(i)}^k - z^k\right) + \lambda_i^k y_{\pi_k(i)}^k = z^k - z^k + \lambda_i^k w_{\pi_k(i)}^k + e_i^k \\ \Leftrightarrow\quad &\left(x_{\pi_k(i)}^k - z^k\right) - e_k^i = \lambda_i^k \left(w_{\pi_k(i)}^k - y_{\pi_k(i)}^k\right). \end{aligned}$$

Dividing by λ_i^k and substituting the definitions of u_i^k and v_i^k yields

$$\left(\frac{1}{\lambda_i^k}\right)\left(u_{\pi_{k(i)}}^k - e_i^k\right) = v_{\pi_{k(i)}}^k. \tag{77}$$

Applying the notation (58) to (77) for $i = 1, \ldots, n$ produces

$$v^k = (\boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\Pi}_k^\top) u^k - (\boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1})\, e^k, \tag{78}$$

where $\boldsymbol{\Pi}_k$ is the $n \times n$ permutation matrix corresponding to the permutation $\pi_k(\cdot)$. Substituting (78) into (71) yields

$$\begin{aligned} \varphi_k(p^k) &= \left\langle u^k, \boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\Pi}_k^\top u^k \right\rangle - \left\langle u^k, \boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} e^k \right\rangle \\ &\geq \left\langle u^k, \boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\Pi}_k^\top u^k \right\rangle - \|u^k\| \|\boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} e^k\| && [\text{ Cauchy - Schwarz }] \\ &\geq \kappa(\boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\Pi}_k^\top) \|u^k\|^2 - \|u^k\| \|\boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k^{-1} e^k\| && [\text{ using (60) }] \\ &= \kappa(\boldsymbol{\Lambda}_k^{-1}) \|u^k\|^2 - \|u^k\| \|\boldsymbol{\Lambda}_k^{-1} e^k\| && [\boldsymbol{\Pi}_k \text{ orthonormal }] \\ &\geq \kappa(\boldsymbol{\Lambda}_k^{-1}) \|u^k\|^2 - \sigma\kappa(\boldsymbol{\Lambda}_k^{-1}) \|u^k\|^2 && [\text{ using (76) }] \\ &= (1-\sigma)\kappa(\boldsymbol{\Lambda}_k^{-1}) \|u^k\|^2 \\ &\geq (1-\sigma)\zeta \|u^k\|^2. && [\text{ using (67) }] \end{aligned} \tag{79}$$

We need to convert this lower bound on $\varphi_k(p^k)$ expressed in terms of $\|u^k\|^2$, to one expressed in terms of $\|\nabla\varphi_k\|^2$, so we can meet hypothesis 3 of Proposition 3.4. First, starting with (78) we obtain

$$\begin{aligned}
\|v^k\|^2 &= \|(\boldsymbol{\Pi}_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Pi}_k^\top)u^k - \boldsymbol{\Pi}_k\boldsymbol{\Lambda}_k^{-1}e^k\|^2 \\
&\leq \left(\|(\boldsymbol{\Pi}_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Pi}_k^\top)u^k\| + \|\boldsymbol{\Pi}_k\boldsymbol{\Lambda}_k^{-1}e^k\|\right)^2 && [\text{ triangle inequality }] \\
&\leq \left(\|\boldsymbol{\Pi}_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Pi}_k^\top\|\|u^k\| + \|\boldsymbol{\Lambda}_k^{-1}e^k\|\right)^2 && [\text{ using (59) }] \\
&\leq \left(\|\boldsymbol{\Lambda}_k^{-1}\|\|u^k\| + \sigma\kappa(\boldsymbol{\Lambda}_k^{-1})\|u^k\|\right)^2 && [\text{ using (76) }] \\
&\leq \left((1+\sigma)\,\|\boldsymbol{\Lambda}_k^{-1}\|\|u^k\|\right)^2 && [\kappa(\boldsymbol{\Lambda}_k^{-1}) \leq \|\boldsymbol{\Lambda}_k^{-1}\|] \\
&\leq \left((1+\sigma)\beta\,\|u^k\|\right)^2 && [\text{ using (67)}] \\
&= (1+\sigma)^2\beta^2\,\|u^k\|^2.
\end{aligned} \tag{80}$$

Now, from (72) in Lemma 3.4 and (80) follows that

$$\|\nabla\varphi_k\| \leq \left(n(1+\sigma)^2\beta^2 + 1\right)\|u^k\|^2.$$

Combining this with the lower bound for $\varphi_k(p^k)$ in (79) yields

$$\begin{aligned}
\varphi_k(p^k) &\geq (1-\sigma)\zeta\|u^k\|^2 \\
&\geq (1-\sigma)\zeta\left[\frac{\|\nabla\varphi_k\|}{n(1+\sigma)^2\beta^2+1}\right] \\
&= \frac{(1-\sigma)\zeta}{n(1+\sigma)^2\beta^2+1}\|\nabla\varphi_k\|.
\end{aligned} \tag{81}$$

Hence, taking

$$\xi = \frac{(1-\sigma)\zeta}{n(1+\sigma)^2\beta^2+1} > 0,$$

as in (81), we obtain Hypothesis 3 of Proposition 3.4, hence $\varphi_k(p^k) \to 0$. Then, we deduce from (79) that $u^k \to 0$, and by (80) this implies that $v^k \to 0$. Thus, Hypotheses 5 and 6 of Proposition 3.4 are satisfied. Then, in virtue of this proposition follows the weak convergence of the sequences $(z^k)_{k\geq 0}$, $(x_i^k)_{k\geq 0}$, $(y_i^k)_{k\geq 0}$, and $(w_i^k)_{k\geq 0}$, for $i = 1, \ldots, n$. □

Remark 3.2. Note that (81) implies that $\varphi_k(p^k)$ is always nonnegative so there is no need of the operation max{0,·} when computing the fraction in the update (29). This is reflected in Algorithm 3, in the computation of the formulas (64)-(66).

4 PROJECTIVE SPLITTING WITH ONE FORWARD STEP FOR COCOERCIVE OPERATORS

In the previous chapter we saw how the general separator-projector framework was applied to an inclusion problem. The resulting algorithm performs only backward steps, and a convergence condition was presented. This chapter, based on (JOHNSTONE; ECKSTEIN, 2021), it is considered an inclusion problem involving cocoercive operators. The problem is presented in Section 4.1. Section 4.2 shows the construction of the extended solution set and the separators for this particular problem. Section 4.3 shows a connection between the way the points are chosen to construct the separators and the FB algorithm. Section 4.4 presents a projective splitting algorithm for the inclusion problem considered in this chapter. Finally, Section 4.5 contains the necessary results to prove the convergence of the algorithm to a solution point.

4.1 PROBLEM STATEMENT

For a collection of real Hilbert spaces $\{\mathcal{H}_i\}_{i=0}^n$ consider the *finite-sum convex minimization problem*:

$$\min_{z\in\mathcal{H}_0} \sum_{i=1}^{n} \big(f_i(G_i z) + h_i(G_i z)\big), \tag{82}$$

where every $f_i \in \Gamma_0(\mathcal{H}_i)$, every $h_i \in \Gamma_0(\mathcal{H}_i)$ is also differentiable with L_i-Lipschitz-continuous gradients, and the operators $G_i \in \mathcal{B}(\mathcal{H}_0,\mathcal{H}_i)$. Under appropriate constraint qualifications, (82) is equivalent to the monotone inclusion problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* \left(A_i + B_i\right) G_i z \tag{83}$$

where all $A_i : \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ and $B_i : \mathcal{H}_i \to \mathcal{H}_i$ are maximal monotone and each B_i is L_i^{-1}-cocoercive. Notice that when $L_i = 0$, B_i must be a constant operator, that is, there is some $v_i \in \mathcal{H}_i$ such that $B_i x = v_i$ for all $x \in \mathcal{H}_i$. Example 2.8 presents an application of some constraint qualification conditions to turn a problem like (82) into one of the form (83).

Defining $T_i = A_i + B_i$ for all i, problem (83) may be written as

$$0 \in \sum_{i=1}^{n} G_i^* T_i G_i z. \tag{84}$$

This more compact problem statement will be used occasionally in our analysis below. We collect here the main assumptions regarding to problem (83).

Assumption 1. *Problem (83) conforms to the following:*

1. $\mathcal{H}_0 = \mathcal{H}_n$ *and* $\mathcal{H}_1, \ldots, \mathcal{H}_{n-1}$ *are real Hilbert spaces.*
2. *For* $i = 1, \ldots, n$*, the operators* $A_i : \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ *and* $B_i : \mathcal{H}_i \to \mathcal{H}_i$ *are monotone. Additionally each* A_i *is maximal.*
3. *Each operator* B_i *is either* L_i^{-1}*-cocoercive for some* $L_i > 0$ *(and thus single-valued) and* $\operatorname{dom} B_i = \mathcal{H}_i$*, or* $L_i = 0$ *and* $B_i x = v_i$ *for all* $x \in \mathcal{H}_i$ *and some* $v_i \in \mathcal{H}_i$*, that is,* B_i *is a constant function.*
4. *Each* $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ *for* $i = 1, \ldots, n-1$ *is linear and bounded.*
5. *Problem (83) has a solution.*

We denote by $\boldsymbol{\mathcal{H}}$ the product space

$$\boldsymbol{\mathcal{H}} = \mathcal{H}_0 \times \cdots \times \mathcal{H}_{n-1},$$

and we will denote any point of $\boldsymbol{\mathcal{H}}$ as

$$p = (z, \boldsymbol{w}) = (z, w_1, \ldots, w_{n-1}), \text{ thus } \boldsymbol{w} = (w_1, \ldots, w_{n-1}).$$

For $\boldsymbol{\mathcal{H}}$, we adopt the following norm and inner product for some $\gamma > 0$:

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

4.2 CONSTRUCTION OF A EXTENDED SOLUTION SET AND THE SEPARATORS

Our first goal is to construct a separator-projector algorithm as in Chapter 3, so that we obtain all the discussed properties of the generated sequence as in Proposition 3.1.

To that end, we would like to devise a construction of a extended solution set and separators inherent to the problem we are considering. We start by making the assumption that there exist a $z \in \mathcal{H}$ such that

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

In what it follows we will impose the assumption that $G_n = I$, this does not represent a restriction since one could redefine the last operator as $T_n = G_n^* \circ T_n \circ G_n^*$, or one could simply append a new operator T_n with $T_n z = \{0\}$ everywhere. Setting $w_i \in T_i G_i z$ we can rewrite the inclusion as follows

$$0 \in \sum_{i=1}^{n-1} G_i^* w_i + T_n z$$

from where we obtain

$$-\sum_{i=1}^{n-1} G_i^* w_i \in T_n z.$$

Hence, we define the *extended solution set* as

$$\mathcal{S}_e := \left\{ (z, w_1, \ldots, w_{n-1}) \in \boldsymbol{\mathcal{H}} \mid w_i \in T_i G_i z, i = 1, \ldots, n-1, -\sum_{i=1}^{n-1} G_i^* w_i \in T_n z \right\} \quad (85)$$

In the previous construction we set $w_n := -\sum_{i=1}^{n-1} G_i^* w_i$, therefore, we may say that $(z, w_n) \in \operatorname{gra} T_n$.

Lemma 4.1. *Suppose Assumption 1 holds. The set $\mathcal{S}_e$ defined in (85) is closed and convex.*

Proof. First, Assumption 1(5) and the construction shown of the set $\mathcal{S}_e$ allow us to conclude that $\mathcal{S}_e \neq \emptyset$. By Proposition 2.11 each B_i is maximal. Since $\operatorname{dom} B_i = \mathcal{H}_i$ applying Proposition 2.5 we obtain that $T_i = A_i + B_i$ is maximal monotone for $i = 1, \ldots, n$. Now, to prove that $\mathcal{S}_e$ is closed and convex we are going to relate it with the set of zeros of a maximal monotone operator, just as in Proposition 3.3. To that end, consider $A = T_n$ and $B = T_1 \times \cdots \times T_{n-1}$ and $L : z \to (G_1 z, \ldots, G_{n-1} z)$, then $L^*(w_1, \ldots, w_n) = \sum_{i=1}^{n-1} G_i^* w_i$ as proved in Example 2.2.

Let $(z, \boldsymbol{w}) \in \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S})$, then

$$\begin{aligned}
0 \in \boldsymbol{M}(z, \boldsymbol{w}) + \boldsymbol{S}(z, \boldsymbol{w}) &\Leftrightarrow 0 \in (Az + L^* \boldsymbol{w}) \times (B^{-1} \boldsymbol{w} - Lz) \\
&\Leftrightarrow 0 \in T_n z + L^* \boldsymbol{w} \;\text{ and }\; Lz \in B^{-1} \boldsymbol{w} \\
&\Leftrightarrow -L^* \boldsymbol{w} \in T_n z \;\text{ and }\; \boldsymbol{w} \in BLz \\
&\Leftrightarrow -L^* \boldsymbol{w} \in T_n z \;\text{ and }\; w_i \in T_i G_i z \\
&\Leftrightarrow (z, \boldsymbol{w}) \in \mathcal{S}_e.
\end{aligned}$$

Consequently, $\mathcal{S}_e$ is closed and convex. □

Now we show how to construct the separators. Let $(z, w_1, \ldots, w_{n-1}) \in \mathcal{S}_e$, from the construction of the extended solution set we had $(G_i z, w_i) \in \operatorname{gra} T_i$ for $i = 1, \ldots, n$. Suppose that for each $i = 1, \ldots, n$, we get a point $(x_i, y_i) \in \operatorname{gra} T_i$, we have from the monotonocity of each T_i that

$$\langle G_i z - x_i, y_i - w_i \rangle \leq 0.$$

Summing over i we obtain

$$\sum_{i=1}^{n} \langle G_i z - x_i, y_i - w_i \rangle \leq 0. \quad (86)$$

This leads to the following

Definition 4.1. Given the points $(x_i^k, y_i^k) \in \operatorname{gra} T_i$, for $i = 1, \dots, n-1$, we define the function $\varphi_k : \mathcal{H} \to \mathbb{R}$ as

$$\varphi_k(z, w_1, \dots, w_n) := \sum_{i=1}^{n} \langle G_i z - x_i^k, y_i^k - w_i \rangle. \tag{87}$$

As we saw in (86) this function has the property of being non-positive in $\mathcal{S}_e$. Some additional properties of this function are presented in the following

Lemma 4.2. *Let φ_k be defined as in (87). Then:*

1. *φ_k is affine on $\mathcal{H}$.*
2. *With respect to inner product $\langle \cdot, \cdot \rangle_\gamma$ on $\mathcal{H}$, the gradient of φ_k is*

$$\nabla \varphi_k = \left(\frac{1}{\gamma} \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), x_1^k - G_1 x_n^k, x_2^k - G_2 x_n^k, \dots, x_{n-1}^k - G_{n-1} x_n^k \right). \tag{88}$$

Proof. Separating terms and using the adjoint of each G_i, the fact that $\sum_{i=1}^{n} G_i^* w_i = 0$, and the definition of w_n we have

$$\begin{aligned}
\varphi_k(z, w_1, \dots, w_{n-1}) &= \sum_{i=1}^{n} \langle G_i z, y_i^k - w_i \rangle - \sum_{i=1}^{n} \langle x_i^k, y_i^k - w_i \rangle \\
&= \left\langle z, \sum_{i=1}^{n} G_i^* y_i^k - \sum_{i=1}^{n} G_i^* w_i \right\rangle - \sum_{i=1}^{n} \langle x_i^k, y_i^k \rangle + \sum_{i=1}^{n} \langle x_i^k, w_i^k \rangle \\
&= \left\langle z, \sum_{i=1}^{n} G_i^* y_i^k \right\rangle - \sum_{i=1}^{n} \langle x_i^k, y_i^k \rangle + \sum_{i=1}^{n-1} \langle x_i^k, w_i \rangle - \left\langle x_n^k, \sum_{i=1}^{n-1} G_i^* w_i \right\rangle.
\end{aligned}$$

It follows that

$$\varphi_k(z, w_1, \dots, w_{n-1}) = \left\langle z, \sum_{i=1}^{n} G_i^* y_i^k \right\rangle + \sum_{i=1}^{n-1} \langle x_i^k - G_i x_n^k, w_i \rangle - \sum_{i=1}^{n} \langle x_i^k, y_i^k \rangle. \tag{89}$$

This equation allows to conclude that φ_k is affine. Now, fix an arbitrary $\tilde{p} \in \mathcal{H}$. Using that φ is affine, we may write

$$\begin{aligned}
\varphi_k(p) &= \langle p - \tilde{p}, \nabla \varphi_k \rangle_\gamma + \varphi_k(\tilde{p}) = \langle p, \nabla \varphi_k \rangle_\gamma + \varphi_k(\tilde{p}) - \langle \tilde{p}, \nabla \varphi_k \rangle_\gamma \\
&= \gamma \langle z, \nabla_z \varphi_k \rangle + \sum_{i=1}^{n-1} \langle w_i, \nabla_{w_i} \varphi_k \rangle + \varphi_k(\tilde{p}) - \langle \tilde{p}, \nabla \varphi_k \rangle_\gamma.
\end{aligned}$$

Equating terms between this expression and (89) yields the claimed expression for the gradient in (88). □

We also use the following notation for $i = 1,\dots,n$:

$$\varphi_{i,k}(z,w_i) := \langle G_i z - x_i^k, y_i^k - w_i\rangle.$$

Note that $\varphi_k(z,w_1,\dots,w_{n-1}) = \sum_{i=1}^n \varphi_{i,k}(z,w_i)$.

4.3 THE NEW PROCEDURE

4.3.1 A Connection with the Forward-Backward Method

Just as it was important in Chapter 3 the way the points are chosen, here we have an specific situation where each T_i is the sum of two maximal monotone operators. In (JOHNSTONE; ECKSTEIN, 2021) it was proposed that at each iteration k and for each $i = 1,\dots,n$ to find a pair $(x_i^k,y_i^k) \in \text{gra}\, T_i = \text{gra}(A_i + B_i)$ conforming the conditions

$$t = (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(B_i x_i^{k-1} - w_i^k) \tag{90}$$

$$x_i^k = J_{\rho A}(t) \tag{91}$$

$$a_i^k = (1/\rho_i)(t - x_i^k) \tag{92}$$

$$b_i^k = B_i x_i^k \tag{93}$$

$$y_i^k = a_i^k + b_i^k. \tag{94}$$

where $\alpha_i \in (0,1)$, $\rho_i \leq 2(1-\alpha_i)/L$ and $b_i^0 = B_i x_i^0$. A resolvent calculation gives us (91), and (92) follows from the relation in (9). To obtain (93) is required only an evaluation (forward step) on B_i, and (94) is a simple vector addition.

Now we show how this proposed updated is related to the FB algorithm. As in (JOHNSTONE; ECKSTEIN, 2020) the pairs $(x_i^k, y_i^k) \in \text{gra}\, T_i$ are solutions of

$$x_i^k + \rho_i y_i^k = G_i z^k + \rho_i w_i^k : \quad y_i^k \in T_i x_i^k \tag{95}$$

for some $\rho_i > 0$, which lead us to a resolvent calculation. Notice the similarity with the way was done in the previous chapter in (54) which leads to $\varphi_k(p^k) > 0$. Now, with problem (83) we have $T_i = A_i + B_i$, with B_i being cocoercive and A_i maximal monotone. For T_i in this form, computing the resolvent as in (95) exactly may be impossible, even when the resolvent of A_i is available. We would like to take advantage of the cocoercivity of each B_i. To that end, since the stepsize ρ_i in (95) can be any positive number, let us replace ρ_i with ρ_i/α_i for some $\alpha_i \in (0,1)$ and rewrite (95) as

$$x_i^k + \frac{\rho_i}{\alpha_i} y_i^k = G_i z^k + \frac{\rho_i}{\alpha_i} w_i^k : \quad y_i^k \in T_i x_i^k. \tag{96}$$

With this structure, x_i^k in (96) satisfies:

$$\begin{aligned} 0 &= \frac{\rho_i}{\alpha_i} y_i^k + x_i^k - \left(G_i z^k + \frac{\rho_i}{\alpha_i} w_i^k\right) \\ \implies 0 &\in \frac{\rho_i}{\alpha_i} A_i x_i^k + \frac{\rho_i}{\alpha_i} B_i x_i^k + x_i^k - \left(G_i z^k + \frac{\rho_i}{\alpha_i} w_i^k\right) \end{aligned}$$

which can be rearranged to $0 \in A_i x_i^k + \tilde{B}_i x_i^k$,where

$$\tilde{B}_i v = B_i v + \frac{\alpha_i}{\rho_i}\left(v - G_i z^k - \frac{\rho_i}{\alpha_i} w_i^k\right).$$

Since B_i is L_i^{-1}-cocoercive, $\tilde{B}_i$ is $(L_i + \alpha_i/\rho_i)^{-1}$-cocoercive by Proposition 2.12. Consider the generic monotone inclusion problem $0 \in A_i x + \tilde{B}_i x$ where A_i is maximal and $\tilde{B}_i$ is cocoercive, and thus one may solve the problem with the FB method as in Theorem 2.8. If one applies a single iteration of FB initialized at x_i^{k-1}, with stepsize ρ_i, to the inclusion $0 \in A_i x + \tilde{B}_i x$, one obtains the calculation:

$$\begin{aligned} x_i^k &= J_{\rho_i A_i}\left(x_i^{k-1} - \rho_i \tilde{B}_i x_i^{k-1}\right) \\ &= J_{\rho_i A_i}\left(x_i^{k-1} - \rho_i\left(B_i x_i^{k-1} + \frac{\alpha_i}{\rho_i}\left(x_i^{k-1} - G_i z^k - \frac{\rho_i}{\alpha_i} w_i^k\right)\right)\right) \\ &= J_{\rho_i A_i}\left((1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(B_i x_i^{k-1} - w_i^k)\right). \end{aligned}$$

So, the proposed calculation is equivalent to *one* iteration of FB initialized at the previous point x_i^{k-1}, applied to the subproblem of computing the resolvent in (96). Prior versions of projective splitting require computing this resolvent either exactly or to within a certain relative error criterion, which may be time consuming. Here, a simple single FB step is made toward computing the resolvent which we will prove is sufficient for the projective splitting method to converge to $\mathcal{S}_e$. Note, however, that the step size restriction in Assumption 2 is stronger than the natural stepsize limit that would arise when applying FB to $0 \in A_i x + \tilde{B}_i x$, which would be

$$\rho_i < \frac{2-\alpha_i}{L_i}.$$

4.4 ALGORITHM DEFINITION

We introduce here a notation for the one-forward-backward step update as follows:

Definition 4.2. Suppose $\mathcal{H}$ and $\mathcal{H}'$ are real Hilbert spaces, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone with nonempty domain, $B : \mathcal{H} \to \mathcal{H}$ is L^{-1}-cocoercive, and $G : \mathcal{H}' \to \mathcal{H}$ is bounded and linear. For $\alpha \in [0,1]$ and $\rho > 0$, define the mapping $\mathcal{F}_{\alpha,\rho}(z,x,w;A,B,G) : \mathcal{H}' \times \mathcal{H}^2 \to \mathcal{H}^2$, with additional parameters A, B, and G, as

$$\mathcal{F}_{\alpha,\rho}\left(\begin{array}{c} z,x,w; \\ A,B,G \end{array}\right) := (x^+,y^+) : \begin{cases} t & := (1-\alpha)x + \alpha G z - \rho(Bx - w) \\ x^+ & = J_{\rho A}(t) \\ y^+ & = \rho^{-1}(t - x^+) + Bx^+. \end{cases} \tag{97}$$

The expression for y^+ follows from (94) and (92). Notice that this avoids the evaluation of Ax^+. To simplify the presentation, we will also use the notation

$$\mathcal{F}^i(z,x,w) := \mathcal{F}_{\alpha_i,\rho_i}(z,x,w;A_i,B_i,G_i). \tag{98}$$

With this notation the step (90)-(94) can be written as

$$(x_i^k, y_i^k) = \mathcal{F}^i(z^k, x_i^{k-1}, w^k). \tag{99}$$

Now, we state the projective splitting algorithm for problem (83).

Algorithm 4: One-Forward-Step Projective Splitting

Data: $p^1 = (z^1, \boldsymbol{w}^1) \in \boldsymbol{\mathcal{H}}$, $\gamma > 0$, $\delta \in (0,1)$, and $\hat{\rho}$. For $i = 1, \ldots, n$: $x_i^0 \in \mathcal{H}_i$ and $0 < \alpha_i \leq 1$ and $\rho_i > 0$.

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

2 Compute $(x_i^k, y_i^k) = \mathcal{F}^i(z^k, x_i^{k-1}, w_i^k)$ with $\mathcal{F}^i$ defined in (97)

/* Projection starts */

3 $u_i^k = x_i^k - G_i x_n^k$, $i = 1, \ldots, n-1$

4 $v^k = \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k$

5 $\pi_k = \|u^k\|^2 + \gamma^{-1}\|v^k\|^2$

6 **if** $\pi_k > 0$ **then**

7 $\varphi_k(p^k) = \langle z^k, v^k \rangle + \sum_{i=1}^{n-1} \langle w_i^k, u_i^k \rangle - \sum_{i=1}^{n} \langle x_i^k, y_i^k \rangle$

8 $\tau = \dfrac{1}{\pi_k} \max\{0, \varphi_k(p^k)\}$

9 $z^{k+1} = z^k - \gamma^{-1}\tau v$

10 $w_i^{k+1} = w_i^k - \tau u_i^k$ $i = 1, \ldots, n-1$

11 $w_n^{k+1} = -\sum_{i=1}^{n-1} G_i^* w_i^{k+1}$

12 **else**

13 **return** $(x_n^k, y_1^k, \ldots, y_n^k)$

14 **end**

15 **end**

Line 5 computes the squared norm of the gradient of the separator expressed in (88) using the norm $\|\cdot\|_\gamma$.

The stepsizes ρ_i for $i = 1, \ldots, n$ are fixed across all iterations, satisfying Assumption 2. The same applies for the averaging parameter α_i.

The parameter $\gamma > 0$ allows for the projection to be performed using a slightly more general primal-dual metric than (37). In effect, this parameter changes the relative size of the primal and dual updates in lines 9-10 of Algorithm 4. As γ increases, a smaller step is taken in the primal and a larger step in the dual. As γ decreases, a smaller step is taken in the dual update and a larger step is taken in the primal. See (ECKSTEIN; SVAITER, 2009, Sec. 5.1) and (ECKSTEIN; SVAITER, 2007, Sec. 4.1) for more details.

Unlike the algorithm in the previous chapter, where was ensured that $\varphi_k(p^k)$ is always nonnegative, here will be established an “ascent lemma” that relates the values $\varphi_k(p^k)$ and $\varphi_{k-1}(p^{k-1})$ in such a way that overall convergence may still be proved, even though it is possible that $\varphi_k(p^k) \leq 0$ at some iterations k. In particular, $\varphi_k(p^k)$ will be larger than the previous value $\varphi_{k-1}(p^{k-1})$, up to some error term that vanishes as $k \to \infty$.

Remark 4.1. The algorithm presented in (JOHNSTONE; ECKSTEIN, 2021) contains a backtracking linesearch for those operators B_i with unknown cocoercivity constant. Here with the objective of simplifying the analysis, we opted to assume that all cocoercivity constants are known, so Algorithm 4 is presented without the backtracking linesearch.

4.4.1 Separator projector properties

Lemma 4.3 details the key results for Algorithm 4 that stem from it being a separator-projector algorithm. While these properties alone do not guarantee convergence, they are important to all of the arguments that follow.

Lemma 4.3. *Suppose that Assumption 1 holds. Then for Algorithm 4*

1. *The sequence* $(p^k)_{k\geq 0}$*, where* $p^k = (z^k, w_1^k, \dots, w_{n-1}^k)$ *is bounded.*
2. *If the algorithm never terminates via line 13,* $p^k - p^{k+1} \to 0$*. Furthermore* $z^k - z^{k-1} \to 0$ *and* $w_i^k - w_i^{k-1} \to 0$ *for* $i = 1, \dots n$*.*
3. *If the algorithm never terminates via line 13 and* $\|\nabla\varphi_k\|_\gamma$ *remains bounded for all* $k \geq 1$*, then* $\limsup_{k\to\infty} \varphi_k(p^k) \leq 0$.

Proof. 1. Notice that line 2 computes for each $i = 1, \dots, n$ a point (x_i^k, y_i^k) in gra$(A_i + B_i)$ according to (98), while lines 7-11 computes the projection of the current iterate $(z^k, w_1^k, \dots, w_n^k)$ onto the hyperplane according to the definition of the separator in (89) and to the basic projection Algorithm 2. Therefore the sequence produced by Algorithm 4 satisfies the properties listed in Proposition 3.1, specifically we obtain that the sequence $(p^k)_{k\geq 0} = (z^k, w_1^k, \dots, w_n^k)_{k\geq 0}$ is a Fejér monotone sequence with respect to $\mathcal{S}_e$, and thus is a bounded sequence. Additionally, we can write

$$p^{k+1} = p^k - \frac{\max\{0, \varphi_k(p^k)\}}{\|\nabla\varphi_k\|_\gamma^2}\nabla\varphi_k.$$

2. If Algorithm 4 never terminates via line 13 this means that $\pi_k \neq 0$, which implies that lines 7-11 were executed. This generates an infinite sequence that as in the previous item is a Fejér monotone sequence, by Proposition 3.1(4) we have from the convergence of the series in (32) that $\|p^k - p^{k+1}\|^2 \to 0$. We conclude from this that

$$z^k - z^{k+1} \to 0 \text{ and } w_i^k - w_i^{k+1} \to 0 \text{ for } i = 1, \dots, n.$$

3. Suppose that $\|\nabla\varphi_k\|_\gamma \leq \xi \; \forall\, k$, from

$$p^{k+1} = p^k - \frac{\max\{\varphi_k(p^k), 0\}}{\|\nabla\varphi_k\|_\gamma^2}\nabla\varphi_k,$$

follows that

$$0 = \lim_{k\to\infty} \|p^{k+1} - p^k\| = \lim_{k\to\infty} \frac{\max\{\varphi_k(p^k),0\}}{\|\nabla\varphi_k\|_\gamma}$$
$$\geq \limsup_{k\to\infty} \frac{\varphi_k(p^k)}{\xi}.$$

Therefore, $\limsup_{k\to\infty} \varphi_k(p^k) \leq 0$.

□

We now precisely state our stepsize assumption for the manually chosen stepsizes, as well as the stepsize upper bound $\hat{\rho}$.

Assumption 2. *If $L_i > 0$, then $0 < \rho_i \leq 2(1-\alpha_i)/L_i$. The parameter $\hat{\rho}$ must satisfy*

$$\hat{\rho} \geq \rho_i. \tag{100}$$

Note that if $L_i > 0$, Assumption 2 effectively limits α_i to be strictly less than 1, otherwise the stepsize ρ_i would be forced to 0, which is prohibited. In this case α_i must be chosen in (0,1). On the other hand, if $L_i = 0$, there is no constraint on ρ_i other than that it is positive and nonzero, and in this case α_i may be chosen in (0,1].

4.5 MAIN PROOF

The following lemma contains a condition to ensure weak convergence. Hence, the core of the proof strategy will be to establish (101) below.

Lemma 4.4. *Suppose Assumption 1 holds and Algorithm 4 produces an infinite sequence of iterations without terminating via Line 13. If*

$$(\forall i = 1,\ldots,n): \quad y_i^k - w_i^k \to 0 \text{ and } G_i z^k - x_i^k \to 0, \tag{101}$$

then there exists $(\bar{z},\overline{\mathbf{w}}) \in \mathcal{S}_e$ such that $(z^k,\mathbf{w}^k) \rightharpoonup (\bar{z},\overline{\mathbf{w}})$. Furthermore, we also have $x_i^k \rightharpoonup G_i\bar{z}$ and $y_i^k \rightharpoonup \overline{w}_i$ for all $i = 1,\ldots,n-1$, $x_n^k \rightharpoonup \bar{z}$, and $y_n^k \rightharpoonup -\sum_{i=1}^{n-1} G_i^*\overline{w}_i$.

Proof. The strategy of the proof follows a similar pattern to the second part of the proof of Proposition 3.4, however this proof is a little more elaborated.

The first tool is to use Lemma 2.2, that is, we need to show that any weak cluster point of the sequence $(p^k)_{k\geq 0}$ belongs to the set $\mathcal{S}_e$. Consider a weak cluster point of $(p^k)_{k\geq 0}$ which exists by Lemma 2.1, hence there exists an increasing sequence of indices $(k_m)_{m\geq 0}$ such that

$$(z^{k_m},\mathbf{w}^{k_m}) \rightharpoonup (z^\infty,\mathbf{w}^\infty) \in \boldsymbol{\mathcal{H}}. \tag{102}$$

In what follows, we consider the subsequences as a sequences, renaming if necessary the indices. In order to apply Theorem 2.3 we consider the sequences

$$\boldsymbol{x}^k = (x_1^k, \dots, x_{n-1}^k) \text{ and } \boldsymbol{y}^k = (y_1^1, \dots, y_{n-1}^k).$$

Recall that they satisfy $y_i^k \in T_i x_i^k$ for $i = 1, \dots, n-1$. From $y_i^k - w_i^k \to 0$ and $w_i^k \rightharpoonup w_i^\infty$ follow that $y_i^k \rightharpoonup w_i^\infty$ for $i = 1, \dots, n$. Similarly, from $G_i z^k - x_i^k \to 0$, $z^k \rightharpoonup z^\infty$ and the boundedness of G_i follow that $x_i^k \rightharpoonup G_i z^\infty$ for $i = 1, \dots, n$.

Now we establish the second part of the Theorem 2.3. Notice that from the first part of the hypothesis in (101), and the boundedness of G_i imply that

$$\sum_{i=1}^{n} G_i^* y_i^k = \sum_{i=1}^{n} G_i^* w_i^k + \sum_{i=1}^{n} G_i^* (y_i^k - w_i^k) \to 0. \tag{103}$$

In addition, the second part of (101), and the boundedness of G_i yield

$$x_i^k - G_i x_n^k = x_i^k - G_i z^k - G_i(x_n^k - z^k) \to 0, \quad \forall\, i = 1, \dots, n-1. \tag{104}$$

Next, we define a projection over a closed subspace, to this end we will follow (ALOTAIBI; COMBETTES; SHAHZAD, 2014, Prop. 2.4) applied to this context. Let $L : \mathcal{H}_0 \to \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ defined by

$$L : z \to (G_1 z, \dots, G_{n-1} z)$$

then by Example 2.2 we have that $L^* : \mathcal{H}_1 x \cdots x \mathcal{H}_{n-1} \to \mathcal{H}_0$ is given by

$$L^*(w_1, w_2, \dots, w_{n-1}) = \sum_{i=1}^{n-1} G_i^* w_i.$$

Notice that in terms of L and L^* we can write (103) and (104) as

$$y_n^k + L^* \boldsymbol{y}^k \to 0 \text{ and } \boldsymbol{x}^k - L x_n^k \to 0, \tag{105}$$

respectively. Setting $\mathcal{K} = \mathcal{H}_0 \times \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ and recalling that $\boldsymbol{w} = (w_1, \cdots, w_{n-1})$ we define

$$V = \{(z, \boldsymbol{w}) \in \mathcal{K} : Lz = \boldsymbol{w}\},$$

its orthogonal complement as seen in (15) is

$$V^\perp = \{(z, \boldsymbol{w}) \in \mathcal{K} \mid z = -L^* \boldsymbol{w}\}.$$

Using the expressions for P_V and $P_{V^\perp}$ in Proposition 2.10 and (105) we obtain that

$$P_V(y_n^k, \boldsymbol{y}^k) = ((I + LL^*)^{-1}(y_n^k + L^* \boldsymbol{y}^k), L(I + L^*L)^{-1}(y_n^k + L^* \boldsymbol{y}^k)) \to 0,$$

and

$$(I - P_V)(x_n^k, \boldsymbol{x}^k) = P_{V^\perp}(x_n^k, \boldsymbol{x}^k) = (L^*(I + LL^*)^{-1}(L x_n^k - \boldsymbol{x}^k), -(I + LL^*)^{-1}(L x_n^k - \boldsymbol{x}^k)) \to 0.$$

Altogether, since L and L^* are weakly continuous we have

$$\begin{aligned}
&((x_n^k, \boldsymbol{x}^k), (y_n^k, \boldsymbol{y}^k)) \in \operatorname{gra} T_n \times T_1 \times \cdots \times T_{n-1} \\
&(x_n^k, \boldsymbol{x}^k) \rightharpoonup (z^\infty, G_1 z^\infty \dots, G_{n-1} z^\infty) = (z^\infty, Lz^\infty) \\
&(y_n^k, \boldsymbol{y}^k) \rightharpoonup (-L^* \boldsymbol{w}^\infty, \boldsymbol{w}^\infty) \\
&P_{V^\perp}(x_n^k, \boldsymbol{x}^k) \to 0 \\
&P_V(y_n^k, \boldsymbol{y}^k) \to 0
\end{aligned}$$

Denoting $T_n \times T_1 \times \cdots \times T_{n-1}$ by T, it follows from Theorem 2.3 that

$$\big((z^\infty, Lz^\infty), (-L^* \boldsymbol{w}^\infty, \boldsymbol{w}^\infty)\big) \in (V \times V^\perp) \cap \operatorname{gra} T.$$

Since $((z^\infty, Lz^\infty), (-L^* \boldsymbol{w}^\infty, \boldsymbol{w}^\infty)) \in \operatorname{gra} T$ we have

$$-\sum_{i=1}^{n-1} G_i^* w_i^\infty = -L^* \boldsymbol{w}^\infty \in T_n z^\infty \text{ and } w_i^\infty \in T_i G_i z^\infty \text{ for } i = 1, \dots, n-1.$$

This implies by the definition of $\mathcal{S}_e$ in (85) that $(z^\infty, w_1^\infty, \dots, w_{n-1}^\infty) \in \mathcal{S}_e$. We have established that the weak cluster point $(z^\infty, \boldsymbol{w}^\infty)$ belongs to $\mathcal{S}_e$, since this point was arbitrary we can conclude via Lemma 2.2 that the whole sequence $(z^k, \boldsymbol{w}^k)_{k \geq 0}$ converges weakly to some $(\bar{z}, \overline{\boldsymbol{w}}) \in \mathcal{S}_e$. For each $i = 1, \dots, n$, we finally observe that since $G_i z^k - x_i^k \to 0$ and $y_i^k - w_i^k \to 0$, we also have $x_i^k \rightharpoonup G_i \bar{z}$ and $y_i^k \rightharpoonup \overline{w}_i$.

□

In order to establish (101), we start by establishing certain contractive and "ascent" properties for the mapping $\mathcal{F}$ in Lemmas 4.9 and 4.13. Then, we prove the boundedness of x_i^k and y_i^k, in turn yielding the boundedness of the gradients $\nabla \varphi_k$ and hence the result that $\limsup \nabla \varphi_k \leq 0$ by Lemma 4.3. Next we establish a "Lyapunov-like" recursion for $\varphi_i^k(z^k, w_i^k)$, relating $\varphi_i^k(z^k, w_i^k)$ to $\varphi_i^{k-1}(z^{k-1}, w_i^{k-1})$. Eventually this result will allow us to establish that $\liminf \nabla \varphi_k \geq 0$ and hence that $\lim \nabla \varphi_k = 0$, which will in turn allow an argument that $y_i^k - w_i^k \to 0$. The proof that $G_i z^k - x_i^k \to 0$ will then follow fairly elementary arguments.

4.5.1 Some Basic Results

We begin by stating three elementary results on sequences, which may be found in (POLYAK, 1987), and a basic, well known nonexpansivity property for forward steps with cocoercive operators.

Lemma 4.5. *Suppose that $a_k \geq 0$ for all $k \geq 1$, $b \geq 0$, $0 \leq \tau < 1$, and $a_{k+1} \leq \tau a_k + b$ for all $k \geq 1$. Then $\{a_k\}$ is a bounded sequence.*

Lemma 4.6. *Suppose that $a_k \geq 0, b_k \geq 0$ for all $k \geq 1$, $b_k \to 0$, and there is some $0 \leq \tau < 1$ such that $a_{k+1} \leq \tau a_k + b_k$ for all $k \geq 1$. Then $a_k \to 0$.*

Lemma 4.7. *Suppose that* $0 \leq \tau < 1$ *and* $\{r_k\},\{b_k\}$ *are sequences in* $\mathbb{R}$ *with the properties* $b_k \to 0$ *and* $r_{k+1} \geq \tau r_k + b_k$ *for all* $k \geq 1$*. Then* $\liminf_{k\to\infty}\{r_k\} \geq 0$*.*

Proof. Negating the assumed inequality yields $-r_{k+1} \leq \tau(-r_k) - b_k$. Applying Lemma 4.6 then yields $\limsup\{-r_k\} \leq 0$. □

Lemma 4.8. *Suppose B is* L^{-1}*-cocoercive and* $0 \leq \rho \leq 2/L$*. Then for all* $x,y \in \text{dom}(B)$

$$\|x - y - \rho(Bx - By)\| \leq \|x - y\|. \tag{106}$$

Proof. Squaring the left hand side of (106) yields

$$\begin{aligned}\|x - y - \rho(Bx - By)\|^2 &= \|x - y\|^2 - 2\rho\langle x - y, Bx - By\rangle + \rho^2\|Bx - By\|^2 \\ &\leq \|x - y\|^2 - \frac{2\rho}{L}\|Bx - By\|^2 + \rho^2\|Bx - By\|^2 \\ &\leq \|x - y\|^2.\end{aligned}$$

□

4.5.2 A Contractive Result

We begin the main proof with a result on the one-forward-step mapping $\mathcal{F}$ from Definition 4.2. The following lemma will ultimately be used to show that the iterates remain bounded in the next subsection.

Lemma 4.9. *Suppose* $(x^+,y^+) = \mathcal{F}_{\alpha,\rho}(z,x,w;A,B,G)$*, where* $\mathcal{F}_{\alpha,\rho}$ *is given in Definition 4.2. Recall that B is* L^{-1}*-cocoercive. If* $L = 0$ *or* $\rho \leq 2(1-\alpha)/L$*, then*

$$\|x^+ - \hat\theta\| \leq (1-\alpha)\|x - \hat\theta\| + \alpha\|Gz - \hat\theta\| + \rho\,\|w - \hat w\| \tag{107}$$

for any $\hat\theta \in \text{dom}\,A$ *and* $\hat w \in A\hat\theta + B\hat\theta$*.*

Proof. Select any $\hat\theta \in \text{dom}(A)$ and $\hat w \in A\hat\theta + B\hat\theta$. Let $\hat a = \hat w - B\hat\theta \in A\hat\theta$. Then we can write

$$\begin{aligned}\hat a = \hat w - B\hat\theta &\Rightarrow \rho\hat a + \hat\theta \in \rho A\hat\theta + \hat\theta = (I + \rho A)\hat\theta \\ &\Rightarrow \hat\theta = (I + \rho A)^{-1}(\rho\hat a + \hat\theta).\end{aligned}$$

In other words

$$\hat\theta = J_{\rho A}(\hat\theta + \rho\hat a), \tag{108}$$

according to the definition of the resolvent. Therefore, the definition of the operator $\mathcal{F}$ in (97), (108) and the non-expansiveness of the operator $J_{\rho A}$ yield

$$\begin{aligned}\|x^+ - \hat\theta\| &= \left\|J_{\rho A}\big((1-\alpha)x + \alpha Gz - \rho(Bx - w)\big) - J_{\rho A}(\hat\theta + \rho\hat a)\right\| \\ &\leq \left\|(1-\alpha)x + \alpha Gz - \rho(Bx - w) - \hat\theta - \rho\hat a\right\| \\ &= \left\|(1-\alpha)\left(x - \hat\theta - \frac{\rho}{1-\alpha}\left(Bx - B\hat\theta\right)\right) + \alpha(Gz - \hat\theta) + \rho\left(w - \hat a - B\hat\theta\right)\right\| \end{aligned} \tag{109}$$

where the last term was obtained by grouping terms, adding and subtracting $B\hat{\theta}$. Notice that if $L > 0$ we have by Assumption 2 that

$$\frac{\rho}{1-\alpha} \leq \frac{2}{L}$$

therefore, applying Lemma 4.8 to the first term on the right hand side in (109) to obtain

$$\left\| x - \hat{\theta} - \frac{\rho}{1-\alpha}\left(Bx - B\hat{\theta}\right) \right\| \leq \|x - \hat{\theta}\|. \tag{110}$$

Finally, we can write (109) using the triangle inequality and (110) as

$$\begin{aligned} \|x^+ - \hat{\theta}\| &\leq (1-\alpha)\left\| x - \hat{\theta} - \frac{\rho}{1-\alpha}\left(Bx - B\hat{\theta}\right)\right\| + \alpha\|Gz - \hat{\theta}\| + \rho\left\| w - (\hat{a} + B\hat{\theta})\right\| \\ &\leq (1-\alpha)\|x - \hat{\theta}\| + \alpha\|Gz - \hat{\theta}\| + \rho\,\|w - \hat{w}\|\,. \end{aligned}$$

Alternatively, if $L = 0$, implying that B is a constant-valued operator, then $Bx = B\hat{\theta}$ and we obtain just an equality in (110). □

The inequality in (107) helps us establish the boundedness of the sequence $(x^k)_{k\geq 0}$, as in the following lemma.

4.5.3 Boundedness Results and their Direct Consequences

Lemma 4.10. *For all $i = 1, \ldots, n$, the sequences $(x_i^k)_{k\geq 0}$ and $(y_i^k)_{k\geq 0}$ are bounded.*

Proof. First, we prove the boundedness of the sequence $(x_i^k)_{k\geq 0}$. To that end, notice that in the context of Algorithm 4, we have for $i = 1, \ldots, n$

$$(x_i^k, y_i^k) = \mathcal{F}_{\alpha_i,\rho_i}(z^k, x_i^{k-1}, w_i^k; A_i, B_i, G_i).$$

Hence, in this context, Lemma 4.9 reduces to

$$\|x_i^k - \hat{\theta}_i\| \leq (1-\alpha_i)\|x_i^{k-1} - \hat{\theta}_i\| + \alpha_i\|G_i z^k - \hat{\theta}_i\| + \rho_i\left\| w_i^k - \hat{w}_i\right\| \tag{111}$$

for $k \geq 1$, and for any $\hat{\theta}_i \in \operatorname{dom} A_i$ fixed. Additionally, Assumption 2 holds, that is, $\rho_i \leq \hat{\rho}$ we arrive from (111) at

$$\|x_i^k - \hat{\theta}_i\| \leq (1-\alpha_i)\|x_i^{k-1} - \hat{\theta}_i\| + \alpha_i\|G_i z^k - \hat{\theta}_i\| + \hat{\rho}\left\| w_i^k - \hat{w}_i\right\| \tag{112}$$

Calling

$$b_i^k = \alpha_i\|G_i z^k - \hat{\theta}_i\| + \hat{\rho}\|w_i^k - \hat{w}_i\|,$$

we deduce it is bounded by, say b, since $0 < \alpha_i \leq 1$, the sequences $(z^k)_{k\geq 0}$, and $(w_i^k)_{k\geq 0}$ are bounded by Lemma 4.3, and G_i is bounded by Assumption 1. Consider now for each $i = 1, \ldots, n$ the sequence $a_i^k = x_i^k - \hat{\theta}_i$, thus we can write (112) as

$$a_i^{k+1} \leq \tau a_i^k + b.$$

Applying Lemma 4.5 with $\tau = 1-\alpha_i < 1$ to this last form of (112) we deduce boundedness of $(x_i^k)_{k\geq 0}$. Next, we establish that the sequence $(y_i^k)_{k\geq 0}$ is bounded. Recall that

$$(x_i^k, y_i^k) = \mathcal{F}_{\alpha_i,\rho_i}(z^k, x_i^{k-1}, w_i^k; A_i, B_i, G_i).$$

Expanding the y^+-update in the definition of $\mathcal{F}$ in (97), we may write

$$y_i^k = (\rho_i)^{-1}\left((1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(B_i x_i^{k-1} - w_i^k) - x_i^k\right) + Bx_i^k. \tag{113}$$

Notice that the sequence $(B_i x_i^k)_{k\geq 0}$ is bounded since B_i is continuous and $(x_i^k)_{k\geq 0}$ is bounded. Additionally, G_i, z^k, and w_i^k are bounded, hence we conclude from the expression of y_i^k in (113) its boundedness □

With $(x_i^k)_{k\geq 0}$ and $(y_i^k)_{k\geq 0}$ bounded for all $i = 1,\ldots,n$, the boundedness of $\nabla\varphi_k$ follows immediately:

Lemma 4.11. *The sequence* $(\nabla\varphi_k)_{k\geq 1}$ *is bounded. If Algorithm 4 never terminates via line 13,* $\limsup_{k\to\infty}\varphi_k(p^k) \leq 0$.

Proof. Recall that by Lemma 4.2(2), we have

$$\nabla_z\varphi_k = \sum_{i=1}^n G_i^* y_i^k,$$

which is bounded since each G_i is bounded by assumption and each $(y_i^k)_{k\geq 0}$ is bounded by Lemma 4.10. Furthermore,

$$\nabla_{w_i}\varphi_k = x_i^k - G_i x_n^k$$

is bounded for each $i = 1,\ldots,n-1$, using the same two lemmas. Therefore, the sequence $(\|\nabla\varphi_k\|)_{k\geq 1}$ is bounded, it follows from Lemma 4.3(3) that

$$\limsup_{k\to\infty}\varphi_k(p^k) \leq 0.$$

□

Once again, using the boundedness of $(x_i^k)_{k\geq 0}$ and $(y_i^k)_{k\geq 0}$, next we can derive the following simple bound relating $\varphi_{i,k-1}(z^k, w_i^k)$ to $\varphi_{i,k-1}(z^{k-1}, w_i^{k-1})$:

Lemma 4.12. *There exists* $M_1, M_2 \geq 0$ *such that for all* $k \geq 2$ *and* $i = 1,\ldots,n$,

$$\varphi_{i,k-1}(z^k, w_i^k) \geq \varphi_{i,k-1}(z^{k-1}, w_i^{k-1}) - M_1\|w_i^k - w_i^{k-1}\| - M_2\|G_i\|\|z^k - z^{k-1}\|.$$

Proof. Boundedness of the sequences $(p^k)_{k\geq 0}$, $(x_i^k)_{k\geq 0}$ and $(y_i^k)_{k\geq 0}$ together with the assumption of each G_i being bounded, allow us to conclude that for each $i \in \{1,\ldots,n\}$, there exist $M_{1,i}$, $M_{2,i} \geq 0$ such that

$$\|G_i z^{k-1} - x_i^{k-1}\| \leq M_{1,i},$$

and

$$\|y_i^{k-1} - w_i^k\| \leq M_{2,i}.$$

Let $M_1 = \max\{M_{1,1}, \ldots, M_{1,n}\}$ and $M_2 = \max\{M_{2,1}, \ldots, M_{2,n}\}$. Now, for any $k \geq 2$ and $i \in \{1, \ldots, n\}$, in order to relate $\varphi_{i,k-1}(z^{k-1}, w_i^{k-1})$ with $\varphi_{i,k-1}(z^k, w_i^k)$, we add and subtract the terms $G_i z^{k-1}$ and w_i^{k-1} in the inner product of $\varphi_{i,k-1}$ yielding

$$\begin{aligned}
\varphi_{i,k-1}(z^k, w_i^k) &= \langle G_i z^k - x_i^{k-1}, y_i^{k-1} - w_i^k \rangle \\
&= \langle G_i z^{k-1} - x_i^{k-1}, y_i^{k-1} - w_i^k \rangle + \langle G_i z^k - G_i z^{k-1}, y_i^{k-1} - w_i^k \rangle \\
&= \langle G_i z^{k-1} - x_i^{k-1}, y_i^{k-1} - w_i^{k-1} \rangle + \langle G_i z^{k-1} - x_i^{k-1}, w_i^{k-1} - w_i^k \rangle \\
&\quad + \langle G_i z^k - G_i z^{k-1}, y_i^{k-1} - w_i^k \rangle
\end{aligned}$$

Next, applying the Cauchy-Schwarz inequality to the last two terms of the right hand side follows that

$$\begin{aligned}
\varphi_{i,k-1}(z^k, w_i^k) &\geq \varphi_{i,k-1}(z^{k-1}, w_i^{k-1}) - \|G_i z^{k-1} - x_i^{k-1}\| \|w_i^{k-1} - w_i^k\| \\
&\quad - \|G_i z^k - G_i z^{k-1}\| \|y_i^{k-1} - w_i^k\| \\
&\geq \varphi_{i,k-1}(z^{k-1}, w_i^{k-1}) - M_1 \|w_i^k - w_i^{k-1}\| - M_2 \|G_i\| \|z^k - z^{k-1}\|,
\end{aligned}$$

where the last step uses the boundedness of each G_i and the definitions of M_1 and M_2. □

4.5.4 Ascent lemma

We now prove the key "ascent lemma". It shows that, while the update step given by (99) is not guaranteed to find a separating hyperplane at each iteration, it does make a certain kind of progress toward separation.

Lemma 4.13. *Suppose $(x^+, y^+) = \mathcal{F}_{\alpha,\rho}(z, x, w; A, B, G)$, where $\mathcal{F}_{\alpha,\rho}$ is given in Definition 4.2. Recall B is L^{-1}-cocoercive. Let $y \in Ax + Bx$ and define $\varphi := \langle Gz - x, y - w \rangle$. Further, define $\varphi^+ := \langle Gz - x^+, y^+ - w \rangle$, t as in (97), and $\hat{y} := \rho^{-1}(t - x^+) + Bx$. If $\alpha \in (0,1]$ and $\rho \leq 2(1-\alpha)/L$ whenever $L > 0$, then*

$$\varphi^+ \geq \frac{\rho}{2\alpha}\left(\|y^+ - w\|^2 + \alpha\|\hat{y} - w\|^2\right) + (1-\alpha)\left(\varphi - \frac{\rho}{2\alpha}\|y - w\|^2\right). \tag{114}$$

Proof. Since $y \in Ax + Bx$, there exists $a \in Ax$ such that $y = a + Bx$. Let $a^+ := \rho^{-1}(t - x^+)$. Note that $a^+ \in Ax^+$ by definition of the resolvent. With this notation, $\hat{y} = a^+ + Bx$, hence

$$(y^+, y) = (a^+ + Bx^+, a + Bx) \text{ where } a \in Ax, a^+ \in Ax^+. \tag{115}$$

Additionally, we may write the x^+-update in (97) as

$$x^+ + \rho a^+ = (1-\alpha)x + \alpha Gz - \rho(Bx - w)$$

which rearranges to

$$x^+ = (1-\alpha)x + \alpha Gz - \rho(\hat{y} - w) \implies -x^+ = -\alpha Gz - (1-\alpha)x + \rho(\hat{y} - w).$$

Adding Gz to both sides yields

$$Gz - x^+ = (1-\alpha)(Gz - x) + \rho(\hat{y} - w). \tag{116}$$

Substituting this equation into the definition of φ^+ yields

$$\begin{aligned}
\varphi^+ &= \langle Gz - x^+, y^+ - w\rangle \\
&= \big\langle (1-\alpha)(Gz - x) + \rho(\hat{y} - w), y^+ - w\big\rangle \\
&= (1-\alpha)\langle Gz - x, y^+ - w\rangle + \rho\langle \hat{y} - w, y^+ - w\rangle \\
&= (1-\alpha)\langle Gz - x, y - w\rangle + (1-\alpha)\langle Gz - x, y^+ - y\rangle + \rho\langle \hat{y} - w, y^+ - w\rangle \qquad (117)\\
&= (1-\alpha)\varphi + (1-\alpha)\langle Gz - x, y^+ - y\rangle + \rho\langle \hat{y} - w, y^+ - w\rangle.
\end{aligned}$$

We now focus on the second term on the right-hand side of (117). First, consider the case where $L > 0$. Adding and subtracting x^+ to the first entry of the inner product we obtain using (115) that

$$\begin{aligned}
\langle Gz - x, y^+ - y\rangle &= \langle x^+ - x, y^+ - y\rangle + \langle Gz - x^+, y^+ - y\rangle \\
&= \langle x^+ - x, a^+ - a\rangle + \langle x^+ - x, Bx^+ - Bx\rangle + \langle Gz - x^+, y^+ - y\rangle \\
&\geq L^{-1}\|Bx^+ - Bx\|^2 + \langle Gz - x^+, y^+ - y\rangle \qquad (118)\\
&= L^{-1}\|Bx^+ - Bx\|^2 + \langle Gz - x^+, y^+ - w\rangle + \langle Gz - x^+, w - y\rangle \\
&= L^{-1}\|Bx^+ - Bx\|^2 + \varphi^+ + \langle Gz - x^+, w - y\rangle. \qquad (119)
\end{aligned}$$

In (118) we applied the monotonicity of A and the L^{-1}-cocoercivity of B to the previous term. Next, we simply add and subtract w to the second entry of the inner product to get (119). Now, we can substitute the resulting inequality back to (117) yielding

$$\begin{aligned}
\varphi^+ &= (1-\alpha)\varphi + (1-\alpha)\langle Gz - x, y^+ - y\rangle + \rho\langle \hat{y} - w, y^+ - w\rangle \\
&\geq (1-\alpha)\varphi + (1-\alpha)\left(L^{-1}\|Bx^+ - Bx\|^2 + \varphi^+ + \langle Gz - x^+, w - y\rangle\right) \\
&\quad + \rho\langle \hat{y} - w, y^+ - w\rangle.
\end{aligned}$$

Subtracting $(1-\alpha)\varphi^+$ from both sides of the above inequality produces

$$\alpha\varphi^+ \geq (1-\alpha)\left(\varphi + L^{-1}\|Bx^+ - Bx\|^2 + \langle Gz - x^+, w - y\rangle\right) + \rho\langle \hat{y} - w, y^+ - w\rangle. \tag{120}$$

Using (116) once again, this time to the third term on the right-hand side of (120), we write

$$\begin{aligned}
\langle Gz - x^+, w - y\rangle &= \big\langle (1-\alpha)(Gz - x) + \rho(\hat{y} - w), w - y\big\rangle \\
&= (1-\alpha)\langle Gz - x, w - y\rangle + \rho\langle \hat{y} - w, w - y\rangle \\
&= (\alpha - 1)\varphi - \rho\langle \hat{y} - w, y - w\rangle. \qquad (121)
\end{aligned}$$

Substituting this equation back into (120) yields

$$\alpha\varphi^+ \geq (1-\alpha)\left(\alpha\varphi + L^{-1}\|Bx^+ - Bx\|^2 - \rho\langle \hat{y}-w, y-w\rangle\right) + \rho\langle \hat{y}-w, y^+-w\rangle. \qquad (122)$$

We next use the identity $\langle x_1, x_2\rangle = \frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_2\|^2 - \frac{1}{2}\|x_1 - x_2\|^2$ on both inner products in (122), as follows:

$$\begin{aligned}\langle \hat{y}-w, y-w\rangle &= \frac{1}{2}\left(\|\hat{y}-w\|^2 + \|y-w\|^2 - \|\hat{y}-y\|^2\right)\\ &= \frac{1}{2}\left(\|\hat{y}-w\|^2 + \|y-w\|^2 - \|a^+-a\|^2\right) \qquad (123)\end{aligned}$$

and

$$\begin{aligned}\langle \hat{y}-w, y^+-w\rangle &= \frac{1}{2}\left(\|\hat{y}-w\|^2 + \|y^+-w\|^2 - \|\hat{y}-y^+\|^2\right)\\ &= \frac{1}{2}\left(\|\hat{y}-w\|^2 + \|y^+-w\|^2 - \|Bx^+-Bx\|^2\right). \qquad (124)\end{aligned}$$

Here we have used the identities

$$\begin{aligned}\hat{y}-y &= a^+ + Bx - (a+Bx) = a^+ - a\\ \hat{y}-y^+ &= a^+ + Bx - (a^+ + Bx^+) = Bx - Bx^+.\end{aligned}$$

Using (123)–(124) in (122) yields

$$\begin{aligned}\alpha\varphi^+ &\geq (1-\alpha)\left(\alpha\varphi + L^{-1}\|Bx^+ - Bx\|^2 - \rho\langle \hat{y}-w, y-w\rangle\right) + \rho\langle \hat{y}-w, y^+-w\rangle\\ &= (1-\alpha)\left(\alpha\varphi + L^{-1}\|Bx^+ - Bx\|^2\right)\\ &\quad - \frac{\rho(1-\alpha)}{2}\left(\|\hat{y}-w\|^2 + \|y-w\|^2 - \|a^+-a\|^2\right)\\ &\quad + \frac{\rho}{2}\left(\|\hat{y}-w\|^2 + \|y^+-w\|^2 - \|Bx^+-Bx\|^2\right)\\ &= (1-\alpha)\left(\alpha\varphi - \frac{\rho}{2}\|y-w\|^2\right) + \frac{\rho}{2}\left(\|y^+-w\|^2 + \alpha\|\hat{y}-w\|^2\right)\\ &\quad + \left(\frac{1-\alpha}{L} - \frac{\rho}{2}\right)\|Bx^+-Bx\|^2 + \frac{(1-\alpha)\rho}{2}\|a^+-a\|^2.\end{aligned}$$

Consider the last two terms in this last expression: since $\alpha \leq 1$, the coefficient $(1-\alpha)\rho/2$ multiplying $\|a^+-a\|^2$ is nonnegative. Furthermore, since $\rho \leq 2(1-\alpha)/L$, the coefficient multiplying $\|Bx^+-Bx\|^2$ is positive. Therefore we may drop these two terms from the above inequality and divide by α to obtain (114).

Now, consider the case where $L = 0$, which implies that $Bx = v$ for some $v \in \mathcal{H}$ for all $x \in \mathcal{H}$. The main difference is that the $\|Bx^+ - Bx\|^2$ terms are no longer present since $Bx^+ = Bx$. The analysis is the same up to (117). Hence, instead of the expression in (119) we obtain

$$\begin{aligned}\langle Gz - x, y^+ - y\rangle &= \langle x^+ - x, y^+ - y\rangle + \langle Gz - x^+, y^+ - y\rangle\\ &= \langle x^+ - x, a^+ - a\rangle + \langle x^+ - x, Bx^+ - Bx\rangle + \langle Gz - x^+, y^+ - y\rangle\\ &\leq \varphi^+ + \langle Gz - x^+, w - y\rangle.\end{aligned}$$

Since $Bx^+ = Bx = v$ is constant we also have that

$$\hat{y} = a^+ + Bx = a^+ + v = a^+ + Bx^+ = y^+$$

Thus, instead of (120) in this case we have the simpler inequality

$$\alpha\varphi^+ \geq (1-\alpha)\left(\varphi + \langle Gz - x^+, w - y\rangle\right) + \rho\|y^+ - w\|^2. \tag{125}$$

The term $\langle Gz - x^+, w - y\rangle$ in (125) is dealt with just as in (120), by substitution of (116). This step now leads via (121) to

$$\alpha\varphi^+ \geq \alpha(1-\alpha)\varphi - \rho(1-\alpha)\langle y^+ - w, y - w\rangle + \rho\|y^+ - w\|^2.$$

Once again using $\langle x_1, x_2\rangle = \frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_2\|^2 - \frac{1}{2}\|x_1 - x_2\|^2$ on the second term on the right hand side above yields

$$\alpha\varphi^+ \geq \alpha(1-\alpha)\varphi + \rho\|y^+ - w\|^2 - \frac{\rho(1-\alpha)}{2}\left(\|y^+ - w\|^2 + \|y - w\|^2 - \|y^+ - y\|^2\right).$$

We can lower-bound the $\|y^+ - y\|^2$ term by 0. Dividing through by α and rearranging, we obtain

$$\varphi^+ \geq \frac{\rho(1+\alpha)}{2\alpha}\|y^+ - w\|^2 + (1-\alpha)\left(\varphi - \frac{\rho}{2\alpha}\|y - w\|^2\right).$$

Since $y^+ = \hat{y}$ in the $L = 0$ case, this is equivalent to (114). □

We now establish a Lyapunov-like recursion for the hyperplane. At iteration k we have from line 2 that

$$(x_i^k, y_i^k) = \mathcal{F}_{\alpha_i,\rho_i}(z^k, x_i^{k-1}, w_i^k; A_i, B_i, G_i)$$

which means in the context of lemma that $x^+ = x^k$ thus by definition of the x^+- update in (97), there exists by (9) $a_i^k \in A_i x_i^k$ such that

$$x_i^k + \rho_i a_i^k = (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(B_i x_i^{k-1} - w_i^k). \tag{126}$$

Additionally we have that

$$\varphi = \langle Gz^k - x^{k-1}, y^{k-1} - w_i^k\rangle = \varphi_{i,k-1}(z^k, w_i^k)$$
$$\varphi^+ = \langle Gz^k - x^k, y^k - w_i^k\rangle = \varphi_{i,k}(z^k, w_i^k).$$

Just as in Lemma 4.13, we define for x_i^{k-1}

$$\hat{y}_i^k := a_i^k + B_i x_i^{k-1}. \tag{127}$$

A direct application of this lemma gives us

$$\varphi_{i,k}(z^k, w_i^k) - \frac{\rho_i}{2\alpha_i}\left(\|y_i^k - w_i^k\|^2 + \alpha_i\|\hat{y}_i^k - w_i^k\|^2\right)$$
$$\geq (1-\alpha_i)\left(\varphi_{i,k-1}(z^k, w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^{k-1} - w_i^k\|^2\right) \tag{128}$$

Since

$$\|y_i^k - w_i^k\|^2 \leq \|y_i^k - w_i^k\|^2 + \alpha_i \|\hat{y}_i^k - w_i^k\|^2$$

We readily obtain from this last inequality applied to (128) that

$$\varphi_{i,k}(z^k, w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^k - w_i^k\|^2 \geq (1-\alpha_i)\left(\varphi_{i,k-1}(z^k, w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^{k-1} - w_i^k\|^2\right). \quad (129)$$

4.5.5 Finishing the Proof

We now work toward establishing the conditions of Lemma 4.4. Unless otherwise specified, we henceforth assume that Algorithm 4 runs indefinitely and does not terminate at line 13. Termination at line 13 is dealt with in Theorem 4.1 to come.

Lemma 4.14. *For all $i = 1, \ldots, n$, we have $y_i^k - w_i^k \to 0$ and $\varphi_k(p^k) \to 0$.*

Proof. Fix any $i \in \{1, \ldots, n\}$. First, note that for all $k \geq 2$,

$$\begin{aligned}\|y_i^{k-1} - w_i^k\|^2 &= \|y_i^{k-1} - w_i^{k-1}\|^2 + 2\langle y_i^{k-1} - w_i^{k-1}, w_i^{k-1} - w_i^k\rangle + \|w_i^{k-1} - w_i^k\|^2 \\ &\leq \|y_i^{k-1} - w_i^{k-1}\|^2 + M_3\|w_i^k - w_i^{k-1}\| + \|w_i^k - w_i^{k-1}\|^2 \\ &= \|y_i^{k-1} - w_i^{k-1}\|^2 + d_i^k, \end{aligned} \quad (130)$$

where

$$d_i^k := M_3\|w_i^k - w_i^{k-1}\| + \|w_i^k - w_i^{k-1}\|^2,$$

and $M_3 \geq 0$ is a bound on $2\|y_i^{k-1} - w_i^{k-1}\|$, which must exist because both $(w_i^k)_{k\geq 0}$ and $(y_i^k)_{k\geq 0}$ are bounded by Lemmas 4.3 and 4.10 respectively. Note that $d_i^k \to 0$ since $w_i^k - w_i^{k-1} \to 0$ by Lemma 4.3. Second, recall Lemma 4.12, which states that there exists $M_1, M_2 \geq 0$ such that for all $k \geq 2$,

$$\varphi_{i,k-1}(z^k, w_i^k) \geq \varphi_{i,k-1}(z^{k-1}, w_i^{k-1}) - M_1\|w_i^{k-1} - w_i^k\| - M_2\|G_i\|\|z^k - z^{k-1}\|. \quad (131)$$

Now let, for all $k \geq 1$,

$$r_i^k := \varphi_{i,k}(z^k, w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^k - w_i^k\|^2, \quad (132)$$

so that

$$\sum_{i=1}^n r_i^k = \varphi_k(p^k) - \sum_{i=1}^n \frac{\rho_i}{2\alpha_i}\|y_i^k - w_i^k\|^2. \quad (133)$$

Notice that the left hand side of (129) is r_i^k, thus we can write

$$r_i^k \geq (1-\alpha_i)\left(\varphi_{i,k-1}(z^k, w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^{k-1} - w_i^k\|^2\right). \quad (134)$$

Using (130) on (134) we obtain

$$r_i^k \geq (1-\alpha_i)\left(\varphi_{i,k-1}(z^k,w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^{k-1}-w_i^{k-1}\|^2 - \frac{\rho_i}{2\alpha_i}d_i^k\right).$$

Applying (131) to the first term of the right hand side of this last inequality yields

$$r_i^k \geq (1-\alpha_i)r_i^{k-1} + e_i^k, \quad \forall k \geq 2, \tag{135}$$

where

$$e_i^k := -(1-\alpha_i)\left(\frac{\rho_i}{2\alpha_i}d_i^k + M_1\|w_i^{k-1}-w_i^k\| + M_2\|G_i\|\|z^k-z^{k-1}\|\right). \tag{136}$$

Note that $e_i^k \to 0$. This follows from the fact that $0 < \alpha_i \leq 1$, the boundedness of ρ_i,the finiteness of $\|G_i\|$, $\|z^k - z^{k-1}\| \to 0$ and $\|w_i^k - w_i^{k-1}\| \to 0$ by Lemma 4.3, and $d_i^k \to 0$.

Since $0 < \alpha_i \leq 1$, we may apply Lemma 4.7 to (135) with $\tau = 1-\alpha_i < 1$, which yields $\liminf_{k\to\infty} r_i^k \geq 0$. Therefore

$$\liminf_{k\to\infty}\sum_{i=1}^n r_i^k \geq \sum_{i=1}^n \liminf_{k\to\infty} r_i^k \geq 0. \tag{137}$$

On the other hand, $\limsup_{k\to\infty}\varphi_k(p^k) \leq 0$ by Lemma 4.11. Therefore, using (133) and (137),

$$\begin{aligned} 0 \leq \liminf_{k\to\infty}\sum_{i=1}^n r_i^k &= \liminf_{k\to\infty}\left\{\varphi_k(p^k) - \sum_{i=1}^n \frac{\rho_i}{2\alpha_i}\|y_i^k - w_i^k\|^2\right\} \\ &\leq \liminf_{k\to\infty}\varphi_k(p^k) \leq \limsup_{k\to\infty}\varphi_k(p^k) \leq 0. \end{aligned}$$

Therefore $\lim_{k\to\infty}\varphi_k(p^k) = 0$. Consider any $i \in \{1,\ldots,n\}$. Since $\liminf_{k\to\infty}\sum_{i=1}^n r_i^k \geq 0$ and $\lim_{k\to\infty}\varphi_k(p^k) = 0$ we have from its definition in (133) that

$$\limsup_{k\to\infty}\left\{\frac{\rho_i}{\alpha_i}\|y_i^k - w_i^k\|^2\right\} \leq 0 \quad \Rightarrow \quad \|y_i^k - w_i^k\|^2 \to 0.$$

□

We have already proved the first requirement of Lemma 4.4, that $y_i^k - w_i^k \to 0$ for all $i \in \{1,\ldots,n\}$. We now work to establish the second requirement, that $G_i z^k - x_i^k \to 0$. In the upcoming lemmas we continue to use the quantity $\hat{y}_i^k$ which is given in (127).

Lemma 4.15. *For all $i = 1,\ldots,n$, $\hat{y}_i^k - w_i^k \to 0$.*

Proof. Fix any $k \geq 1$. For all $i = 1,\ldots,n$. Starting with the inequality in (128) and using (130) and (131) we obtain

$$\begin{aligned} \varphi_{i,k}(z^k,w_i^k) \geq\ & (1-\alpha_i)\left(\varphi_{i,k-1}(z^k,w_i^k) - \frac{\rho_i}{2\alpha_i}\|y_i^{k-1}-w_i^k\|^2\right) \\ &+ \frac{\rho_i}{2\alpha_i}\left(\|y_i^k - w_i^k\|^2 + \alpha_i\|\hat{y}_i^k - w_i^k\|^2\right) \\ \geq\ & (1-\alpha_i)r_i^{k-1} + \frac{\rho_i}{2}\|\hat{y}_i^k - w_i^k\|^2 + e_i^k \end{aligned}$$

where r_i^k is defined in (132) and e_i^k is defined in (136). Notice that this is the same argument used in Lemma 4.14, but applied to (128), rather than (129), so that we can upper bound the $\|\hat{y}_i^k - w_i^k\|^2$ term. Summing over $i = 1,\ldots,n$, yields

$$\varphi_k(p^k) = \sum_{i=1}^n \varphi_{i,k}(z^k, w_i^k) \geq \sum_{i=1}^n (1-\alpha_i) r_i^{k-1} + \sum_{i=1}^n \frac{\rho_i}{2}\|\hat{y}_i^k - w_i^k\|^2 + \sum_{i=1}^n e_i^k.$$

Since $\varphi_k(p^k) \to 0$, $e_i^k \to 0$, and $\liminf_{k\to\infty} r_i^k \geq 0$, the above inequality implies that $\hat{y}_i^k - w_i^k \to 0$. □

Lemma 4.16. *For $i = 1\ldots,n$, $x_i^k - x_i^{k-1} \to 0$.*

Proof. Fix $i \in \{1,\ldots,n\}$. Using the definition of a_i^k in (126) and the definition of $\hat{y}_i^k$ in (127), we have for $k \geq 1$ that

$$\begin{aligned} x_i^k + \rho_i a_i^k &= (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(B_i x_i^{k-1} - w_i^k) \\ &= (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(a_i^k + B_i x_i^{k-1} - w_i^k) \\ &= (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(\hat{y}_i^k - w_i^k). \end{aligned}$$

This implies that

$$\begin{aligned} x_i^k &= (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(\hat{y}_i^k - w_i^k), && \forall\, k \geq 1 \qquad (138)\\ x_i^{k-1} &= (1-\alpha_i)x_i^{k-2} + \alpha_i G_i z^{k-1} - \rho_i(\hat{y}_i^{k-1} - w_i^{k-1}), && \forall\, k \geq 2. \end{aligned}$$

Subtracting the second of these equations from the first yields, for all $k \geq 2$,

$$\begin{aligned} x_i^k - x_i^{k-1} = {}&(1-\alpha_i)(x_i^{k-1} - x_i^{k-2}) + \alpha_i(G_i z^k - G_i z^{k-1}) - \rho_i(\hat{y}_i^k - w_i^k) \\ &+ \rho_i(\hat{y}_i^{k-1} - w_i^{k-1}) \end{aligned}$$

Taking norms and using the triangle inequality yields, for all $k \geq 2$, that

$$\|x_i^k - x_i^{k-1}\| \leq (1-\alpha_i)\,\|x_i^{k-1} - x_i^{k-2}\| + \tilde{e}_i^k \qquad (139)$$

where

$$\tilde{e}_i^k = \|G_i\|\,\|z_i^k - z_i^{k-1}\| + \rho_i\|\hat{y}_i^k - w_i^k\| + \rho_i\|\hat{y}_i^{k-1} - w_i^{k-1}\|$$

Since ρ_i is bounded from above, $\tilde{e}_i^k \to 0$ using Lemma 4.15, the finiteness of $\|G_i\|$, and Lemma 4.3. Furthermore, $\alpha_i > 0$, so we may apply Lemma 4.6 to (139) to conclude that $x_i^k - x_i^{k-1} \to 0$. □

Lemma 4.17. *For $i = 1,\ldots,n$, $G_i z^k - x_i^k \to 0$.*

Proof. Recalling (138), we first write

$$\begin{aligned} && x_i^k &= (1-\alpha_i)x_i^{k-1} + \alpha_i G_i z^k - \rho_i(\hat{y}_i^k - w_i^k) \\ \Leftrightarrow && \alpha_i\left(G_i z^k - x_i^k\right) &= (1-\alpha_i)(x_i^k - x_i^{k-1}) + \rho_i(\hat{y}_i^k - w_i^k). \qquad (140) \end{aligned}$$

Lemma 4.16 implies that the first term on the right-hand side of (140) converges to zero. Since ρ_i is bounded from above, Lemma 4.15 implies that the second term on the right-hand side also converges to zero. Since $\alpha_i > 0$, we conclude that $\|G_i z^k - x_i^k\| \to 0$. □

Finally, we can state the convergence result for Algorithm 4:

Theorem 4.1. *Suppose that Assumptions 1-2 hold. If Algorithm 4 terminates by reaching line 13, then its final iterate is a member of the extended solution set $\mathcal{S}_e$. Otherwise, the sequence $(z^k, \boldsymbol{w}^k)_{k\geq 0}$ generated by Algorithm 4 converges weakly to some point $(\overline{z}, \overline{\boldsymbol{w}})$ in the extended solution set $\mathcal{S}_e$ of (83) defined in (85). Furthermore, $x_i^k \rightharpoonup G_i\overline{z}$ and $y_i^k \rightharpoonup \overline{w}_i$ for all $i = 1,\ldots,n-1$, $x_n^k \rightharpoonup \overline{z}$, and $y_n^k \rightharpoonup -\sum_{i=1}^{n-1} G_i^* \overline{w}_i$.*

Proof. If Algorithm 4 terminates via line 13 implies that $\pi_k = 0$ for some k. By definition of π_k in line 5 we have that

$$u^k = x_i^k - G_i x_n^k = 0 \ \forall\ i = 1,\ldots,n-1, \tag{141}$$

and

$$v^k = \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0. \tag{142}$$

Since by construction we have that $y_i^k \in T_i x_i^k$ we obtain from (141) that

$$x_i^k = G_i x_n^k \ \forall i = 1,\ldots,n-1 \Rightarrow y_i^k \in T_i G_i x_n^k \ \forall i = 1,\ldots,n-1,$$

and from (142) that

$$y_n^k = -\sum_{i=1}^{n-1} G_i^* y_i^k \Rightarrow \sum_{i=1}^{n} G_i^* y_i^k = 0.$$

Hence, taking $z^{k+1} = x_n^k$, $w_i^{k+1} = y_i^k$ for $i = 1,\ldots,n-1$ we obtain by definition of the extended solution set in (85) that $(z^{k+1}, w_1^{k+1},\ldots,w_{n-1}^{k+1}) \in \mathcal{S}_e$

Now, if the algorithm never terminates via line 13 then Lemmas 4.14 and 4.17 imply that the hypotheses of Lemma 4.4 hold, hence the conclusion follows. □

Remark 4.2. The special case where $n = 1$ turns out to be an application of the FB algorithm for a forbidden boundary case. Indeed, in this case we have by assumption that $G_1 = I$, $w_1^k = 0$, and we are solving the problem $0 \in Az + Bz$, where both operators are maximal monotone and B is L^{-1}-cocoercive. Let $x^k := x_1^k$, $y^k := y_1^k$, $\alpha := \alpha_1$, and $\rho := \rho_1$. Then the updates carried out by the algorithm are

$$\begin{aligned} x^k &= J_{\rho A}\left((1-\alpha)x^{k-1} + \alpha z^k - \rho B x^{k-1}\right) \\ y^k &= Bx^k + \frac{1}{\rho}\left((1-\alpha)x^{k-1} + \alpha z^k - \rho B x^{k-1} - x^k\right) \\ z^{k+1} &= z^k - \tau^k y^k, \quad \text{where } \tau^k = \frac{\max\{\langle z^k - x^k, y^k\rangle, 0\}}{\|y^k\|^2}. \end{aligned} \tag{143}$$

If $\alpha = 0$, then for all $k \geq 2$, the iterates computed in (143) reduce simply to

$$x^k = J_{\rho A}\left(x^{k-1} - \rho B x^{k-1}\right)$$

which is exactly FB. However, as seen in Assumption 2 the case $\alpha = 0$ is not allowed since would imply that $\rho_i = 0$. Thus, FB is a forbidden boundary case which may be approached by setting α arbitrarily close to 0. As α approaches 0, the stepsize constraint $\rho \leq 2(1-\alpha)/L$ approaches the classical stepsize constraint for FB: $\rho \leq 2/L - \varepsilon$ for some arbitrarily small constant $\varepsilon > 0$.

A potential benefit of Algorithm 4 over FB in the $n = 1$ case is that it does allow for backtracking when L is unknown or only a conservative estimate is available.

5 CONCLUSION

The simplicity of the Algorithm 2 is attractive, leaving the crucial part of the algorithm to the mechanism to choose the points. In the construction of Algorithm 3, we see great flexibility in what the choice of parameters refers to, in contrast with, for instance, the Douglas-Rachford method (Theorem 2.7) where the corresponding λ parameter is fixed in all the iterations and for both operators.

Two proofs in (ECKSTEIN; SVAITER, 2009) were modified employing two more recent results. Those results were also present in the related results for (JOHNSTONE; ECKSTEIN, 2021). For instance, Proposition 3.4 was updated using Theorem 2.3, but this theorem was also used in the related result in Lemma 4.4, which is related to the fact that the solution points are also in the defined sets V and $V^{\perp}$. In the same spirit, we saw that the operator $\boldsymbol{M} + \boldsymbol{S}$ used to prove the closedness and convexity of the extended solution set for problem (84) was applicable to the extended solution set of problem (33) too.

It is also interesting that both algorithms perform low-complexity projections over a half-space for which a simple formula is known, involving only inner products, norms, matrix multiplication by G_i (when applicable), and sums of scalars. Even when linear operators are involved there is no need to estimate their norms.

Additionally, Theorems 3.1 and 4.1 show that not only the generated sequence by the algorithm $(p^k)_{k\geq 0}$ converges to a point in the solution set, but also the sequence of points chosen $(x_i^k, y_i^k) \in \text{gra}\, T_i$, showing that the way of choosing them and the projection performed relates these sequences.

We decided to study the algorithm developed in (JOHNSTONE; ECKSTEIN, 2021) without the backtracking procedure to make simpler the exposition, however, it is of interest because in certain cases the cocoercivity constant is unknown or hard to estimate. In our first reading, we found it interesting that the backtracking termination conditions are related to the obtained inequalities in (107) and (114) showing that the properties of the operator $\mathcal{F}$ allow this backtracking search. We wish to do a deeper review of this procedure to understand the full capacities of this proposed method.

A potential benefit of Algorithm 1 over FB in the n = 1 case is that it does allow for backtracking when L is unknown or only a conservative estimate is available.

In addition to this, just as in (JOHNSTONE; ECKSTEIN, 2020) a block-iterative feature could be interesting to implement in Algorithm 4.

REFERENCES

ALOTAIBI, A.; COMBETTES, P. L.; SHAHZAD, N. Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn–Tucker set. **SIAM Journal on Optimization**, v. 24, n. 4, p. 2076–2095, 2014.

ALVES, M. M. Another proof and a generalization of a theorem of H. H. Bauschke on monotone operators. **Optimization**, p. 1–6, 2020.

BAILLON, J. B.; HADDAD, G. Quelques propriétés des opérateurs angle-bornés etn-cycliquement monotones. **Israel Journal of Mathematics**, v. 26, p. 137–150, 1977.

BAUSCHKE, H. H. A Note on the Paper by Eckstein and Svaiter on “General Projective Splitting Methods for Sums of Maximal Monotone Operators”. **SIAM Journal on Control and Optimization**, v. 48, 2009.

BAUSCHKE, H. H.; COMBETTES, P. L. **Convex analysis and monotone operator theory in Hilbert spaces**. 2nd. [S.l.]: Springer, 2017.

BOYD, S.; PARIKH, N.; CHU, E.; PELEATO, B.; ECKSTEIN, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. **Foundations and Trends in Machine Learning**, v. 3, p. 1–122, 2011.

BRICEÑO-ARIAS, L.; COMBETTES, P. L. A Monotone+Skew Splitting Model for Composite Monotone Inclusions in Duality. **SIAM Journal on Optimization**, v. 21, p. 1230–1250, 2011.

COMBETTES, P. L. Fejér monotonicity in convex optimization. In: ENCYCLOPEDIA of optimization. [S.l.]: Springer Science & Business Media, 2001. v. 2, p. 106–114.

COMBETTES, P. L.; PESQUET, J. C. Proximal Splitting Methods in Signal Processing. **Fixed-Point Algorithms for Inverse Problems in Science and Engineering**, v. 49, 2009.

COMBETTES, P. L.; WAJS, V. R. Signal Recovery by Proximal Forward-Backward Splitting. **Multiscale Modeling & Simulation**, v. 4, 2005.

DOUGLAS, J.; RACHFORD, H. H. Jr. On the numerical solution of heat conduction problems in two and three space variables. **Transactions of the American Mathematical Society**, v. 82, p. 421–439, 1956.

ECKSTEIN, J.; SVAITER, B. F. A family of projective splitting methods for the sum of two maximal monotone operators. **Mathematical Programming**, v. 111, p. 173–199, 2007.

ECKSTEIN, J.; SVAITER, B. F. General projective splitting methods for sums of maximal monotone operators. **SIAM Journal on Control and Optimization**, v. 48, n. 2, p. 787–811, 2009.

IUSEM, A. N.; SVAITER, B. F. A variant of Korpelevich's method for variational inequalities with a new search strategy. **Optimization**, v. 42, n. 4, p. 309–321, 1997.

JOHNSTONE, P. R.; ECKSTEIN, J. Projective splitting with forward steps. **Mathematical Programming**, p. 1–40, 2020.

JOHNSTONE, P. R.; ECKSTEIN, J. Single-forward-step projective splitting: exploiting cocoercivity. **Computational Optimization and Applications**, v. 78, p. 1–42, 2021.

KRASNOSEL'SKII, M. A. Two remarks on the method of successive approximations. **Uspekhi Matematicheskikh Nauk**, v. 10, n. 1, p. 123–127, 1955.

LIONS, P. L.; MERCIER, B. Splitting Algorithms for the Sum of Two Nonlinear Operators. **SIAM Journal on Numerical Analysis**, v. 16, n. 6, p. 964–979, 1979.

MANN, W. R. Mean value methods in iteration. In: PROC. Am. Math. Soc. [S.l.: s.n.], 1953. v. 4. (3), p. 506–510.

MINTY, G. J. Monotone (nonlinear) operators in Hilbert space. **Duke Mathematical Journal**, Duke University Press, v. 29, n. 3, p. 341–346, 1962.

POLYAK, B. T. **Introduction to Optimization**. [S.l.]: Optimization Software Inc. Publications Division, New York, 1987.

ROCKAFELLAR, R. T. Monotone Operators and the Proximal Point Algorithm. **SIAM Journal on Control and Optimization**, v. 14, p. 877–898, 1976.

SOLODOV, M. V.; SVAITER, B. F. A Hybrid Approximate Extragradient-Proximal Point Algorithm Using The Enlargement Of A Maximal Monotone Operator. **Set-Valued Analysis**, v. 7, p. 323–345, 1999.

SOLODOV, M. V.; SVAITER, B. F. A Hybrid Projection-Proximal Point Algorithm. **Journal of Convex Analysis**, v. 6, p. 59–70, 1999.

SOLODOV, M. V.; SVAITER, B. F. A New Projection Method for Variational Inequality Problems. **SIAM Journal on Control and Optimization**, v. 37, p. 765–776, 1999.

SOLODOV, M. V.; SVAITER, B. F. A Unified Framework For Some Inexact Proximal Point Algorithms. **Numerical Functional Analysis and Optimization**, v. 22, p. 1013–1035, 2001.

APPENDIX A – PROOF OF LEMMA 3.3

Proof. If $u = 0$, then $\boldsymbol{L}u = 0$ and (59)–(60) hold trivially, so it remains to consider the case that at least one u_i is nonzero. Given any such u, let $v = (v_1, \ldots, v_n)$ and l_{ij} be defined as in (58). Define $U \subseteq \mathcal{H}$ to be the finite-dimensional subspace spanned by $u_1, \ldots, u_n$ in $\mathcal{H}$. From (58), we have $v_i \in U$ for $i = 1, \ldots, n$, and, thus, $u, v \in U^n$. Letting $n' \leq n$ denote the dimension of U, take $B = (b_1, \ldots, b_{n'})$ to be some orthonormal basis for U. From B, we may create an orthonormal basis $\overline{B} = (\overline{b}_1, \ldots, \overline{b}_{n'n})$ for U^n via

$$\begin{array}{rclrclcrcl} \overline{b}_1 &=& (b_1, 0, 0, \ldots, 0), & \overline{b}_{n+1} &=& (b_2, 0, 0, \ldots, 0) & \cdots & \overline{b}_{(n'-1)n+1} &=& (b_{n'}, 0, 0, \ldots, 0), \\ \overline{b}_2 &=& (0, b_1, 0, \ldots, 0), & \overline{b}_{n+2} &=& (0, b_2, 0, \ldots, 0) & \cdots & \overline{b}_{(n'-1)n+2} &=& (0, b_{n'}, 0, \ldots, 0), \\ & \vdots & & & \vdots & & & & \vdots & \\ \overline{b}_n &=& (0, 0, \ldots, 0, b_1), & \overline{b}_{2n} &=& (0, 0, \ldots, 0, b_2) & \cdots & \overline{b}_{n'n} &=& (0, 0, \ldots, 0, b_{n'}). \end{array}$$

Let $\overline{\boldsymbol{u}} \in \mathbb{R}^{n'n}$ be the unique representation of u with respect to this basis, that is, its elements $\overline{u}_m, m = 1, \ldots, n'n$, are such that $u = \sum_{m=1}^{n'n} \overline{u}_m \overline{b}_m$. Similarly, let $\overline{\boldsymbol{v}} \in \mathbb{R}^{n'n}$ be the unique representation of v. By the orthonormality of the basis $\overline{B}$, it follows that $\|u\| = \|\overline{\boldsymbol{u}}\|$, $\|v\| = \|\overline{\boldsymbol{v}}\|$, and $\langle u, \boldsymbol{L}u \rangle = \langle u, v \rangle = \overline{\boldsymbol{u}}^\top \overline{\boldsymbol{v}}$. Let us now examine the action of the linear mapping defined by (58) on the basis $\overline{B}$, namely,

$$\begin{array}{rclcl} \overline{b}_1 &=& (b_1, 0, 0, \ldots, 0) & \mapsto & (\ell_{11} b_1, \ell_{21} b_1, \ldots, \ell_{n1} b_1) = \sum_{i=1}^{n} \ell_{i1} \overline{b}_i, \\ \overline{b}_2 &=& (0, b_1, 0, \ldots, 0) & \mapsto & (\ell_{12} b_1, \ell_{22} b_1, \ldots, \ell_{n2} b_1) = \sum_{i=1}^{n} \ell_{i2} \overline{b}_i, \\ && & \vdots & \\ \overline{b}_n &=& (0, 0, \ldots, 0, b_1) & \mapsto & (\ell_{1n} b_1, \ell_{2n} b_1, \ldots, \ell_{nn} b_1) = \sum_{i=1}^{n} \ell_{in} \overline{b}_i, \\ \overline{b}_{n+1} &=& (b_2, 0, 0, \ldots, 0) & \mapsto & (\ell_{11} b_2, \ell_{21} b_2, \ldots, \ell_{n1} b_2) = \sum_{i=1}^{n} \ell_{i1} \overline{b}_{n+i}, \\ \overline{b}_{n+2} &=& (0, b_2, 0, \ldots, 0) & \mapsto & (\ell_{12} b_2, \ell_{22} b_2, \ldots, \ell_{n2} b_2) = \sum_{i=1}^{n} \ell_{i2} \overline{b}_{n+i}, \\ && & \vdots & \\ \overline{b}_{n'n} &=& (0, 0, \ldots, 0, b_{n'}) & \mapsto & (\ell_{1n} b_{n'}, \ell_{2n} b_{n'}, \ldots, \ell_{nn} b_{n'}) = \sum_{i=1}^{n} \ell_{in} \overline{b}_{(n'-1)n+i}. \end{array}$$

Thus, in terms of the basis $\overline{B}$, the action of the linear mapping (58) is that of the $n'n \times n'n$ block-diagonal matrix

$$\overline{\boldsymbol{L}} := \underbrace{\begin{bmatrix} \boldsymbol{L} & & & \\ & \boldsymbol{L} & & \\ & & \ddots & \\ & & & \boldsymbol{L} \end{bmatrix}}_{n' \text{ times}}$$

and we must have $\overline{\boldsymbol{v}} = \overline{\boldsymbol{L}}\overline{\boldsymbol{u}}$. It is easily seen that $\operatorname{sym} \boldsymbol{L}$ has the same eigenvalues as $\operatorname{sym} \boldsymbol{L}$, so $\kappa(\overline{\boldsymbol{L}}) = \kappa(\boldsymbol{L})$. Using standard eigenvalue analysis in $\mathbb{R}^{n'n}$, we therefore have

$$\overline{\boldsymbol{u}}^\top \overline{\boldsymbol{L}}\overline{\boldsymbol{u}} \geq \kappa(\overline{\boldsymbol{L}}) \|\overline{\boldsymbol{u}}\|^2 = \kappa(\boldsymbol{L}) \|\overline{\boldsymbol{u}}\|^2.$$

Substituting $\|u\| = \|\overline{\boldsymbol{u}}\|$ and $\langle u, \boldsymbol{L}u\rangle = \langle u, v\rangle = \overline{\boldsymbol{u}}^\top \overline{\boldsymbol{v}} = \overline{\boldsymbol{u}}^\top \overline{\boldsymbol{L}}\overline{\boldsymbol{u}}$ into this relation yields (60). To establish (59), we observe that

$$
\begin{aligned}
\|\overline{\boldsymbol{L}}\|^2 &= \max\left\{ \|\overline{\boldsymbol{L}}\overline{\boldsymbol{x}}\|^2 \;\middle|\; \overline{\boldsymbol{x}} \in \mathbb{R}^{n'n},\ \|\overline{\boldsymbol{x}}\| = 1 \right\} \\
&= \max\left\{ \sum_{j=1}^{n'} \|\boldsymbol{L}x_j\|^2 \;\middle|\; x_1, \ldots, x_{n'} \in \mathbb{R}^n, \sum_{j=1}^{n'} \|x_j\|^2 = 1 \right\} \\
&= \max\left\{ \sum_{j=1}^{n'} \max\left\{ \|\boldsymbol{L}x\|^2 \;\middle|\; \begin{array}{c} x \in \mathbb{R}^n \\ \|x\|^2 = \nu_j \end{array} \right\} \;\middle|\; \begin{array}{c} \nu_1, \ldots, \nu_{n'} \geq 0 \\ \nu_1 + \cdots + \nu_{n'} = 1 \end{array} \right\} \\
&= \max\left\{ \sum_{j=1}^{n'} \nu_j \|\boldsymbol{L}\|^2 \;\middle|\; \begin{array}{c} \nu_1, \ldots, \nu_{n'} \geq 0 \\ \nu_1 + \cdots + \nu_{n'} = 1 \end{array} \right\} = \|\boldsymbol{L}\|^2.
\end{aligned}
$$

Thus, we may substitute $\|\overline{\boldsymbol{L}}\| = \|\boldsymbol{L}\|$ into the inequality $\|\overline{\boldsymbol{L}}\overline{\boldsymbol{u}}\| \leq \|\boldsymbol{L}\|\|\overline{\boldsymbol{u}}\|$, along with $\|\overline{\boldsymbol{L}}\overline{\boldsymbol{u}}\| = \|\overline{\boldsymbol{v}}\| = \|v\| = \|\boldsymbol{L}u\|$ and $\|u\| = \|\overline{\boldsymbol{u}}\|$ to obtain (59). □