PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

Universidad de Valparaíso CHILE

COMPOSITE MONOTONE INCLUSIONS IN VECTOR SUBSPACES: THEORY, SPLITTING, AND APPLICATIONS

Fernando Roldán Contreras

A Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics

Advisor

Dr. Luis Briceño-Arias.

Evaluation committee:
Dr. Patrick L. Combettes
Dr. Julio Deride
Dr. Jean-Christophe Pesquet
Dr. Juan Peypouquet

Valparaíso, Chile.
June, 2022

To my family and Jesus

Agradecimientos

En primer lugar quiero agradecer a mi profesor guía, Luis Briceño-Arias, quien durante estos años me ha ayudado con gran disposición, diligencia, paciencia, buscando siempre mi desarrollo académico. Agradezco que me haya introducido en una de sus áreas de investigación, compartiendo sus ideas y guiándome a presentar el escrito actual. Destaco la paciencia para transmitir su conocimiento e ideas, repitiéndolas de ser necesario (sobre todo con el inglés). Muchas gracias por todo su trabajo.

Agradezco a Juan Peypouquet por aceptar ser el revisor de esta tesis, quien adicionalmente ha sido parte importante en mi formación académica. I also thank the evaluation committee for their time, interest, suggestions and corrections.

I thank Jinjian Chen and Yuchao Tang for their collaboration and ideas.

Agradezco a las universidades Pontificia Universidad Católica de Valparaíso, Universidad de Valparaíso, y Universidad Técnica Federico Santa María, que conforman el programa en consorcio del doctorado en Matemática. Agradezco a los directores y diferentes profesores del programa, muchas gracias por su trabajo y dedicación. Agradezco en particular a Alberto Mercado, Cristopher Hermosilla, y Rodrigo Lecaros, quienes me enseñaron en los cursos de primer año. Quiero agradecer también a Zahamara por su importante ayuda con todas las gestiones administrativas, siempre entregó muy buena disposición.

Dentro de la Universidad Técnica Federico Santa María quiero, agradecer a todas las personas que me ayudaron y acompañaron de manera más cercana en este proceso. Don Augusto, don Manuel y Fabián, quienes son fundamentales para el funcionamiento del departamento de matemáticas y siempre están dispuestos para una conversación de pasillo. Andrea, Astrik y Macarena, quienes fueron de mucha ayuda, pude compartir muchos momentos con ellas, son esenciales para el departamento. A mi amigo César, en quien he encontrado una muy buena compañía, compartiendo buenos momentos dentro y fuera de la universidad, nunca faltó un tema en el que darnos vueltas o de que reírse. A todos los demás con quienes compartí durante este tiempo Patricio, Esteban, Daniel, Javier, Francisco, Rogelio, Emilio, Lisbeth, Odalis, Sebastián, Wilmer, Sergio. Muchas Gracias.

Quiero agradecer también a mis amigos en la fe Carlos, Cristian, Fredi, Nicole, Paula, Esteban, Martín, Nate, Juan Carlos, Felipe, y Brian. Quienes han sido una compañía fundamental, gracias por su tiempo, oraciones, ánimo en momentos difíciles, y siempre ayudarme a mantenerme enfocado.

Agradezco a mi papá y abuelo por su constante dedicación y esfuerzo, por su ayuda durante toda mi vida, siempre me han apoyado. Mucho de lo que soy hoy es gracias a ellos.

Agradezco a mi hermana por su amor y compañía, siempre dispuesta a ayudar. Agradezco también a mi mamá, mis tíos, primos, y cuñado por los momentos compartidos.

Agradezco a mi ayuda idónea, Mariel, quien ha presenciado de manera más íntima este proceso. Gracias por su amor, dedicación, paciencia, atenciones, y siempre apoyarme en mis quehaceres.

Todo lo anterior lo agradezco a mi Padre Celestial, quien me ha dado todo por añadidura, siendo el mayor y principal regalo de todos la salvación mediante Cristo Jesús. Agradezco a Jesús, mi salvador, por quien todo subsiste e intercede por mi ante Dios Padre, anhelo su próxima venida. Agradezco al Espíritu Santo quien me guía a toda esta verdad.

Contents

Composite Monotone Inclusions in Vector Subspaces: Theory, Splitting, and Applications

Fernando Roldán Contreras

Abstract

In this thesis we aim at solving primal-dual coupled inclusions in real Hilbert spaces involving the sum of different types of monotone operators. Our objective is to propose globally convergent splitting algorithms for solving the coupled inclusions, which take advantage of the intrinsic properties of each operator in the inclusion. We split this thesis in two main parts.

In the first part, we propose a splitting method for solving primal-dual inclusions involving general maximally monotone operators and linear compositions. First, we provide a generalization of the Douglas-Rachford splitting and the primal-dual algorithm, including critical step-sizes. We also derive a new Split-ADMM by applying our method to the dual of a convex optimization problem. Next, we derive an extension on Krasnosel'skiĭ-Mann (KM) iterations defined in the range of linear operators. We prove that the relaxed primal-dual algorithm with critical step-sizes defines KM iterations in the range of a particular linear operator and we derive its convergence. At the end of the first part, we provide the resolvent computation of the parallel composition of a maximally monotone operator by a linear operator under mild assumptions. This operation naturally appears when dealing with linearly composed maximally monotone operators. Additionally, in the context of convex optimization, we obtain the proximity operator of the infimal postcomposition under mild qualification conditions.

In the second part, we propose splitting methods for solving inclusions involving the sum of cocoercive, monotone-Lipschitzian, monotone continuous operators, and the normal cone to a closed vector subspace of a real Hilbert space. First, we suppose that the monotone continuous operator is zero and we provide a method generalizing the method of partial inverses and the forward-backward-half forward splitting, among others. Finally, we provide a splitting method for solving the case when the monotone continuous operator is not zero but the vector subspace is the whole Hilbert space. We obtain a generalization of the forward-back-half forward and Tseng's splitting algorithms involving line search. Also, we derive a method for solving non-linearly constrained composite convex optimization problem.

Additionally, in each section we provide numerical simulations to compare our methods with best competitors in literature. We provide simulations in total variation image restoration, sparse minimization, and constrained total variation least-square problems.

Chapter 1

Introduction

1.1 Problem and State-of-the-Art

In this thesis we aim at solving coupled inclusions in real Hilbert spaces involving different operators including maximally monotone, linear compositions, cocoercive, monotone-Lipschitzian, monotone-continuous, and normal cones to closed vector subspaces. Our objective is to propose efficient algorithms taking advantage of the intrinsic properties of each operator in the inclusion. More precisely, in this thesis we aim at solving instances of the following general problem. The reader is referred to Section 1.3 for notation.

Problem 1.1.1. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone operators, let $B_2\colon \mathcal{H} \to \mathcal{H}$ be a monotone-Lipschitzian operator, let $B_3\colon \mathcal{H} \to \mathcal{H}$ be a monotone-continuous operator, let $L : \mathcal{H} \to \mathcal{G}$ be a linear bounded operator, and let V be a closed vector subspace of $\mathcal{H}$. The problem is to*

$$\text{find } (x,u) \in \mathcal{H} \times \mathcal{G} \text{ such that } \begin{cases} 0 \in Ax + L^*u + B_2x + B_3x + N_Vx \\ 0 \in B^{-1}u - Lx, \end{cases} \tag{1.1.1}$$

under the assumption that the solution set of (1.1.1) is nonempty.

We will explore applications of Problem 1.1.1 in next section. We split this thesis in two main parts, in which we solve particular instances of Problem 1.1.1.

1.1.1 Case I: Primal-Dual Algorithms with Critical Step-Sizes when $B_2 = B_3 = 0$ and $V = \mathcal{H}$

In the first part of this thesis, we solve the following problem.

Problem 1.1.2. *In the context of Problem 1.1.1 assume that $B_2 = B_3 = 0$ and $V = \mathcal{H}$. Then, Problem 1.1.1 reduces to*

$$\text{find} \quad (x,u) \in \mathcal{H} \times \mathcal{G} \quad \text{such that} \quad \begin{cases} 0 \in Ax + L^*u \\ 0 \in B^{-1}u - Lx. \end{cases} \tag{1.1.2}$$

We denote by $\boldsymbol{Z} \neq \varnothing$ its set of solutions.

This inclusion arises naturally in several problems in partial differential equations coming from mechanical models [39, 41, 43], differential inclusions [2, 58], game theory [16], among other disciplines. It follows from [15, Proposition 2.8] that any solution $(\hat{x}, \hat{u})$ to Problem 1.1.2 satisfies that $\hat{x}$ is a solution to the primal inclusion

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + L^*BLx \tag{1.1.3}$$

and $\hat{u}$ is solution to the dual inclusion

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad 0 \in B^{-1}u - LA^{-1}(-L^*u). \tag{1.1.4}$$

Conversely, if $\hat{x}$ is a solution to (1.1.3) then there exists $\tilde{u}$ solution to (1.1.4) such that $(\hat{x}, \tilde{u}) \in \boldsymbol{Z}$ and the dual argument also holds. In the particular case when $A = \partial f$ and $B = \partial g^*$, for proper convex lower semicontinuous functions $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ and $g \colon \mathcal{G} \to \left]-\infty, +\infty\right]$, any solution $\hat{x}$ to (1.1.3) is a solution to the primal convex optimization problem

$$\min_{x \in \mathcal{H}} \big(f(x) + g(Lx)\big), \tag{1.1.5}$$

any solution $\hat{u}$ to (1.1.4) is a solution to the dual problem

$$\min_{u \in \mathcal{G}} \big(g^*(u) + f^*(-L^*u)\big), \tag{1.1.6}$$

and the converse holds under standard qualification conditions (see, e.g., [15]). Problems (1.1.5) and (1.1.6) model several image processing problems as image restoration and denoising [20, 23, 30, 47, 50, 57], traffic theory [12, 38, 40], among others.

In the case when $B = 0$, Problem 1.1.2 reduces to find $x \in \operatorname{zer} A := \big\{x \in \mathcal{H} \;\big|\; 0 \in Ax\big\}$ and the Proximal Point algorithm (PPA) [49, 55] generates a sequence $(x_n)_{n \in \mathbb{N}}$ converging weakly to a point in $\operatorname{zer} A$ [55, Theorem 1]. Considering the operator $J_A = (\mathrm{Id} + A)^{-1}$, PPA is detailed below.

Algorithm 1.1.3 (Proximal Point Algorithm (PPA))**.** *Let $x_0 \in \mathcal{H}$ and $\tau > 0$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \qquad \big\lfloor\; x_{n+1} = J_{\tau A} x_n. \tag{1.1.7}$$

In the case when $B \neq 0$, for solving (1.1.3), PPA can be applied to $A + L^*BL$ if it is maximally monotone. The difficulty here arises from the computation of the resolvent $J_{\tau(A+L^*BL)}$ which can be numerically expensive. For this reason, several methods solve monotone inclusions by splitting the influence of each operator involved in Problem 1.1.2. Some of this methods, which are known as splitting algorithms, are described below.

When $L = \mathrm{Id}$, Problem 1.1.2 reduces to find $x \in \mathrm{zer}(A+B)$. In this case, this problem can be solved by using the Douglas-Rachford splitting algorithm (DRS) [36, 46]. This method was introduced first in [33] to solve some discretized partial differential equations and it is detailed below.

Algorithm 1.1.4 (Douglas-Rachford Splitting (DRS))**.** *Let $x_0 \in \mathcal{H}$ and let $\tau > 0$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \qquad \lfloor\; z_n = J_{\tau A}(2J_{\tau B}z_n - z_n) + z_n - J_{\tau B}z_n. \tag{1.1.8}$$

It is proved that $(z_n)_{n\in\mathbb{N}}$ converges weakly to a point $z \in \mathcal{H}$ such that $J_{\tau A}z \in \mathrm{zer}(A+B)$ [46]. In [36] was shown that the DRS correspond to the PPA applied to a particular maximally monotone operator. Later, in [60], the weak convergence of the *shadow sequence* $(J_{\tau A}z_n)_{n\in\mathbb{N}}$ to an element in $\mathrm{zer}(A+B)$ is proved (see also [4] and [3, Section 26.3]).

In the case when $L^*L = \alpha\mathrm{Id}$, since $J_{L^*BL} = \mathrm{Id} - \frac{1}{\alpha}L^* \circ (\mathrm{Id} - J_{\alpha B}) \circ L$ [3, Proposition 23.25], DRS applied to A and L^*BL is a splitting algorithm. However, for a general operator L, J_{L^*BL} is not explicit in general. The Monotone+Skew Algorithm (MSA) [15] is a splitting algorithm proposed for solving Problem 1.1.2.

Algorithm 1.1.5 (Monotone+Skew Algorithm (MSA))**.** *Let $x_0 \in \mathcal{H}$, let $v_0 \in \mathcal{G}$, and let $\tau \in \left]0, 1/\|L\|\right[$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \qquad \left\lfloor \begin{aligned} y_{1,n} &= x_n - \tau L^* v_n \\ y_{2,n} &= v_n + \tau L v_n \\ p_{1,n} &= J_{\tau A} y_{1,n} \\ p_{2,n} &= J_{\tau B^{-1}} y_{2,n} \\ q_{1,n} &= p_{1,n} - \tau L^* p_{2,n} \\ q_{2,n} &= p_{2,n} + \tau L p_{1,n} \\ x_{n+1} &= x_n - y_{1,n} + q_{1,n} \\ v_{n+1} &= v_n - y_{2,n} + q_{2,n}. \end{aligned} \right. \tag{1.1.9}$$

The sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge weakly to a point in $\boldsymbol{Z}$ [15, Theorem 3.1]. A disadvantage of MSA is that the linear operators L and L^* are computed twice by iteration, which is numericallly costly in large scale problems.

In the optimization context on finite dimensional spaces (i.e. $A = \partial f$, $B = \partial g$, see (1.1.5) and (1.1.6)), the Primal-Dual splitting (PDS) was proposed in [21]. Later on, in [28] and [62], for optimization and monotone inclusion problems, respectively, the PDS was generalized to infinite dimensions spaces and including cocoercive operators in the inclusion (or a functions with Lipschitzian gradient in the optimization context). The primal-dual splitting in the context of Problem 1.1.2 is detailed below.

Algorithm 1.1.6 (Primal-Dual Splitting (PDS))**.** *Let $x_0 \in \mathcal{H}$, let $u_0 \in \mathcal{G}$, let $(\sigma, \tau) \in]0, +\infty[^2$ be such that $\sigma\tau\|L\|^2 < 1$, and let $(\lambda_n)_{n\in\mathbb{N}}$ in $]0, 2[$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\tau A}(x_n - \tau L^* u_n) \\ q_{n+1} = J_{\sigma B^{-1}}(u_n + \sigma L(2p_{n+1} - x_n)) \\ (x_{n+1}, u_{n+1}) = (1 - \lambda_n)(x_n, u_n) + \lambda_n(p_{n+1}, q_{n+1}). \end{array} \right. \tag{1.1.10}$$

The sequence $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$ converges weakly to a point in $\boldsymbol{Z}$ [62, Theorem 3.1]. The relaxation sequence $(\lambda_n)_{n\in\mathbb{N}}$ can be used to accelerate the numerical convergence of PDS. In [21], the DRS was deduced from PDS in the case when $L = \mathrm{Id}$ and $\sigma\tau = 1$. However, the convergence of DRS is not guaranteed in [21] because the convergence of PDS holds when $\sigma\tau < 1$. In [28] this result was also proved in the case when $\sigma\tau\|L\| \leq 1$ for optimization problems in finite dimensions. A variable metric version of the PDS (VMPDS) proposed in [27] solves Problem 1.1.2 (see [52] for a similar method in the optimization context). The method reads as follows.

Algorithm 1.1.7 (Variable Metric Primal-Dual Algorithm (VMPDS))**.** *Let $x_0 \in \mathcal{H}$, let $u_0 \in \mathcal{G}$, let $\Sigma\colon \mathcal{G} \to \mathcal{G}$ and $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ be self-adjoint linear strongly monotone operators such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\|^2 < 1$, and let $(\lambda_n)_{n\in\mathbb{N}}$ in $]0, 2[$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon L^* u_n) \\ q_{n+1} = J_{\Sigma B^{-1}}(u_n + \Sigma L(2p_{n+1} - x_n)) \\ (x_{n+1}, u_{n+1}) = (1 - \lambda_n)(x_n, u_n) + \lambda_n(p_{n+1}, q_{n+1}). \end{array} \right. \tag{1.1.11}$$

The sequences $(x_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ converge weakly to a point in $\boldsymbol{Z}$, respectively [27, Corollary 6.2]. This algorithm include self-adjoint linear strongly monotone operators Σ and Υ, which can accelerate the convergence in numerical implementations. Note that, in the case when $\Sigma = \sigma\mathrm{Id}$ and $\Upsilon = \tau\mathrm{Id}$, the VMPDS reduces to the PDS. In this context, given the connection between PDS and DRS reveled in [21] in the critical step-size case $\sigma\tau\|L\|^2 = 1$, which is studied in [28, Theorem 3.3] for optimization problems in finite dimensions, a detailed study of the limit case $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| = 1$ for VMPDS in arbitrary real Hilbert spaces is missing.

In the optimization context, the DRS was applied to the optimization dual problem (1.1.6) to derive the Alternating Direction Method of Multipliers (ADMM) (considering the operators $A = \partial g^*$ and $B = \partial(f^* \circ (-L^*))$) [39]. ADMM can also be interpreted as alternating minimization-maximization of the Augmented Lagrangian:

$$\begin{aligned} \mathcal{L}\colon\, & \mathcal{H} \times \mathcal{G} \times \mathcal{G} \to\,]-\infty, +\infty] \\ & (x, u, z) \mapsto f(x) + g(u) + \langle z \mid Lx - u\rangle + \frac{\tau}{2}\|Lx - u\|^2, \end{aligned} \tag{1.1.12}$$

where $\tau \in [0,+\infty[$. The ADMM algorithm iterates as follows.

Algorithm 1.1.8 (Alternating Direction Method of Multipliers (ADMM))**.** *Let* $(u_0, z_0) \in \mathcal{G}^2$ *and* $\tau > 0$. *Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \qquad \left\lfloor \begin{array}{l} x_{n+1} \in \arg\min_{x\in\mathcal{H}} f(x) + \frac{\tau}{2}\|Lx - (u_n - z_n/\tau)\|^2 \\ u_{n+1} = \arg\min_{u\in\mathcal{G}} g(u) + \frac{\tau}{2}\|Lx_{n+1} - (u - z_n/\tau)\|^2 \\ z_{n+1} = z_n + \tau(Lx_{n+1} - u_{n+1}). \end{array} \right. \tag{1.1.13}$$

Since the term Lx appears in the first minimization problem in (1.1.13), additional assumptions are needed to guarantee the existence of $(x_n)_{n\in\mathbb{N}}$. For example, in finite dimensional spaces, is it required L^*L to be invertible and that $(\operatorname{ri}\operatorname{dom} g) \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing$ (see [25]). This issue is because, as we claim before, the ADMM algorithm can be deduced from the DRS applying to the operator $B = \partial(f^* \circ (-L^*))$, in which ∂f^* has not been split from L^*. Similar methods which relax the hypothesis on L, and additionally include variable metric, are proposed in [1, 11, 22, 37]. Others versions of ADMM can also solve problems as in (1.1.14), which consider two linear operators $A\colon \mathcal{H} \to \mathcal{K}$ and $B\colon \mathcal{G} \to \mathcal{K}$ (see, e.g., [10]).

$$\min_{\substack{x\in\mathcal{H}, y\in\mathcal{G} \\ Ax+By=0}} f(x) + g(y). \tag{1.1.14}$$

In this context, a detailed study of weak conditions ensuring that ADMM algorithm is well defined is not available in the literature, as far as we know.

1.1.2 Case II: Splitting Algorithms when $L = \mathrm{Id}$ and B Cocoercive

The second part is dedicated to the numerical resolution of Problem 1.1.1 when $L = \mathrm{Id}$ and B is β-cocoercive[1] for some $\beta \in]0,+\infty[$. In particular, we aim at solving the following inclusion.

Problem 1.1.9. *In the context of Problem 1.1.1, assume that* $L = \mathrm{Id}$ *and* B *is* β*-cocoercive for some* $\beta \in]0,+\infty[$. *Then, Problem 1.1.1 reduces to*

$$\text{find} \quad x \in \mathcal{H} \ \text{such that} \quad 0 \in Ax + Bx + B_2x + B_3x + N_V x. \tag{1.1.15}$$

Is worth to notice that if $x \in \mathcal{H}$ is a solution to Problem 1.1.9, then, there exists $u \in \mathcal{H}$, such that (x,u) is a solution to Problem 1.1.1 in this setting.

This inclusion encompasses several problems in partial differential equations coming from mechanical models [39, 42, 43], differential inclusions [2, 58], game theory [16], among other disciplines.

[1] Let $\beta \in]0,+\infty[$. $B\colon \mathcal{H} \to \mathcal{H}$ is β-cocoercive if for every $(x,y) \in \mathcal{H}^2$, $\langle x - y \mid Bx - By\rangle \geq \beta\|Bx - By\|^2$.

In the case when $B_2 = B_3 = 0$ and $V = \mathcal{H}$ Problem 1.1.9 reduces to find $x \in \operatorname{zer}(A+B)$, which is solved by the Forward-Backward Splitting (FBS) detailed below.

Algorithm 1.1.10 (Forward-Backward Splitting (FBS))**.** *Let $x_0 \in \mathcal{H}$ and let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2\beta - \varepsilon]$ with $\varepsilon \in]0, \beta[$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \lfloor\ x_{n+1} = J_{\tau_n A}(x_n - \tau_n B x_n). \tag{1.1.16}$$

The sequence $(x_n)_{n\in\mathbb{N}}$ generated by FBS converges weakly to some $x \in \operatorname{zer}(A + B)$. This method was proposed in [43] in a optimization context, where $B = \nabla f$ and A is the normal cone to a convex set. This algorithm also was studied in [46, 51] in a monotone inclusions context.

When $B_2 = B_3 = 0$ and V is properly contained in $\mathcal{H}$, Problem 1.1.9 is reduced to find $x \in \operatorname{zer}(A + B + N_V)$. In this setting, Problem 1.1.9 is solved by the *Forward–Partial Inverse–Splitting* (FPS) proposed in [13], which is described below.

Algorithm 1.1.11 (Forward–Partial Inverse–Splitting (FPS))**.** *Let $\gamma \in]0, 2\beta[$, let $\varepsilon \in]0, 1[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$, let $x_0 \in V$, and let $y_0 \in V^\perp$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} s_n = x_n - \gamma P_V B x_n + \gamma y_n \\ p_n = J_{\gamma A} s_n \\ y_{n+1} = y_n + (\lambda_n/\gamma)(P_V p_n - p_n) \\ x_{n+1} = x_n + \lambda_n (P_V p_n - x_n). \end{array} \right. \tag{1.1.17}$$

The sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to $x \in V$ such that $x \in \operatorname{zer}(A + B + N_V)$ [13, Corollary 5.5]. In the case when $B = 0$ and $\lambda_n \equiv 1$, the FPS method coincides with the Spingarn's partial inverse method with constant step-sizes [59]. Additionally, when $V = \mathcal{H}$ the FPS method reduces to the FBS algorithm (Algorithm 1.1.10).

On the other hand, in the case $B = B_3 = 0$, Problem 1.1.9 reduces to find $x \in \operatorname{zer}(A+B_2+N_V)$. In this context, Problem 1.1.9 is solved by the *Forward–Partial–Forward Splitting* (FPFS) method proposed in [14], which is described below.

Algorithm 1.1.12 (Forward–Partial–Forward Splitting (FPFS))**.** *Let $\gamma \in]0, 1/L[$, let $\varepsilon \in]0, 1[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$, and let $z_0 \in \mathcal{H}$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} r_n = z_n - \gamma P_V B_2 P_V z_n \\ p_n = J_{\gamma A} r_n \\ s_n = 2P_V p_n - p_n + r_n - P_V r_n \\ t_n = s_n - \gamma P_V B_2 P_V s_n \\ z_{n+1} = z_n + \lambda_n (t_n - r_n). \end{array} \right. \tag{1.1.18}$$

By setting, for every $n \in \mathbb{N}$, $x_n = P_V z_n$, we have that $x_n \rightharpoonup x \in \operatorname{zer}(A + B_2 + N_V)$ [14, Corollary 3.1]. Note that, when $V = \mathcal{H}$, $\lambda_n \equiv 1$ the FPFS method reduces to the Tseng's method [61]. On the other hand, when $B_2 = 0$ and $\lambda_n \equiv 1$, the method coincides with the Spingarn's partial inverse method with constant step size [59].

In the case when $V = \mathcal{H}$, Problem 1.1.9 reduces to find $x \in \operatorname{zer}(A + B + B_2 + B_3)$. The *forward–backward–half forward* splitting (FBHF), proposed in [17], solves this problem and is detailed below.

Algorithm 1.1.13 (Forward–Backward–Half Forward (FBHF))**.** *Let $z_0 \in \operatorname{dom} A \cup X$. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \qquad \left\lfloor \begin{array}{l} x_n = J_{\gamma_n A}(z_n - \gamma_n(B + B_2 + B_3)z_n) \\ z_{n+1} = P_X(x_n + \gamma_n((B_2 + B_3)z_n - (B_2 + B_3)x_n)), \end{array} \right. \tag{1.1.19}$$

where $X \subset \mathcal{H}$ is nonempty, closed, and convex and $(\gamma_n)_{n\in\mathbb{N}}$ is a sequence in $]0, +\infty[$ satisfying one of the following conditions.

1. *Suppose that $B_3 = 0$ and, for every $n \in \mathbb{N}$, $\gamma_n \in [\eta, \chi - \eta]$, where $\eta \in]0, \chi/2[$ and $\chi = 4\beta/(1 + \sqrt{1 + 16\beta^2 L^2})$.*

2. *Suppose that $B_2 = 0$, that $X \subset \operatorname{dom} A$, let $\varepsilon \in]0, 1[$, let $\delta \in]0, 1[$, let $\theta \in \left]0, \sqrt{1 - \varepsilon}\right[$, and, for every $n \in \mathbb{N}$, let γ_n be the largest $\gamma \in \{2\beta\varepsilon\sigma, 2\beta\varepsilon\sigma^2, \ldots\}$ satisfying*

$$\gamma\|B_3 z_n - B_3 J_{\gamma A}(z_n - \gamma(B_1 + B_3)z_n)\| \leq \theta\|z_n - J_{\gamma A}(z_n - \gamma(B_1 + B_3)z_n)\|. \tag{1.1.20}$$

   Additionally, suppose that at least one of the following conditions holds:

   (i) $\liminf_{n\to\infty} \gamma_n = \delta > 0$.

   (ii) *B_3 is uniformly continuous in any weakly compact subset of $\overline{\operatorname{dom} A}$.*

The sequence $(z_n)_{n\in\mathbb{N}}$ generated by the FBHF algorithm converges weakly to $z \in X \cap \operatorname{zer}(A + B + B_2 + B_3)$ [17, Theorem 2.3]. Is worth to notice that when $B_3 = 0$, $B_2 = 0$, and $L \to 0$ the FBHF reduces to FBS (Algorithm 1.1.10). On the other hand, in the case when $B_3 = 0$, $B_1 = 0$, and $\beta \to \infty$ the Tseng's splitting algorithm [61] is recovered from FBHF. Additionally, when $B = 0$, $B_2 = 0$, $L \to 0$, and $\varepsilon \to 0$, the Tseng's splitting algorithm with backtracking [61] is also recovered.

Methods proposed in [6, 27, 31, 32, 45, 53, 62] take advantage of the cocoercivity of B, but they do not exploit the Lipschitzian property of B_2, the continuity of B_3, and the normal cone N_V, hence these methods need to compute the resolvents of B_2, B_3, and N_V. The schemes proposed in [5, 26, 34, 44] take advantage of the monotone-Lipschitzian property of B_2, but the cocoercivity of B, the continuity of B_3, and the normal cone N_V are not leveraged. If we apply the above mentioned algorithms in [5, 26, 34, 44] to Problem 1.1.9, we may consider B_3 and N_V as maximally monotone operators and $B + B_2$

as a monotone and Lipschitzian operator and activates it twice by iteration. On the other hand, Algorithms in [19, 29, 48, 54, 56] can also consider $B_1 + B_2$ as monotone and Lipschitzian activating it only once by iteration, but they need to store in the memory the two past iterations and the step-size is reduced significantly. Furthermore, methods in [19, 29, 48] consider only one maximally monotone operator, hence, the resolvent of $A + B_3 + N_V$ must be calculated; additionally, methods in [54, 56] need to calculate the resolvent of B_3 and N_V. The method in [18] exploits the properties of B and B_2 but not considers the properties of B_3 and N_V which should be treated as a maximally monotone operators. Other methods can solve Problem 1.1.9 by calculating the resolvents of B, B_2, B_3, and N_V [7, 8, 9, 15, 24, 35]. These calculation, in general, are not explicit or they can be numerically expensive.

In this thesis we split Problem 1.1.9 in two sub-problems, tackling separately the cases $B_3 = 0$ and $N_V = \mathcal{H}$, respectively. In Chapters 5 and 6. In each of these chapters, we propose a fully split method, which takes advantage of each of the intrinsic properties of the operators, overcoming the drawbacks of the methods mentioned above.

1.2 Organization and Contributions

In this section we describe the main contributions of each part.

1.2.1 Case I

In the first part of this work we aim at solving numerically Problem 1.1.2. This part consists in three chapters and each chapter contains an introduction and a main article self contained. The main contributions of each chapter are:

Chapter 2

- We propose and study the Split-Douglas-Rachford (SDR) algorithm for solving Problem 1.1.2. This algorithm generalizes DRS (Algorithm 1.1.4) splitting the influence of the linear operator from the monotone operators and include variable metrics.

- We show that VMPDS (Algorithm 1.1.7) can be deduced from SDR and it convergence is extended to operators $\Sigma\colon \mathcal{G} \to \mathcal{G}$ and $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\|^2 \leq 1$ when $\lambda_n \equiv 1$.

- In the convex optimization context, we propose the Split-ADMM algorithm (SADMM) which splits the influence of the linear operator in the first step of ADMM (Algorithm 1.1.8) and includes non-standard metrics.

- We show that the SADMM method is equivalent to the SDR applied to the optimization problem in equation (1.1.6).

- We provide a numerical comparison of SDR with several methods available in the literature in a total variation image reconstruction problem.
- We also illustrate the efficiency of SADMM by comparing its numerical performance in an academical sparse minimization example.

Chapter 3

- We obtain the weak convergence of Krasnosel'skiĭ-Mann iterations governed by firmly quasinonexpansive and averaged operators in the space $(\operatorname{ran} V, \langle \cdot \mid V \cdot \rangle)$.
- We generalize VMPDS (Algorithm 1.1.7) including critical preconditioners (i.e. linear operators Σ and Υ such that $\|\sqrt{\Sigma} L \sqrt{\Upsilon}\|^2 \leq 1$) and relaxation parameters. We obtain the weak convergence of a *shadow* sequence of the VMPDS method in the range of a linear suitable operator.
- We provide a detailed analysis of the case $L = \mathrm{Id}$ in Problem 1.1.2 and relations of primal-dual algorithms with the relaxed Douglas-Rachford splitting (DRS). We give a primal-dual version with preconditioners of DRS derived from VMPDS when $L = \mathrm{Id}$.
- We provide a numerical experiment on total variation image reconstruction. With that experiment, we illustrate the advantages of using critical preconditioners and relaxation steps.

Chapter 4

- We derive a formula for the resolvent of the parallel composition in a real Hilbert space with non-standard metrics under mild assumptions.
- We also derive a formula for the proximity operator of the infimal postcomposition in a real Hilbert space with non-standard metrics under mild assumptions.
- By using a generalization of the proximity operator, we derive a generalization of Moreau's decomposition for composite maximally monotone operators and subdifferentials of composite convex functions.

1.2.2 Case II

In the second part, we aim at solving numerically Problem 1.1.9 in the cases $B_3 = 0$ and $N_V = \mathcal{H}$, presented in chapters 5 and 6, respectively. Each chapter contains an introduction and an article self contained. The main contributions of this part are:

Chapter 5

- We propose a splitting algorithm which fully exploits the structure and the operators's properties in Problem 1.1.9 when $B_3 = 0$. We generalize the Spingarn's splitting with constant step size [59], the FPFS Algorithm 1.1.12 [14], and the FPS Algorithm 1.1.11 [13].

- By using product space techniques, we apply our algorithm to solve composite primal-dual inclusion involving Lipschitzian-monotone operators, cocoercive operators, and a normal cone to a closed vector subspace.

- We derive an algorithm for solving convex composite optimization problems under vector subspace constraints.

- We implement our method in a TV-regularized least-squares problem with constraints. We compare it performance with previous methods in the literature.

Chapter 6

- We derive a fully split method for solving Problem 1.1.9 when $V = \mathcal{H}$, which takes advantage of the intrinsic properties of the operators. In particular instances we recover the FB method (Algorithm 1.1.10), the Tseng's splitting [61], and the FBHF method (Algorithm 1.1.13) proposed in [17].

- We derive an algorithm for solving optimization problems involving convex Gâteaux differentiable functions, linear compositions, and Gâteaux differentiable nonlinear convex constraints.

- We compare numerically our method with previous methods in literature. We test it in a regularized least-squares problem with constraints, illustrating the efficiency of our method.

1.3 Notation

Throughout this thesis $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces with the scalar product $\langle\cdot\mid\cdot\rangle$ and associated norm $\|\cdot\|$. The identity operator on $\mathcal{H}$ is denoted by Id. The symbols $\rightharpoonup$ and $\to$ denote the weak and strong convergence, respectively. Let $D\subset\mathcal{H}$ be non-empty and let $T\colon D\to\mathcal{H}$. The set of fixed points of T is $\operatorname{Fix}T=\big\{x\in D\;\big|\;x=Tx\big\}$. Let $\beta\in\,]0,+\infty[$. The operator T is β-cocoercive if

$$(\forall x\in D)(\forall y\in D)\quad \langle x-y\mid Tx-Ty\rangle\geq\beta\|Tx-Ty\|^2, \tag{1.3.1}$$

it is β-Lipschitzian if

$$(\forall x\in D)(\forall y\in D)\quad \|Tx-Ty\|\leq\beta\|x-y\|, \tag{1.3.2}$$

it is nonexpansive if it is 1-Lipschitzian, it is firmly nonexpansive if

$$(\forall x\in D)(\forall y\in D)\ \ \|Tx-Ty\|^2\leq\|x-y\|^2-\|(\mathrm{Id}-T)x-(\mathrm{Id}-T)y\|^2, \tag{1.3.3}$$

it is firmly quasinonexpansive if

$$(x\in D)(y\in\operatorname{Fix}T)\quad \|Tx-y\|^2\leq\|x-y\|^2-\|Tx-x\|^2,$$

and it is $\beta-$strongly monotone if

$$(\forall x\in D)(\forall y\in D)\quad \langle x-y\mid Tx-Ty\rangle\geq\beta\|x-y\|^2.$$

Let $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ be a set-valued operator. The domain, range, graph, and zeros of A are $\operatorname{dom}A=\big\{x\in\mathcal{H}\;\big|\;Ax\neq\varnothing\big\}$, $\operatorname{ran}A=\big\{u\in\mathcal{H}\;\big|\;(\exists x\in\mathcal{H})\ u\in Ax\big\}$, $\operatorname{gra}A=\big\{(x,u)\in\mathcal{H}\times\mathcal{H}\;\big|\;u\in Ax\big\}$, and $\operatorname{zer}A=\big\{x\in\mathcal{H}\;\big|\;0\in Ax\big\}$, respectively. The inverse of A is $A^{-1}\colon u\mapsto\big\{x\in\mathcal{H}\;\big|\;u\in Ax\big\}$. The operator A is monotone if, for every (x,u) and (y,v) in $\operatorname{gra}A$, we have $\langle x-y\mid u-v\rangle\geq 0$ and A is maximally monotone if it is monotone and its graph is maximal in the sense of inclusions among the graphs of monotone operators. The resolvent of a maximally monotone operator A is $J_A=(\mathrm{Id}+A)^{-1}$, which is firmly nonexpansive and satisfies $\operatorname{Fix}J_A=\operatorname{zer}A$. The partial inverse of A with respect to a closed vector subspace V of $\mathcal{H}$, denoted by A_V, is defined by

$$(\forall(x,y)\in\mathcal{H}^2)\quad y\in A_Vx\quad\Leftrightarrow\quad (P_Vy+P_{V^\perp}x)\in A(P_Vx+P_{V^\perp}y). \tag{1.3.4}$$

Note that $A_{\mathcal{H}}=A$ and $A_{\{0\}}=A^{-1}$. Given a linear bounded operator $L\colon\mathcal{H}\to\mathcal{G}$, we denote its adjoint by $L^*\colon\mathcal{G}\to\mathcal{H}$, its kernel by $\ker L$, and its range by $\operatorname{ran}L$, and if $\operatorname{ran}L$ is closed, its *Moore-Penrose inverse* by

$$L^\dagger\colon\mathcal{G}\to\mathcal{H}\colon y\mapsto P_{C_y}0, \tag{1.3.5}$$

where $C_y = \{x \in \mathcal{H} \mid L^*Lx = L^*y\}$. If L^*L is invertible, we have [3, Example 3.29]

$$L^\dagger = (L^*L)^{-1}L^*. \tag{1.3.6}$$

For every self-adjoint monotone linear operator $U\colon \mathcal{H} \to \mathcal{H}$, we define $\|\cdot\|_U = \sqrt{\langle \cdot \mid \cdot \rangle_U}$, where $\langle \cdot \mid \cdot \rangle_U \colon (x, y) \to \langle x \mid Uy \rangle$ is bilinear, positive semi-definite, symmetric. For every x and y in $\mathcal{H}$, we have

$$\|x - y\|_U^2 = \|x\|_U^2 - 2\langle x \mid y \rangle_U + \|y\|_U^2. \tag{1.3.7}$$

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is defined by $f^*\colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$, $f^* \in \Gamma_0(\mathcal{H})$, the subdifferential of f is the maximally monotone operator $\partial f\colon x \mapsto \big\{u \in \mathcal{H} \bigm| (\forall y \in \mathcal{H})\ f(x) + \langle y - x \mid u\rangle \leq f(y)\big\}$, $(\partial f)^{-1} = \partial f^*$, and we have that $\operatorname{zer}\partial f$ is the set of minimizers of f, which is denoted by $\arg\min_{x\in\mathcal{H}} f$. Given a strongly monotone self-adjoint linear operator $\Upsilon\colon \mathcal{H} \to \mathcal{H}$, we denote by

$$\operatorname{prox}_f^\Upsilon\colon x \mapsto \underset{y\in\mathcal{H}}{\arg\min}\,\big(f(y) + \frac{1}{2}\|x - y\|_\Upsilon^2\big), \tag{1.3.8}$$

and by $\operatorname{prox}_f = \operatorname{prox}_f^{\mathrm{Id}}$. We have $\operatorname{prox}_f^\Upsilon = J_{\Upsilon^{-1}\partial f}$ [3, Proposition 24.24(i)] and it is single valued since the objective function in (1.3.8) is strongly convex. Moreover, it follows from [3, Proposition 24.24] that

$$\operatorname{prox}_f^\Upsilon = \mathrm{Id} - \Upsilon^{-1}\operatorname{prox}_{f^*}^{\Upsilon^{-1}}\Upsilon = \Upsilon^{-1}\,(\mathrm{Id} - \operatorname{prox}_{f^*}^{\Upsilon^{-1}})\,\Upsilon. \tag{1.3.9}$$

Given a non-empty set $C \subset \mathcal{H}$, we denote by $\overline{\operatorname{span}}\,C$ the closed span of C, by $\operatorname{cone} C$ its conical hull. Let C be a non-empty closed convex subset of $\mathcal{H}$. We denote by $\operatorname{sri} C = \big\{x \in C \bigm| \operatorname{cone}(C - x) = \overline{\operatorname{span}}\,(C - x)\big\}$ its strong relative interior, by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise, by $P_C^U = \operatorname{prox}_{\iota_C}^U$ the projection onto C with respect to $(\mathcal{H}, \langle \cdot \mid \cdot\rangle_U)$, and we denote $P_C = P_C^{\mathrm{Id}}$. For further properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [3].

Bibliography

[1] H. Attouch and M. Soueycatt, *Augmented lagrangian and proximal alternating direction methods of multipliers in hilbert spaces, applications to games, pde's and control*, Pacific. J. Optim., 5 (2008), pp. 17–37.

[2] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2009, https://doi.org/10.1007/978-0-8176-4848-0.

[3] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5. With a foreword by Hédy Attouch.

[4] H. H. Bauschke and W. M. Moursi, *On the order of the operators in the Douglas-Rachford algorithm*, Optim. Lett., 10 (2016), pp. 447–455, https://doi.org/10.1007/s11590-015-0920-5.

[5] R. I. Boţ and E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algorithms, 71 (2016), pp. 519–540, https://doi.org/10.1007/s11075-015-0007-5.

[6] R. I. Boţ and E. R. Csetnek, *ADMM for monotone operators: convergence analysis and rates*, Adv. Comput. Math., 45 (2019), pp. 327–359, https://doi.org/10.1007/s10444-018-9619-3.

[7] R. I. Boţ, E. R. Csetnek, and A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM J. Optim., 23 (2013), pp. 2011–2036, https://doi.org/10.1137/12088255X.

[8] R. I. Boţ, E. R. Csetnek, and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Appl. Math. Comput., 256 (2015), pp. 472–487, https://doi.org/10.1016/j.amc.2015.01.017.

[9] R. I. Boţ and C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565, https://doi.org/10.1137/120901106.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3 (2011), pp. 1–122, https://doi.org/10.1561/2200000016.

[11] K. Bredies and H. Sun, *A proximal point analysis of the preconditioned alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 878–907, https://doi.org/10.1007/s10957-017-1112-5.

[12] L. Briceño, R. Cominetti, C. E. Cortés, and F. Martínez, *An integrated behavioral model of land use and transport system: a hyper-network equilibrium approach*, Netw. Spat. Econ., 8 (2008), pp. 201–224, https://doi.org/10.1007/s11067-007-9052-5.

[13] L. M. Briceño-Arias, *Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions*, Optimization, 64 (2015), pp. 1239–1261, https://doi.org/10.1080/02331934.2013.855210.

[14] L. M. Briceño-Arias, *Forward-partial inverse-forward splitting for solving monotone inclusions*, J. Optim. Theory Appl., 166 (2015), pp. 391–413, https://doi.org/10.1007/s10957-015-0703-2.

[15] L. M. Briceño-Arias and P. L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250, https://doi.org/10.1137/10081602X.

[16] L. M. Briceño-Arias and P. L. Combettes, *Monotone operator methods for Nash equilibria in non-potential games*, in Computational and analytical mathematics, vol. 50 of Springer Proc. Math. Stat., Springer, New York, 2013, pp. 143–159, https://doi.org/10.1007/978-1-4614-7621-4_9.

[17] L. M. Briceño-Arias and D. Davis, *Forward-backward-half forward algorithm for solving monotone inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871, https://doi.org/10.1137/17M1120099.

[18] M. N. Bùi and P. L. Combettes, *Multivariate monotone inclusions in saddle form*, Mathematics of Operations Research, to appear (2022).

[19] V. Cevher and B. Vũ, *A reflected forward-backward splitting method for monotone inclusions involving lipschitzian operators*, Set-Valued Var. Anal., (2020), https://doi.org/10.1007/s11228-020-00542-4.

[20] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188, https://doi.org/10.1007/s002110050258.

[21] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[22] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*, Math. Programming, 64 (1994), pp. 81–101, https://doi.org/10.1007/BF01582566.

[23] J. Colas, N. Pustelnik, C. Oliver, P. Abry, J.-C. Géminard, and V. Vidal, *Nonlinear denoising for characterization of solid friction under low confinement pressure*, Physical Review E , 42 (2019), p. 91, https://doi.org/10.1103/PhysRevE.100.032803, https://hal.archives-ouvertes.fr/hal-02271333.

[24] P. L. Combettes and J. Eckstein, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions*, Math. Program., 168 (2018), pp. 645–672, https://doi.org/10.1007/s10107-016-1044-0.

[25] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, vol. 49 of Springer Optim. Appl., Springer, New York, 2011, pp. 185–212, https://doi.org/10.1007/978-1-4419-9569-8_10.

[26] P. L. Combettes and J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued Var. Anal., 20 (2012), pp. 307–330, https://doi.org/10.1007/s11228-011-0191-y.

[27] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318, https://doi.org/10.1080/02331934.2012.733883.

[28] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479, https://doi.org/10.1007/s10957-012-0245-9.

[29] E. Csetnek, Y. Malitsky, and M. Tam, *Shadow Douglas-Rachford splitting for monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 665–678.

[30] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, https://doi.org/10.1002/cpa.20042.

[31] D. Davis and W. Yin, *A three-operator splitting scheme and its optimization applications*, Set-Valued Var. Anal., 25 (2017), pp. 829–858, https://doi.org/10.1007/s11228-017-0421-z.

[32] Y. Dong, *Weak convergence of an extended splitting method for monotone inclusions*, J. Global Optim., 79 (2021), pp. 257–277, https://doi.org/10.1007/s10898-020-00940-w.

[33] J. Douglas, Jr. and H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439, https://doi.org/10.2307/1993056.

[34] D. Dũng and B. C. Vũ, *A splitting algorithm for system of composite monotone inclusions*, Vietnam J. Math., 43 (2015), pp. 323–341, https://doi.org/10.1007/s10013-015-0121-7.

[35] J. Eckstein, *A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 155–182, https://doi.org/10.1007/s10957-017-1074-7.

[36] J. Eckstein and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[37] J. Eckstein and W. Yao, *Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives*, Pac. J. Optim., 11 (2015), pp. 619–644.

[38] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Programming, 72 (1996), pp. 1–15, https://doi.org/10.1016/0025-5610(95)00012-7.

[39] D. Gabay, *Chapter IX applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, New York, 1983, pp. 299 – 331, https://doi.org/10.1016/S0168-2024(08)70034-1.

[40] E. M. Gafni and D. P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM J. Control Optim., 22 (1984), pp. 936–964, https://doi.org/10.1137/0322061.

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

[42] R. Glowinski and A. Marroco, *Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite, d'une classe de problemes de dirichlet non lineares*, Revue Francaise d'Automatique, Informatique et Recherche Operationelle, 9 (1975), pp. 41–76, https://doi.org/10.1051/M2AN/197509R200411.

[43] A. A. Goldstein, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society, 70 (1964), pp. 709 – 710, https://doi.org/bams/1183526263, https://doi.org/.

[44] P. R. Johnstone and J. Eckstein, *Projective splitting with forward steps only requires continuity*, Optim. Lett., 14 (2020), pp. 229–247, https://doi.org/10.1007/s11590-019-01509-7.

[45] P. R. Johnstone and J. Eckstein, *Single-forward-step projective splitting: exploiting cocoercivity*, Comput. Optim. Appl., 78 (2021), pp. 125–166, https://doi.org/10.1007/s10589-020-00238-3.

[46] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.

[47] X. Liu, D. Zhai, D. Zhao, G. Zhai, and W. Gao, *Progressive image denoising through hybrid graph Laplacian regularization: a unified framework*, IEEE Trans. Image Process., 23 (2014), pp. 1491–1503, https://doi.org/10.1109/TIP.2014.2303638.

[48] Y. Malitsky and M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim., 30 (2020), pp. 1451–1472, https://doi.org/10.1137/18M1207260.

[49] B. Martinet, *Brève communication. régularisation d'inéquations variationnelles par approximations successives*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, 4 (1970), pp. 154–158.

[50] J. Pang and G. Cheung, *Graph Laplacian regularization for image denoising: analysis in the continuous domain*, IEEE Trans. Image Process., 26 (2017), pp. 1770–1785, https://doi.org/10.1109/TIP.2017.2651400.

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

[52] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, Nov. 6-13 2011, pp. 1762–1769.

[53] H. Raguet, J. Fadili, and G. Peyré, *A generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226, https://doi.org/10.1137/120872802.

[54] J. Rieger and M. K. Tam, *Backward-forward-reflected-backward splitting for three operator monotone inclusions*, Appl. Math. Comput., 381 (2020), pp. 125248, 10, https://doi.org/10.1016/j.amc.2020.125248.

[55] R. Rockafellar, *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.

[56] E. K. Ryu and B. C. Vũ, *Finding the forward-Douglas-Rachford-forward method*, J. Optim. Theory Appl., 184 (2020), pp. 858–876, https://doi.org/10.1007/s10957-019-01601-z.

[57] L. Sha, D. Schonfeld, and J. Wang, *Graph laplacian regularization with sparse coding for image restoration and representation*, IEEE Transactions on Circuits and Systems for Video Technology, 30 (2020), pp. 2000–2014, https://doi.org/10.1109/TCSVT.2019.2913411.

[58] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1997, https://doi.org/10.1090/surv/049.

[59] J. E. Spingarn, *Partial inverse of a monotone operator*, Appl. Math. Optim., 10 (1983), pp. 247–265, https://doi.org/10.1007/BF01448388.

[60] B. F. Svaiter, *On weak convergence of the Douglas-Rachford method*, SIAM J. Control Optim., 49 (2011), pp. 280–287, https://doi.org/10.1137/100788100.

[61] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446, https://doi.org/10.1137/S0363012998338806.

[62] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

Part I

Primal-Dual Algorithms with Critical Step-Sizes when $B_2 = B_3 = 0$ and $V = \mathcal{H}$.

Preface to Part I

In this first part we aim at solving the Problem 1.1.2 proposing a generalization of PDS (Algorithm 1.1.6) including critical pre-conditioners and non-standard metrics.

In Chapter 2 we present the Split Douglas-Rachford algorithm (SDR) and we prove its convergence to a solution to Problem 1.1.2. We also show the relation of SDR with the PDS including critical pre-conditioners and non-standard metrics without relaxation parameters. Additionally, by applying the SDR algorithm to a particular optimization dual problem we derive the Split ADMM algorithm. We include numerical experiments.

In Chapter 3 we study the Primal-Dual Algorithm with critical pre-conditioners and non-standard metrics in the range of a linear operator, including relaxation parameters. We obtain a solution to Problem 1.1.2 through the convergence of a *shadow sequence*. Additionally, we deduce the convergence of the Douglas-Rachford algorithm with non-standard metrics. We provide some numerical simulations.

In Chapter 4 we obtain a formula for the resolvent of the parallel composition and the proximity operator of the infimal postcomposition, under mild assumptions. These operations arise, for example, in Chapter 2 when we derive Split ADMM.

Chapter 2

Primal-Dual Algorithm with Critical Step-Sizes

2.1 Introduction and Main Results

In this chapter we aim at solving the following problem.

Problem 2.1.1. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone operators, and let $L\colon \mathcal{H} \to \mathcal{G}$ be a non-zero linear bounded operator. The problem is to find $(x,u) \in \boldsymbol{Z}$, where*

$$\boldsymbol{Z} = \big\{(x,u) \in \mathcal{H} \times \mathcal{G} \;\big|\; 0 \in Ax + L^*u,\ 0 \in B^{-1}u - Lx\big\} \tag{2.1.1}$$

is assumed to be non-empty.

This problem arises naturally in several problems in partial differential equations coming from mechanical problems [35, 38, 39], differential inclusions [2, 53], game theory [13], among other disciplines.

In order to solve Problem 2.1.1 we propose the following algorithm.

Algorithm 2.1.2 (Split-Douglas-Rachford (SDR))**.** *In the context of Problem 2.1.1, let $(x_0,u_0) \in \mathcal{H} \times \mathcal{G}$, let $\Sigma\colon \mathcal{G} \to \mathcal{G}$ and $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ be strongly monotone self-adjoint linear operators such that $U = \Upsilon^{-1} - L^*\Sigma L$ is monotone. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} v_n = \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_n + \Sigma^{-1}u_n) \\ z_n = x_n - \Upsilon L^* v_n \\ x_{n+1} = J_{\Upsilon A} z_n \\ u_{n+1} = \Sigma L(x_{n+1} - x_n) + v_n. \end{array} \right. \tag{2.1.2}$$

We obtain the following convergence result.

Theorem 2.1.3. *In the context of Problem 2.1.1, let* $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$ *and consider the sequence* $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$ *defined by the Algorithm 2.1.2. Then, the following assertions hold:*

1. $\sum_{n\geq 1} \|x_{n+1} - x_n\|^2 < +\infty$ *and* $\sum_{n\geq 1} \|u_{n+1} - u_n\|^2 < +\infty$.
2. *There exists* $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$ *such that* $(x_n, u_n) \rightharpoonup (\hat{x}, \hat{u})$ *in* $\mathcal{H} \oplus \mathcal{G}$.

We show that VMPDS (Algorithm 1.1.7) can be deduced from SDR and its convergence is extended to operators $\Sigma\colon \mathcal{G} \to \mathcal{G}$ and $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\|^2 \leq 1$. In this chapter we also present the following proposition which shows the reduction of Algorithm 2.1.2 to DRS (see Algorithm 1.1.4) in the case when $\operatorname{ran} L = \mathcal{G}$.

Proposition 2.1.4. *In the context of Problem 2.1.1, assume* $\operatorname{ran} L = \mathcal{G}$ *and set* $\Sigma = (L\Upsilon L^*)^{-1}$. *Then, Algorithm 2.1.2 with starting point* $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$ *reduces to the recurrence*

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = J_{\Upsilon L^*BL}(2J_{\Upsilon A} z_n - z_n) + z_n - J_{\Upsilon A} z_n, \tag{2.1.3}$$

where $z_0 = x_0 - \Upsilon L^*\Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_0 + \Sigma^{-1}u_0)$.

Proposition 2.1.4 motivates us to derive an explicit computation of $J_{\tau L^*BL}$ and the resolvent of the parallel composition studied in Chapter 4.

In the optimization context, we present the following problems in order to apply the SDR algorithm to the dual problem on equation (D) to derive the Split-Alternating Direction Method of Multipliers (SADMM).

Problem 2.1.5. *Let* $\mathcal{H}$, $\mathcal{G}$, *and* $\mathcal{K}$ *be real Hilbert spaces. Let* $g \in \Gamma_0(\mathcal{K})$, *let* $f \in \Gamma_0(\mathcal{H})$, *and let* $T : \mathcal{K} \to \mathcal{G}$ *and* $K : \mathcal{G} \to \mathcal{H}$ *be non-zero bounded linear operators such that* $\operatorname{ran} T^* \cap \operatorname{dom} g^* \neq \varnothing$. *Consider the following optimization problem*

$$\min_{y\in\mathcal{K}} \big(g(y) + f(KTy)\big) \tag{P}$$

together with the associated Fenchel-Rockafellar dual

$$\min_{x\in\mathcal{H}} \big(f^*(x) + g^*(-T^*K^*x)\big). \tag{D}$$

Moreover, consider the following Fenchel-Rockafellar dual problem associated to (D)

$$\min_{u\in\mathcal{G}} \big((g^* \circ -T^*)^*(u) + f(-Ku)\big). \tag{P^*}$$

We denote by S_P, S_D, *and* S_{P^*} *the set of solutions to* (P), (D), *and* (P^*), *respectively.*

The proposed SADMM iterates as follows.

Algorithm 2.1.6 (Split-Alternating Direction Method of Multipliers (SADMM))**.** *In the context of Problem 2.2.10, let* $\Sigma\colon \mathcal{G}\to\mathcal{G}$ *and* $\Upsilon\colon \mathcal{H}\to\mathcal{H}$ *be strongly monotone self-adjoint linear operators such that* $\Sigma^{-1}-K^*\Upsilon K$ *is monotone, let* $p_0\in\mathcal{K}$*, and let* $(q_0,x_0)\in\mathcal{H}\times\mathcal{H}$*. Consider, the sequences defined by the recurrence*

$$(\forall n\in\mathbb{N})\quad \left\lfloor \begin{aligned} &y_n = x_n+\Upsilon(KTp_n-q_n)\\ &p_{n+1}\in\operatorname*{arg\,min}_{p\in\mathcal{K}}\left(g(p)+\tfrac{1}{2}\|Tp-(Tp_n-\Sigma K^*y_n)\|^2_{\Sigma^{-1}}\right)\\ &q_{n+1}=\operatorname{prox}^{\Upsilon}_{f}(\Upsilon^{-1}x_n+KTp_{n+1})\\ &x_{n+1}=x_n+\Upsilon(KTp_{n+1}-q_{n+1}). \end{aligned}\right. \tag{2.1.4}$$

We obtain the following result which establishes the relation between SDR applied to (D) and SADMM and the weak convergence of SADMM.

Theorem 2.1.7. *In the context of Problem 2.1.5, suppose that there exists* $(x,u)\in\mathcal{H}\times\mathcal{G}$ *such that*

$$\begin{cases} 0\in\partial f^*(\hat{x})+K\hat{u}\\ 0\in\partial(g^*\circ -T^*)^*(\hat{u})-K^*\hat{x}, \end{cases}$$

set

$$A=\partial f^*,\quad B=\partial(g^*\circ(-T^*)),\quad \textit{and}\quad L=K^*, \tag{2.1.5}$$

and assume that $0\in\operatorname{sri}(\operatorname{dom}g^*-\operatorname{ran}T^*)$*. Then,* $(p_n)_{n\in\mathbb{N}}$ *defined in (2.1.4) exists and the following statements hold.*

1. *(SDR reduces to SADMM) Let* $(\tilde{x}_n)_{n\in\mathbb{N}}$*,* $(\tilde{u}_n)_{n\in\mathbb{N}}$*, and* $(\tilde{v}_n)_{n\in\mathbb{N}}$ *be the sequences generated by Algorithm 2.1.2 and set*

$$(\forall n\in\mathbb{N})\quad \begin{cases} \tilde{p}_{n+1}\in T^{-1}(-\tilde{v}_n)\\ \tilde{q}_{n+1}=\Upsilon^{-1}(\tilde{x}_n-\tilde{x}_{n+1}-\Upsilon K\tilde{v}_n). \end{cases} \tag{2.1.6}$$

   Moreover, set $p_1\in\mathcal{K}$ *such that* $Tp_1=T\tilde{p}_1$*, and* $q_1=\tilde{q}_1$*,* $x_1=\tilde{x}_1$*. Then, sequences* $(p_n)_{n\geq1}$*,* $(q_n)_{n\geq1}$*, and* $(x_n)_{n\geq1}$ *generated by Algorithm 2.1.6 satisfy, for every* $n\geq1$*,* $T\tilde{p}_n=Tp_n$*,* $\tilde{q}_n=q_n$*, and* $\tilde{x}_n=x_n$*.*

2. *(SADMM reduces to SDR) Let* $(p_n)_{n\geq1}$*,* $(q_n)_{n\geq1}$*, and* $(x_n)_{n\geq1}$ *be sequences generated by Algorithm 2.1.6 and define*

$$(\forall n\in\mathbb{N})\quad u_{n+1}=\Sigma K^*(x_{n+1}-x_n)-Tp_{n+1}. \tag{2.1.7}$$

   Moreover, set $\tilde{x}_0=x_1$*,* $\tilde{u}_0=u_1$*, and let* $(\tilde{x}_n)_{n\in\mathbb{N}}$ *and* $(\tilde{u}_n)_{n\in\mathbb{N}}$ *be the sequences generated by Algorithm 2.1.2. Then, for all* $n\in\mathbb{N}$*,* $\tilde{x}_n=x_{n+1}$ *and* $\tilde{u}_n=u_{n+1}$*.*

3. *Let $(p_n)_{n\in\mathbb{N}}$, $(q_n)_{n\in\mathbb{N}}$, and $(x_n)_{n\in\mathbb{N}}$ be sequences generated by Algorithm 2.1.6. Then, the following hold:*
   (a) *There exists $(\hat{y},\hat{x},\hat{u}) \in S_P \times S_D \times S_{P^*}$ such that $(x_n, -Tp_n, q_n) \rightharpoonup (\hat{x},\hat{u},-K\hat{u})$ and $\hat{u} = -T\hat{y}$.*
   (b) *Suppose that $\operatorname{ran} T^* = \mathcal{K}$. Then, there exists $\hat{y} \in S_P$ such that $p_n \rightharpoonup \hat{y}$.*

We show that SADMM algorithm can be reduced to the ADMM algorithm (Algorithm 1.1.8) including preconditioners. We also present a version of SADMM which allows to deal with more general formulations involving two linear operators as in (1.1.14).

We finalize this chapter with numerical experiments. First, we present a comparison of SDR with several methods available in the literature in a total variation image reconstruction problem. Later we illustrate the efficiency of SADMM in an academical example.

2.2 Article: Split-Douglas–Rachford Algorithm for Composite Monotone Inclusions and Split-ADMM[1]

Abstract In this paper we provide a generalization of the Douglas-Rachford splitting (DRS) and the primal-dual algorithm [25, 56] for solving monotone inclusions in a real Hilbert space involving a general linear operator. The proposed method allows for primal and dual non-standard metrics and activates the linear operator separately from the monotone operators appearing in the inclusion. In the simplest case when the linear operator has full range, it reduces to classical DRS. Moreover, the weak convergence of primal-dual sequences to a Kuhn-Tucker point is guaranteed, generalizing the main result in [54]. Inspired by [35], we also derive a new Split-ADMM (SADMM) by applying our method to the dual of a convex optimization problem involving a linear operator which can be expressed as the composition of two linear operators. The proposed SADMM activates one linear operator implicitly and the other one explicitly, and we recover ADMM when the latter is set as the identity. Connections and comparisons of our theoretical results with respect to the literature are provided for the main algorithm and SADMM. The flexibility and efficiency of both methods is illustrated via a numerical simulations in total variation image restoration and a sparse minimization problem.

2.2.1 Introduction

In this paper we focus on a splitting algorithm for solving the following primal-dual monotone inclusion.

[1] [16] Luis M. Briceño-Arias and Fernando Roldán. Split-Douglas–Rachford algorithm for composite monotone inclusions and split-ADMM. *SIAM J. Optim.*, 31(4), 2987-3013, 2021.

Problem 2.2.1. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone operators, and let $L\colon \mathcal{H} \to \mathcal{G}$ be a non-zero linear bounded operator. The problem is to find $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$, where*

$$\boldsymbol{Z} = \big\{(\hat{x}, \hat{u}) \in \mathcal{H} \times \mathcal{G} \;\big|\; 0 \in A\hat{x} + L^*\hat{u},\ 0 \in B^{-1}\hat{u} - L\hat{x}\big\} \tag{2.2.1}$$

is assumed to be non-empty.

This problem arises naturally in several problems in partial differential equations coming from mechanical problems [35, 38, 39], differential inclusions [2, 53], game theory [13], among other disciplines. The set $\boldsymbol{Z}$ is the collection of Kuhn-Tucker points [3, Problem 26.30], which is also known as *extended solution set* (see, e.g., [26] and [31, 54] for the case when $L = \mathrm{Id}$).

It follows from [12, Proposition 2.8] that any solution $(\hat{x}, \hat{u})$ to Problem 2.2.1 satisfies that $\hat{x}$ is a solution to the primal inclusion

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + L^*BLx \tag{2.2.2}$$

and $\hat{u}$ is solution to the dual inclusion

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad 0 \in B^{-1}u - LA^{-1}(-L^*u). \tag{2.2.3}$$

Conversely, if $\hat{x}$ is a solution to (2.2.2) then there exists $\tilde{u}$ solution to (2.2.3) such that $(\hat{x}, \tilde{u}) \in \boldsymbol{Z}$ and the dual argument also holds. In the particular case when $A = \partial f$ and $B = \partial g^*$, for proper convex lower semicontinuous functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ and $g\colon \mathcal{G} \to \left]-\infty, +\infty\right]$, any solution $\hat{x}$ to (2.2.2) is a solution to the primal convex optimization problem

$$\min_{x \in \mathcal{H}} \big(f(x) + g(Lx)\big), \tag{2.2.4}$$

any solution $\hat{u}$ to (2.2.3) is a solution to the dual problem

$$\min_{u \in \mathcal{G}} \big(g^*(u) + f^*(-L^*u)\big), \tag{2.2.5}$$

and the converse holds under standard qualification conditions (see, e.g., [12]). Problems (2.2.4) and (2.2.5) model several image processing problems as image restoration and denoising [19, 22, 27, 43, 47, 51], traffic theory [11, 34, 37], among others.

In the case when $L = \mathrm{Id}$, Problem 2.2.1 is solved by the Douglas-Rachford splitting (DRS) [42], which is a classical algorithm inspired from a numerical method for solving linear systems appearing in discretizations of PDEs [28]. Given $z_0 \in \mathcal{H}$ and $\tau > 0$, DRS generates the sequence $(z_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ via the recurrence

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = J_{\tau B}(2J_{\tau A}z_n - z_n) + z_n - J_{\tau A}z_n, \tag{2.2.6}$$

and $z_n \rightharpoonup \hat{z}$ for some $\hat{z} \in \mathcal{H}$ such that $J_{\tau A}\hat{z}$ is a zero of $A+B$ [42, Theorem 1], where we denote the resolvent of $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ by $J_M = (\mathrm{Id}+M)^{-1}$. Under additional assumptions, such as weak lower semicontinuity of $J_{\tau A}$ or maximal monotonicity of $A+B$, the weak convergence of the *shadow sequence* $(J_{\tau A} z_n)_{n\in\mathbb{N}}$ to a zero of $A+B$ is guaranteed in [42, Theorem 1]. More than thirty years later, the weak convergence of the shadow sequence to a solution is proved in [54] without any further assumption.

In the general case when $L \neq \mathrm{Id}$, a drawback of DRS is that the maximal monotonicity of L^*BL is needed in order to ensure the weak convergence of $(z_n)_{n\in\mathbb{N}}$ and the computation of its resolvent at each iteration usually leads to sub-iterations, at exception of very particular cases. Several algorithms in the literature including [4, 5, 6, 12, 14, 56] split the influence of the linear operator L from the monotone operators, avoiding sub-iterations. In particular, we highlight the primal-dual splitting (PDS) proposed in [56], which generates a sequence in $\mathcal{H}\times\mathcal{G}$ via the recurrence

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x_{n+1} = J_{\tau A}(x_n - \tau L^* v_n) \\ v_{n+1} = J_{\sigma B^{-1}}(v_n + \sigma L(2x_{n+1} - x_n)), \end{array} \right. \tag{2.2.7}$$

for some initial point $(x_0, v_0) \in \mathcal{H}\times\mathcal{G}$ and strictly positive step-sizes satisfying $\tau\sigma\|L\|^2 < 1$.

In the context of convex optimization, it is well known that DRS applied to (2.2.5) leads to the alternating direction method of multipliers (ADMM) [35, 36, 38], whose first step needs sub-iterations in general. This drawback is overcome by the splitting methods proposed in [4, 5, 6, 14, 20, 41, 45]. In particular, the algorithm proposed in [20] coincides with PDS in (2.2.7) in the optimization setting and its convergence is guaranteed if $\tau\sigma\|L\|^2 < 1$. In [25], the convergence of the sequences generated by (2.2.7) with step-sizes satisfying the limit condition $\tau\sigma\|L\|^2 = 1$ is studied in finite dimensions. This limit case is important because the algorithm improves its efficiency as the parameters approach the boundary (see Section 2.2.5.1), it has the advantage of tuning only one parameter, and the algorithm reduces to DRS and ADMM when $L = \mathrm{Id}$ and $\tau\sigma = 1$ [20, Section 4.2]. Furthermore, a preconditioned version of (2.2.7) in the optimization context is proposed in [48]. In this extension, $\tau\mathrm{Id}$ and $\sigma\mathrm{Id}$ are generalized to strongly monotone self-adjoint linear operators Υ and Σ, respectively, and the convergence is guaranteed under the condition $\|\Sigma^{\frac{1}{2}} L \Upsilon^{\frac{1}{2}}\| < 1$. A preconditioned version of (2.2.7) for monotone inclusions is derived in [24].

In this paper we propose and study the following splitting algorithm for solving Problem 2.2.1, which is a generalization of DRS when $L \neq \mathrm{Id}$ and of [24, 56].

Algorithm 2.2.2 (Split-Douglas-Rachford (SDR))**.** *In the context of Problem 2.2.1, let $(x_0, u_0) \in \mathcal{H}\times\mathcal{G}$, let $\Sigma\colon \mathcal{G} \to \mathcal{G}$ and $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ be strongly monotone self-adjoint linear* *operators such that* $U = \Upsilon^{-1} - L^*\Sigma L$ *is monotone. Consider the recurrence:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} v_n = \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_n + \Sigma^{-1}u_n) \\ z_n = x_n - \Upsilon L^* v_n \\ x_{n+1} = J_{\Upsilon A} z_n \\ u_{n+1} = \Sigma L(x_{n+1} - x_n) + v_n. \end{array} \right. \tag{2.2.8}$$

Note that Algorithm 2.2.2 splits the influence of the linear operator from the monotone operators and, by storing $(Lx_n)_{n\in\mathbb{N}}$, only one activation of L is needed at each iteration. Moreover, in the case when $\operatorname{ran} L = \mathcal{G}$, we prove in Proposition 2.2.8 that (2.2.8) reduces to a preconditioned version of DRS in (2.2.6), in which case $J_{\Upsilon L^*BL}$ has a closed formula depending on the resolvent of B. Other preconditioned versions of DRS are used for solving structured convex optimization problems in [6, 8, 10, 58], but they do not reduce to DRS when $L = \mathrm{Id}$. Without any further assumptions than those in Problem 2.2.1, we guarantee the weak convergence of the sequence $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$ generated by Algorithm 2.2.2 to a point in $\boldsymbol{Z}$, generalizing the result in [54] to the case when $L \neq \mathrm{Id}$. In the particular case when $\|\Sigma^{\frac{1}{2}} L \Upsilon^{\frac{1}{2}}\| < 1$, we obtain a reduction of Algorithm 2.2.2 to the preconditioned PDS in [48] and, when $\|\Sigma^{\frac{1}{2}} L \Upsilon^{\frac{1}{2}}\| = 1$, we generalize [25, Theorem 3.3] to monotone inclusions and infinite dimensions considering non-standard metrics. We also provide a numerical comparison of Algorithm 2.2.2 with several methods available in the literature in a total variation image reconstruction problem.

Another contribution of this manuscript is a generalization of ADMM in the convex optimization context, by applying Algorithm 2.2.2 to the dual problem of (2.2.4) when $L = KT$, for some non-trivial linear operators T and K. This splitting, called Split-ADMM (SADMM), allows us to solve (2.2.4) by activating T implicitly and K explicitly. SADMM reduces to the classical ADMM in the case when $K = \mathrm{Id}$, $\Sigma = \sigma\mathrm{Id}$, and $\Upsilon = \tau\mathrm{Id}$ and, in the case when $T = \mathrm{Id}$, it is a fully explicit algorithm which splits the influence of the linear operator in the first step of ADMM. We prove the weak convergence of SADMM, generalizing results in [29, 35, 36]. We also prove the equivalence between SDR and SADMM, generalizing some results in [1, 29, 35, 36, 46] to the case when $L \neq \mathrm{Id}$. In addition, we provide a version of SADMM able to deal with two linear operators as in [9]. The resulting method is a non-standard metric version of several ADMM-type algorithms in [4, 9, 52, 59] and it can be seen as an augmented Lagrangian method with a non-standard metric. We also illustrate the efficiency of SADMM by comparing its numerical performance in an academical sparse minimization example in which the matrix L be factorized as $L = KT$ from its singular value decomposition (SVD). We show that the computational time may be drastically reduced by using SADMM with a suitable factorization of L.

The paper is organized as follows. In Section 2.2.2 we set our notation. In Section 2.2.3 we provide the proof of convergence of SDR and we connect our results with the literature. In Section 2.2.4 we derive the SADMM, we provide several theoretical results, and we compare them with the literature in convex optimization. Finally, in Section 2.2.5 we provide numerical simulations illustrating the efficiency of SDR and SADMM.

2.2.2 Notations and Preliminaries

Throughout this paper $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces with the scalar product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\|$. The identity operator on $\mathcal{H}$ is denoted by Id. Given a linear bounded operator $L : \mathcal{H} \to \mathcal{G}$, we denote its adjoint by $L^*\colon \mathcal{G} \to \mathcal{H}$, its kernel by $\ker L$, and its range by $\operatorname{ran} L$. The symbols $\rightharpoonup$ and $\to$ denote the weak and strong convergence, respectively. Let $D \subset \mathcal{H}$ be non-empty and let $T : D \to \mathcal{H}$. The set of fixed points of T is $\operatorname{Fix} T = \big\{x \in D \;\big|\; x = Tx\big\}$. Let $\beta \in \left]0, +\infty\right[$. The operator T is $\beta-$strongly monotone if, for every x and y in D, we have $\langle x - y \mid Tx - Ty\rangle \geq \beta\|x - y\|^2$, it is nonexpansive if, for every x and y in D, we have $\|Tx - Ty\| \leq \|x - y\|$, it is firmly nonexpansive if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2, \tag{2.2.9}$$

and it is firmly quasinonexpansive if, for every $x \in D$ and $y \in \operatorname{Fix} T$, we have $\|Tx - y\|^2 \leq \|x - y\|^2 - \|Tx - x\|^2$. Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The inverse of A is $A^{-1}\colon u \mapsto \big\{x \in \mathcal{H} \;\big|\; u \in Ax\big\}$. The domain, range, graph, and zeros of A are $\operatorname{dom} A = \big\{x \in \mathcal{H} \;\big|\; Ax \neq \varnothing\big\}$, $\operatorname{ran} A = \big\{u \in \mathcal{H} \;\big|\; (\exists x \in \mathcal{H})\; u \in Ax\big\}$, $\operatorname{gra} A = \big\{(x, u) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$, and $\operatorname{zer} A = \big\{x \in \mathcal{H} \;\big|\; 0 \in Ax\big\}$, respectively. The operator A is monotone if, for every (x, u) and (y, v) in $\operatorname{gra} A$, we have $\langle x - y \mid u - v\rangle \geq 0$ and A is maximally monotone if it is monotone and its graph is maximal in the sense of inclusions among the graphs of monotone operators. The resolvent of a maximally monotone operator A is $J_A = (\mathrm{Id} + A)^{-1}$, which is firmly nonexpansive and satisfies $\operatorname{Fix} J_A = \operatorname{zer} A$.

For every self-adjoint monotone linear operator $U\colon \mathcal{H} \to \mathcal{H}$, we define $\|\cdot\|_U = \sqrt{\langle \cdot \mid \cdot \rangle_U}$, where $\langle \cdot \mid \cdot \rangle_U\colon (x, y) \to \langle x \mid Uy\rangle$ is bilinear, positive semi-definite, symmetric. For every x and y in $\mathcal{H}$, we have

$$\|x - y\|_U^2 = \|x\|_U^2 - 2\langle x \mid y\rangle_U + \|y\|_U^2. \tag{2.2.10}$$

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is defined by $f^*\colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$, $f^* \in \Gamma_0(\mathcal{H})$, the subdifferential of f is the maximally monotone operator $\partial f\colon x \mapsto \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H})\;\; f(x) + \langle y - x \mid u\rangle \leq f(y)\big\}$, $(\partial f)^{-1} = \partial f^*$, and we have that $\operatorname{zer} \partial f$ is the set of minimizers of f, which is denoted by $\arg\min_{x\in\mathcal{H}} f$. Given a strongly monotone self-adjoint linear operator $\Upsilon\colon \mathcal{H} \to \mathcal{H}$, we denote by

$$\operatorname{prox}_f^{\Upsilon}\colon x \mapsto \underset{y\in\mathcal{H}}{\arg\min}\,\Big(f(y) + \frac{1}{2}\|x - y\|_{\Upsilon}^2\Big), \tag{2.2.11}$$

and by $\operatorname{prox}_f = \operatorname{prox}_f^{\mathrm{Id}}$. We have $\operatorname{prox}_f^{\Upsilon} = J_{\Upsilon^{-1}\partial f}$ [3, Proposition 24.24(i)] and it is single valued since the objective function in (2.2.11) is strongly convex. Moreover, it follows from [3, Proposition 24.24] that

$$\operatorname{prox}_f^{\Upsilon} = \mathrm{Id} - \Upsilon^{-1} \operatorname{prox}_{f^*}^{\Upsilon^{-1}} \Upsilon = \Upsilon^{-1} (\mathrm{Id} - \operatorname{prox}_{f^*}^{\Upsilon^{-1}}) \Upsilon. \tag{2.2.12}$$

Given a non-empty closed convex set $C \subset \mathcal{H}$, we denote by P_C the projection onto C, by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise, we denote by $N_C = \partial \iota_C$ the normal cone to C, and by $\operatorname{sri} C$ its strong relative interior. For further properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [3].

We finish this section with a result involving monotone linear operators, which is useful for the connection of our algorithm and [48].

Proposition 2.2.3. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ and $\Sigma\colon \mathcal{G} \to \mathcal{G}$ be strongly monotone self-adjoint linear operators, and set*

$$\boldsymbol{V}\colon \mathcal{H} \oplus \mathcal{G} \to \mathcal{H} \oplus \mathcal{G}\colon (x,u) \mapsto (\Upsilon^{-1}x - L^*u, \Sigma^{-1}u - Lx). \tag{2.2.13}$$

Then, the following statements are equivalent.

1. $\Upsilon^{-1} - L^* \circ \Sigma \circ L$ *is monotone.*
2. $\|\Sigma^{\frac{1}{2}} \circ L \circ \Upsilon^{\frac{1}{2}}\| \leq 1$.
3. $\|\Upsilon^{\frac{1}{2}} \circ L^* \circ \Sigma^{\frac{1}{2}}\| \leq 1$.
4. $\Sigma^{-1} - L \circ \Upsilon \circ L^*$ *is monotone.*
5. *For every* $(x,u) \in \mathcal{H} \times \mathcal{G}$,

$$\langle (x,u) \mid \boldsymbol{V}(x,u) \rangle \geq \max\big\{\|\Upsilon^{-1}u - L^*x\|_{\Upsilon}^2, \|\Sigma^{-1}u - Lx\|_{\Sigma}^2\big\}. \tag{2.2.14}$$

Moreover, if any of the statements above holds, $\boldsymbol{V}$ is $\frac{\tau\sigma}{\tau+\sigma}$−cocoercive, where $\tau > 0$ and $\sigma > 0$ are the strong monotonicity constants of Υ and Σ, respectively.

Proof. 1⇔2: Since Σ and Υ are strongly monotone, linear, and self-adjoint, it follows from [49, Theorem p. 265] that there exists strongly monotone, linear, self-adjoint operators $\Sigma^{\frac{1}{2}}$ and $\Upsilon^{\frac{1}{2}}$ such that $\Sigma = \Sigma^{\frac{1}{2}} \circ \Sigma^{\frac{1}{2}}$ and $\Upsilon = \Upsilon^{\frac{1}{2}} \circ \Upsilon^{\frac{1}{2}}$. Moreover, Υ, Σ, $\Upsilon^{\frac{1}{2}}$, and $\Sigma^{\frac{1}{2}}$ are invertible. Hence, we have

$$\begin{aligned}(\forall x \in \mathcal{H}) \quad \big\langle (\Upsilon^{-1} - L^* \circ \Sigma \circ L)x \mid x \big\rangle &= \|\Upsilon^{-\frac{1}{2}}x\|^2 - \|\Sigma^{\frac{1}{2}}Lx\|^2 \\ &= \|\Upsilon^{-\frac{1}{2}}x\|^2 \left(1 - \frac{\|\Sigma^{\frac{1}{2}}L\Upsilon^{\frac{1}{2}}\Upsilon^{-\frac{1}{2}}x\|^2}{\|\Upsilon^{-\frac{1}{2}}x\|^2}\right). \end{aligned} \tag{2.2.15}$$

Therefore, since $\Upsilon^{-\frac{1}{2}}$ is a bijection, by denoting $y = \Upsilon^{-\frac{1}{2}}x$, 1 yields

$$\|\Sigma^{\frac{1}{2}} \circ L \circ \Upsilon^{\frac{1}{2}}\| = \sup_{y \in \mathcal{H}} \frac{\|\Sigma^{\frac{1}{2}} L \Upsilon^{\frac{1}{2}} y\|}{\|y\|} \leq 1. \tag{2.2.16}$$

The converse clearly holds by using the norm inequality in the right hand side of (2.2.15). 2$\Leftrightarrow$3: Clear from $(\Sigma^{\frac{1}{2}} \circ L \circ \Upsilon^{\frac{1}{2}})^* = \Upsilon^{\frac{1}{2}} \circ L^* \circ \Sigma^{\frac{1}{2}}$. 3$\Leftrightarrow$4: It follows from 1$\Leftrightarrow$2 replacing Σ by Υ and L by L^*, respectively. 1$\Leftrightarrow$5: For every $(x,u) \in \mathcal{H} \times \mathcal{G}$,

$$\begin{aligned}\langle (x,u) \mid \boldsymbol{V}(x,u)\rangle &= \big\langle x \mid \Upsilon^{-1}x - L^*u\big\rangle + \big\langle u \mid \Sigma^{-1}u - Lx\big\rangle \\ &= \big\langle x \mid (\Upsilon^{-1} - L^*\Sigma L)x\big\rangle + \langle \Sigma L x - u \mid Lx\rangle + \big\langle u \mid \Sigma^{-1}u - Lx\big\rangle \\ &= \big\langle x \mid (\Upsilon^{-1} - L^*\Sigma L)x\big\rangle + \|\Sigma^{-1}u - Lx\|_{\Sigma}^2\end{aligned} \tag{2.2.17}$$

and, by symmetry, we analogously obtain

$$\langle (x,u) \mid \boldsymbol{V}(x,u)\rangle = \big\langle u \mid (\Sigma^{-1} - L\Upsilon L^*)u\big\rangle + \|\Upsilon^{-1}x - L^*u\|_{\Upsilon}^2. \tag{2.2.18}$$

Hence, it follows from 1 and (2.2.17) that $\langle (x,u) \mid \boldsymbol{V}(x,u)\rangle \geq \|\Sigma^{-1}u - Lx\|_{\Sigma}^2$. Since 1 is equivalent to 4, (2.2.18) yields $\langle (x,u) \mid \boldsymbol{V}(x,u)\rangle \geq \|\Upsilon^{-1}x - L^*u\|_{\Upsilon}^2$ and we obtain (2.2.14). For the converse implication it is enough to combine (2.2.17) with (2.2.14).

For the last assertion, note that (2.2.14) implies, for every $(x,u) \in \mathcal{H} \times \mathcal{G}$,

$$\begin{cases} \langle (x,u) \mid \boldsymbol{V}(x,u)\rangle \geq \tau\|\Upsilon^{-1}x - L^*u\|^2 \\ \langle (x,u) \mid \boldsymbol{V}(x,u)\rangle \geq \sigma\|\Sigma^{-1}u - Lx\|^2. \end{cases} \tag{2.2.19}$$

By multiplying the first equation in (2.2.19) by $\lambda \in [0,1]$ and the second by $(1-\lambda)$ and summing up we obtain

$$\begin{aligned}\langle (x,u) \mid \boldsymbol{V}(x,u)\rangle &\geq \lambda\tau\|\Upsilon^{-1}x - L^*u\|^2 + (1-\lambda)\sigma\|\Sigma^{-1}u - Lx\|^2 \\ &\geq \min\{\lambda\tau, (1-\lambda)\sigma\}\|\boldsymbol{V}(x,u)\|^2.\end{aligned} \tag{2.2.20}$$

The result follows by noting that $\lambda \mapsto \min\{\lambda\tau, (1-\lambda)\sigma\}$ is maximized at $\lambda^* = \sigma/(\tau + \sigma)$. ☐

2.2.3 Convergence of Algorithm 2.2.2

Denote by $\boldsymbol{M} \colon \mathcal{H} \oplus \mathcal{G} \to 2^{\mathcal{H} \oplus \mathcal{G}}$ the maximally monotone operator [12, Proposition 2.7]

$$\boldsymbol{M} \colon (x,u) \mapsto (Ax + L^*u) \times (B^{-1}u - Lx). \tag{2.2.21}$$

For every strongly monotone self-adjoint linear operators $\Upsilon \colon \mathcal{H} \to \mathcal{H}$ and $\Sigma \colon \mathcal{G} \to \mathcal{G}$, consider the real Hilbert space $\boldsymbol{\mathcal{H}}$ obtained by endowing $\mathcal{H} \times \mathcal{G}$ with the inner product $\langle \cdot \mid \cdot \rangle_{\boldsymbol{U}}$, where $\boldsymbol{U} \colon (x,u) \mapsto (\Upsilon^{-1}x, \Sigma^{-1}u)$. More precisely,

$$\langle \cdot \mid \cdot \rangle_{\boldsymbol{U}} \colon \big((x,u),(y,v)\big) \mapsto \big\langle x \mid \Upsilon^{-1}y\big\rangle + \big\langle u \mid \Sigma^{-1}v\big\rangle, \tag{2.2.22}$$

and we denote the associated norm by $\|\cdot\|_{\boldsymbol{U}} = \sqrt{\langle\cdot\mid\cdot\rangle_{\boldsymbol{U}}}$. Observe that, since Υ and Σ are strongly monotone, the topologies of $\boldsymbol{\mathcal{H}}$ and $\mathcal{H}\oplus\mathcal{G}$ are equivalent.

Proposition 2.2.4. *In the context of Problem 2.2.1, let $\Sigma\colon\mathcal{G}\to\mathcal{G}$ and $\Upsilon\colon\mathcal{H}\to\mathcal{H}$ be strongly monotone self-adjoint linear operators such that $U=\Upsilon^{-1}-L^*\Sigma L$ is monotone, and define $\boldsymbol{T}:\boldsymbol{\mathcal{H}}\to\boldsymbol{\mathcal{H}}$ by*

$$\boldsymbol{T}\colon \begin{pmatrix} x\\ u\end{pmatrix}\mapsto\begin{pmatrix} x_+\\ u_+\end{pmatrix}=\begin{pmatrix} J_{\Upsilon A}\big(x-\Upsilon L^*\Sigma(\mathrm{Id}-J_{\Sigma^{-1}B})(Lx+\Sigma^{-1}u)\big)\\ \Sigma L(x_+-x)+\Sigma(\mathrm{Id}-J_{\Sigma^{-1}B})(Lx+\Sigma^{-1}u)\end{pmatrix}. \tag{2.2.23}$$

Then, the following hold:

1. *For every $(x,u)\in\boldsymbol{\mathcal{H}}$, we have*
$$\big(\Upsilon^{-1}(x-x_+),\Sigma^{-1}(u-u_+)\big)\in\boldsymbol{M}\big(x_+,u_+-\Sigma L(x_+-x)\big). \tag{2.2.24}$$

2. $\mathrm{Fix}\,\boldsymbol{T}=\boldsymbol{Z}=\mathrm{zer}\boldsymbol{M}$*.*

3. *For every $(\hat{x},\hat{u})\in\boldsymbol{Z}$ and $(x,u)\in\boldsymbol{\mathcal{H}}$ we have*
$$\begin{aligned}\|\boldsymbol{T}(x,u)-(\hat{x},\hat{u})\|_{\boldsymbol{U}}^2\le\|(x,u)-(\hat{x},\hat{u})\|_{\boldsymbol{U}}^2-\|(x,u)-\boldsymbol{T}(x,u)\|_{\boldsymbol{U}}^2&\\ +2\langle u_+-u\mid L(x_+-x)\rangle.&\end{aligned} \tag{2.2.25}$$

Proof. 1: From (2.2.23) and [3, Proposition 23.34(iii)] we obtain

$$\begin{aligned}\begin{pmatrix} x_+\\ u_+\end{pmatrix}=\boldsymbol{T}\begin{pmatrix} x\\ u\end{pmatrix}\ &\Leftrightarrow\ \begin{cases} x_+=J_{\Upsilon A}\big(x-\Upsilon L^*\Sigma(\mathrm{Id}-J_{\Sigma^{-1}B})(Lx+\Sigma^{-1}u)\big)\\ u_+=\Sigma L(x_+-x)+\Sigma(\mathrm{Id}-J_{\Sigma^{-1}B})(Lx+\Sigma^{-1}u)\end{cases}\\ &\Leftrightarrow\ \begin{cases} x_+=J_{\Upsilon A}\big(x-\Upsilon L^*(u_+-\Sigma L(x_+-x))\big)\\ u_+-\Sigma L(x_+-x)=J_{\Sigma B^{-1}}(\Sigma Lx+u)\end{cases}\\ &\Leftrightarrow\ \begin{cases} \Upsilon^{-1}(x-x_+)-L^*(u_+-\Sigma L(x_+-x))\in Ax_+\\ \Sigma^{-1}(u-u_+)+Lx_+\in B^{-1}\big(u_+-\Sigma L(x_+-x)\big),\end{cases}\end{aligned} \tag{2.2.26}$$

and the result follows from (2.2.21). 2: It follows from 1 and (2.2.1) that $\boldsymbol{T}(\hat{x},\hat{u})=(\hat{x},\hat{u})\Leftrightarrow(0,0)\in\boldsymbol{M}(\hat{x},\hat{u})\Leftrightarrow(\hat{x},\hat{u})\in\boldsymbol{Z}$. 3: Let $(\hat{x},\hat{u})\in\boldsymbol{Z}$. It follows from 2 that $(0,0)\in\boldsymbol{M}(\hat{x},\hat{u})$. Hence, 1 and the monotonicity of $\boldsymbol{M}$ in $\mathcal{H}\oplus\mathcal{G}$ yield

$$\begin{aligned}0&\le\big\langle\Upsilon^{-1}(x-x_+)\mid x_+-\hat{x}\big\rangle+\big\langle\Sigma^{-1}(u-u_+)\mid u_+-\hat{u}+\Sigma L(x-x_+)\big\rangle\\ &\overset{(2.2.22)}{=}\langle(x,u)-(x_+,u_+)\mid(x_+,u_+)-(\hat{x},\hat{u})\rangle_{\boldsymbol{U}}+\langle u-u_+\mid L(x-x_+)\rangle\\ &\overset{(2.2.10)}{=}\frac{1}{2}\big(\|(x,u)-(\hat{x},\hat{u})\|_{\boldsymbol{U}}^2-\|(x,u)-(x_+,u_+)\|_{\boldsymbol{U}}^2-\|(x_+,u_+)-(\hat{x},\hat{u})\|_{\boldsymbol{U}}^2\big)\\ &\qquad+\langle u-u_+\mid L(x-x_+)\rangle\end{aligned}$$

and the result follows. □

Remark 2.2.5. *1. Note that (2.2.23) and Algorithm 2.2.2 yield, for every $n \in \mathbb{N}$, $(x_{n+1}, u_{n+1}) = (x_{n+}, u_{n+}) = \boldsymbol{T}(x_n, u_n)$. This observation and the properties of $\boldsymbol{T}$ in Proposition 2.2.4 are crucial for the convergence of Algorithm 2.2.2 in Theorem 2.2.6 below.*

2. Proposition 2.2.4(3) can be written equivalently as, for every $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$ and $(x, u) \in \boldsymbol{\mathcal{H}}$, $\|\boldsymbol{T}(x,u) - (\hat{x}, \hat{u})\|_{\boldsymbol{U}}^2 \leq \|(x,u) - (\hat{x}, \hat{u})\|_{\boldsymbol{U}}^2 - \|(x,u) - \boldsymbol{T}(x,u)\|_{\boldsymbol{V}}^2$, where $\boldsymbol{V} \colon (x,u) \mapsto (\Upsilon^{-1}x - L^*u, \Sigma^{-1}u - Lx)$. Since $\Upsilon^{-1} - L^*\Sigma L$ is monotone, Proposition 2.2.3 asserts that $\boldsymbol{V}$ is self-adjoint, linear, and cocoercive, but not strongly monotone and, thus, $\|\cdot\|_{\boldsymbol{V}}^2$ does not define a norm.

Theorem 2.2.6. *In the context of Problem 2.2.1, let $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$ and consider the sequence $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$ defined by the Algorithm 2.2.2. Then, the following assertions hold:*

1. *$\sum_{n\geq 1} \|x_{n+1} - x_n\|^2 < +\infty$ and $\sum_{n\geq 1} \|u_{n+1} - u_n\|^2 < +\infty$.*
2. *There exists $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$ such that $(x_n, u_n) \rightharpoonup (\hat{x}, \hat{u})$ in $\mathcal{H} \oplus \mathcal{G}$.*

Proof. Let $\boldsymbol{x} = (x, u) \in \operatorname{Fix} \boldsymbol{T}$, for every $n \in \mathbb{N}$, denote by $\boldsymbol{x}_n = (x_n, u_n)$, and fix $n \geq 1$. It follows from Remark 2.2.5(1) that $\boldsymbol{x}_{n+1} = \boldsymbol{T}\boldsymbol{x}_n$ and from Proposition 2.2.4(2) that $\boldsymbol{x} \in \boldsymbol{Z}$. Therefore, Proposition 2.2.4(3) yields

$$\|\boldsymbol{x}_{n+1} - \boldsymbol{x}\|_{\boldsymbol{U}}^2 \leq \|\boldsymbol{x}_n - \boldsymbol{x}\|_{\boldsymbol{U}}^2 - \|\boldsymbol{x}_n - \boldsymbol{x}_{n+1}\|_{\boldsymbol{U}}^2 + 2\langle u_{n+1} - u_n \mid L(x_{n+1} - x_n)\rangle. \tag{2.2.27}$$

Hence, we deduce from the firm non-expansiveness of $J_{\Upsilon A}$ in $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_{\Upsilon^{-1}})$ [3, Proposition 23.34(i)] and the monotonicity of $U = \Upsilon^{-1} - L^*\Sigma L$ that

$$\begin{aligned}
&\langle u_{n+1} - u_n \mid L(x_{n+1} - x_n)\rangle \\
&\overset{(2.2.8)}{=} \langle \Sigma L(x_{n+1} - x_n) + v_n - \Sigma L(x_n - x_{n-1}) - v_{n-1} \mid L(x_{n+1} - x_n)\rangle \\
&= \langle x_{n+1} - x_n \mid L^*\Sigma L(x_{n+1} - x_n)\rangle + \langle L^*(v_n - v_{n-1}) \mid x_{n+1} - x_n\rangle \\
&\qquad - \langle \Sigma L(x_n - x_{n-1}) \mid L(x_{n+1} - x_n)\rangle \\
&= \langle x_{n+1} - x_n \mid L^*\Sigma L(x_{n+1} - x_n)\rangle + \big\langle \Upsilon^{-1}(x_n - x_{n-1}) \mid x_{n+1} - x_n\big\rangle \\
&\qquad - \langle (x_n - \Upsilon L^* v_n - (x_{n-1} - \Upsilon L^* v_{n-1})) \mid x_{n+1} - x_n\rangle_{\Upsilon^{-1}} \\
&\qquad - \langle \Sigma L(x_n - x_{n-1}) \mid L(x_{n+1} - x_n)\rangle \\
&\leq \langle x_{n+1} - x_n \mid L^*\Sigma L(x_{n+1} - x_n)\rangle + \big\langle \Upsilon^{-1}(x_n - x_{n-1}) \mid x_{n+1} - x_n\big\rangle \\
&\qquad - \|x_{n+1} - x_n\|_{\Upsilon^{-1}}^2 - \langle L^*\Sigma L(x_n - x_{n-1}) \mid x_{n+1} - x_n\rangle \\
&= -\|x_{n+1} - x_n\|_U^2 + \langle x_n - x_{n-1} \mid x_{n+1} - x_n\rangle_U \\
&\overset{(2.2.10)}{=} -\frac{1}{2}\|x_{n+1} - x_n\|_U^2 + \frac{1}{2}\|x_n - x_{n-1}\|_U^2 - \frac{1}{2}\|x_{n+1} + x_{n-1} - 2x_n\|_U^2 \\
&\leq -\frac{1}{2}\|x_{n+1} - x_n\|_U^2 + \frac{1}{2}\|x_n - x_{n-1}\|_U^2.
\end{aligned} \tag{2.2.28}$$

Therefore, it follows from (2.2.27) that

$$(\forall n \geq 1) \quad \|\boldsymbol{x}_{n+1} - \boldsymbol{x}\|_{\boldsymbol{U}}^2 + \|x_{n+1} - x_n\|_U^2 \leq \|\boldsymbol{x}_n - \boldsymbol{x}\|_{\boldsymbol{U}}^2 + \|x_n - x_{n-1}\|_U^2 - \|\boldsymbol{x}_n - \boldsymbol{x}_{n+1}\|_{\boldsymbol{U}}^2. \quad (2.2.29)$$

Thus, [23, Lemma 3.1] asserts that

$$(\forall \boldsymbol{x} \in \boldsymbol{Z}) \quad \big(\|\boldsymbol{x}_n - \boldsymbol{x}\|_{\boldsymbol{U}}^2 + \|x_n - x_{n-1}\|_U^2\big)_{n\geq 1} \quad \text{converges}, \quad (2.2.30)$$

that

$$\sum_{n\geq 1} \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\|_{\boldsymbol{U}}^2 < +\infty, \quad (2.2.31)$$

and 1 follows from (2.2.22) and the strong monotonicity of Υ^{-1} and Σ^{-1} [49, p.266].

In order to prove 2, note that, from 1 and the uniform continuity of U, we deduce $\|x_n - x_{n-1}\|_U^2 \to 0$. Hence, (2.2.30) implies that, for every $\boldsymbol{x} \in \boldsymbol{Z}$, $(\|\boldsymbol{x}_n - \boldsymbol{x}\|_{\boldsymbol{U}}^2)_{n\in\mathbb{N}}$ converges. Now, let $(\overline{x}, \overline{u}) \in \boldsymbol{\mathcal{H}}$ be a weak sequential cluster point of $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$, say $(x_{k_n}, u_{k_n}) \rightharpoonup (\overline{x}, \overline{u})$ in $\boldsymbol{\mathcal{H}}$. It is clear from (2.2.22) that we have $x_{k_n} \rightharpoonup \overline{x}$ in $\mathcal{H}$ and $u_{k_n} \rightharpoonup \overline{u}$ in $\mathcal{G}$ and from 1 that $x_{k_n+1} \rightharpoonup \overline{x}$ and $u_{k_n+1} \rightharpoonup \overline{u}$. Hence, since Proposition 2.2.4(1) yields

$$\big(\Upsilon^{-1}(x_{k_n} - x_{k_n+1}), \Sigma^{-1}(u_{k_n} - u_{k_n+1})\big) \in \boldsymbol{M}\big(x_{k_n+1}, u_{k_n+1} - \Sigma L(x_{k_n+1} - x_{k_n})\big), \quad (2.2.32)$$

we deduce from 1, the uniform continuity of ΣL, Υ^{-1}, and Σ^{-1}, and [3, Proposition 20.38(ii)], that $(0,0) \in \boldsymbol{M}(\overline{x}, \overline{u})$. Therefore, we conclude from [3, Lemma 2.47] that there exists $\hat{\boldsymbol{x}} \in \operatorname{Fix} \boldsymbol{T}$ such that $\boldsymbol{x}_n \rightharpoonup \hat{\boldsymbol{x}}$ and the result follows from the equivalence of the topologies of $\boldsymbol{\mathcal{H}}$ and $\mathcal{H} \oplus \mathcal{G}$. □

Remark 2.2.7. 1. *In the proof of Theorem 2.2.6, we can also deduce that any weak accumulation point of $((x_n, u_n))_{n\in\mathbb{N}}$ is in $\boldsymbol{Z}$ by using the points in the graph of A and B obtained from (2.2.26) and [3, Proposition 26.5(i)].*

2. *The method can include summable errors in the computation of resolvents and linear operators, by using standard Quasi-Féjer sequences. We prefer to not include this extension for simplicity of our algorithm formulation.*

3. *Consider the sequences $(v_n)_{n\in\mathbb{N}}$, $(z_n)_{n\in\mathbb{N}}$, $(x_n)_{n\in\mathbb{N}}$, $(u_n)_{n\in\mathbb{N}}$ defined by Algorithm 2.2.2 with starting point $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$. It follows from (2.2.8) and [3, Proposition 23.34(iii)] that, for every $n \in \mathbb{N}$,*

$$\begin{aligned} v_{n+1} &= \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_{n+1} + \Sigma^{-1}u_{n+1}) \\ &= J_{\Sigma B^{-1}}(\Sigma L x_{n+1} + u_{n+1}) \\ &= J_{\Sigma B^{-1}}(v_n + \Sigma L(2x_{n+1} - x_n)), \end{aligned}$$

leading to

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x_{n+1} = J_{\Upsilon A}(x_n - \Upsilon L^* v_n) \\ v_{n+1} = J_{\Sigma B^{-1}}(v_n + \Sigma L(2x_{n+1} - x_n)), \end{array} \right. \tag{2.2.33}$$

with starting point $(x_0, \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_0 + \Sigma^{-1}u_0)) \in \mathcal{H} \times \mathcal{G}$. *When* $\|\Sigma^{\frac{1}{2}} \circ L \circ \Upsilon^{\frac{1}{2}}\| < 1$, (2.2.33) *is equivalent to the proximal point algorithm applied to* $\boldsymbol{V}^{-1}\boldsymbol{M}$ *in* $(\mathcal{H}\times\mathcal{G}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{V}})$, *where* $\boldsymbol{V}\colon (x,u) \mapsto (\Upsilon^{-1}x - L^*u, \Sigma^{-1}u - Lx)$ *is strongly monotone in view of [48, Lemma 1]. Moreover, when* $\Upsilon = \tau\mathrm{Id}$, $\Sigma = \sigma\mathrm{Id}$, *and* $\sigma\tau\|L\|^2 < 1$, (2.2.33) *coincides with the PDS in* (2.2.7) *[20, 25, 41, 56]. As stated in Remark 2.2.5, under our assumptions* $\boldsymbol{V}$ *is no longer strongly monotone and the same approach cannot be used. On the other hand, a generalization of the previous approach is provided in [56] using the forward-backward splitting in order to allow cocoercive operators in the monotone inclusion when* $\boldsymbol{V}$ *is strongly monotone. In the optimization context, the inclusion of cocoercive operators allows for convex differentiable functions with* $\beta^{-1}-$*Lipschitzian gradients in the objective function and the convergence results are guaranteed under the more restrictive assumption* $\sigma\tau\|L\|^2 < 1 - \tau/2\beta$ *[25, Theorem 3.1]. Hence, the inclusion of cocoercive operators modifies our monotonicity assumption on* U *in Algorithm 2.2.2 distancing us from our main results. This leads us to consider this extension as part of further research.*

4. *We deduce from* (2.2.33) *and* (2.2.8) *that the primal iterates of SDR coincides with those of PDS in* (2.2.33) *and SDR includes an additional inertial step in the dual updates, more precisely,*

$$(\forall n \in \mathbb{N}) \quad u_{n+1} = \Sigma L(x_{n+1} - x_n) + v_n. \tag{2.2.34}$$

Hence, it follows from Theorem 2.2.6(1)&(2) and the uniform continuity of ΣL *that* $v_n \rightharpoonup \hat{u}$. *As a consequence, we obtain the primal-dual weak convergence of* (2.2.33) *when* $\|\Sigma^{\frac{1}{2}} \circ L \circ \Upsilon^{\frac{1}{2}}\| \leq 1$, *which generalizes [48, Theorem 1] and [25, Theorem 3.3], in the case when* $\Sigma = \sigma\mathrm{Id}$ *and* $\Upsilon = \tau\mathrm{Id}$, *to monotone inclusions and infinite dimensions.*

5. *By using product space techniques, Algorithm 2.2.2 allows us to solve*

$$\text{find } \hat{x} \in \mathcal{H} \text{ such that } \quad 0 \in A\hat{x} + \sum_{i=1}^{m} L_i^* B_i L_i \hat{x}, \tag{2.2.35}$$

where, for every $i \in \{1, \ldots, m\}$, $\mathcal{G}_i$ *is a real Hilbert space,* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *and* $B_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ *are maximally monotone, and* $L_i\colon \mathcal{H} \to \mathcal{G}_i$ *is a linear bounded operator. Indeed, by setting* $\mathcal{G} = \oplus_{1\leq i\leq m}\mathcal{G}_i$, $B\colon (u_i)_{1\leq i\leq m} \mapsto \times_{i=1}^{m} B_i u_i$, *and* $L\colon x \mapsto (L_i x)_{1\leq i\leq m}$, (2.2.35) *is equivalent to* (2.2.2). *Hence, by setting* $\Sigma\colon (u_i)_{1\leq i\leq m} \mapsto (\Sigma_i u_i)_{1\leq i\leq m}$, *where* $(\Sigma_i)_{1\leq i\leq m}$ *are strongly monotone operators, previous remark allows us to write Algorithm 2.2.2 as*

$$(\forall n\in\mathbb{N})\quad \left\lfloor \begin{array}{l} x_{n+1}=J_{\Upsilon A}(x_n-\Upsilon\sum_{i=1}^m L_i^* v_{i,n})\\ v_{1,n+1}=J_{\Sigma_1 B_1^{-1}}(v_{1,n}+\Sigma_1 L_1(2x_{n+1}-x_n))\\ \qquad\vdots\\ v_{m,n+1}=J_{\Sigma_m B_m^{-1}}(v_{m,n}+\Sigma_1 L_m(2x_{n+1}-x_n)), \end{array}\right. \tag{2.2.36}$$

and the weak convergence of $(x_n)_{n\in\mathbb{N}}$ *to a solution to* (2.2.35) *is guaranteed by Theorem 2.2.6, assuming that*

$$\Upsilon^{-1}-\sum_{i=1}^m L_i^*\Sigma_i L_i\quad \textit{is monotone.} \tag{2.2.37}$$

Note that (2.2.36) *has the same structure as the algorithm in [24, Corollary 6.2] without considering cocoercive operators or relaxation steps, but the convergence is guaranteed under the weaker assumption* (2.2.37)*.*

6. *Suppose that* $\operatorname{ran} L^*=\mathcal{H}$ *and that* $\Upsilon=(L^*\Sigma L)^{-1}$*. Then,* $U=\Upsilon^{-1}-L^*\Sigma L=0$ *and the operator* $\boldsymbol{T}$ *defined in* (2.2.23) *is firmly quasinonexpansive in* $\boldsymbol{\mathcal{H}}$*, in view of Proposition 2.2.4(3) and* (2.2.29)*. We thus generalize [54, Corollary 3]. Observe that, in the particular case when* $L=\mathrm{Id}$*, we have* $\Upsilon=\Sigma^{-1}$ *and the operator* $\boldsymbol{T}$ *defined in* (2.2.23) *reduces to* $\boldsymbol{T}\colon(x,u)\mapsto\Phi_A^{\Upsilon}(J_{\Upsilon B}(x+\Upsilon u)-\Upsilon u)$*, where*

$$\Phi_A^{\Upsilon}\colon\mathcal{H}\mapsto\mathcal{H}\times\mathcal{H}\colon z\mapsto(J_{\Upsilon A}z,\Upsilon^{-1}(J_{\Upsilon A}-\mathrm{Id})z). \tag{2.2.38}$$

In the case when $\Upsilon=\tau\mathrm{Id}$*, we recover the operator in [15, Proposition 5.18], which is inspired by [54]. Moreover, note that the inner product* $\langle\cdot\mid\cdot\rangle_{\boldsymbol{U}}$ *defined in* (2.2.22) *coincides with that in [54] (up to a multiplicative constant). Altogether, Theorem 2.2.6 generalizes [54] for an arbitrary operator* L *and non-standard metrics. It also generalizes [35, Theorem 5.1] from variational inequalities to arbitrary monotone inclusions and it provides the weak convergence of shadow sequences* $(J_{\tau A}z_n)_{n\in\mathbb{N}}$ *(not guaranteed in [35]).*

7. *Note that, by storing* $(Lx_n)_{n\in\mathbb{N}}$*, Algorithm 2.2.2 only needs to compute* L *once at each iteration. This observation is important in high dimensional problems in which the computation of* L *is numerically expensive.*

The following result establishes the reduction of Algorithm 2.2.2 to Douglas-Rachford splitting [30, 42] in the case when $\operatorname{ran} L=\mathcal{G}$.

Proposition 2.2.8. *In the context of Problem 2.2.1, assume* $\operatorname{ran} L = \mathcal{G}$ *and set* $\Sigma = (L\Upsilon L^*)^{-1}$. *Then, Algorithm 2.2.2 with starting point* $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$ *reduces to the recurrence*

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = J_{\Upsilon L^* B L}(2J_{\Upsilon A} z_n - z_n) + z_n - J_{\Upsilon A} z_n, \tag{2.2.39}$$

where $z_0 = x_0 - \Upsilon L^* \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_0 + \Sigma^{-1}u_0)$.

Proof. Note that $\operatorname{ran} L = \mathcal{G}$ yields, for every $u \in \mathcal{G}$, $\langle L\Upsilon L^* u \mid u\rangle \geq \tau\|L^*u\|^2 \geq \tau\alpha^2\|u\|^2$, where $\tau > 0$ is the strong monotonicity parameter of Υ and the existence of $\alpha > 0$ is guaranteed by [3, Fact 2.26]. Moreover, it follows from [3, Proposition 23.34(iii)&(ii)] that, for every $n \in \mathbb{N}$,

$$\begin{aligned}
v_{n+1} &\overset{(2.2.8)}{=} \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_{n+1} + \Sigma^{-1}u_{n+1}) \\
&= (\Sigma^{-1} + B^{-1})^{-1}(Lx_{n+1} + \Sigma^{-1}u_{n+1}) \\
&\overset{(2.2.8)}{=} (\Sigma^{-1} + B^{-1})^{-1}(L(2x_{n+1} - x_n) + \Sigma^{-1}v_n) \\
&= (L\Upsilon L^* + B^{-1})^{-1}L(2x_{n+1} - x_n + \Upsilon L^* v_n),
\end{aligned} \tag{2.2.40}$$

where the last equality follows from $\Sigma^{-1} = L\Upsilon L^*$. On the other hand, [3, Proposition 23.34(iii)] yields

$$\begin{aligned}
J_{\Upsilon L^* B L} &= \Upsilon^{\frac{1}{2}} J_{\Upsilon^{\frac{1}{2}} L^* B L \Upsilon^{\frac{1}{2}}} \Upsilon^{-\frac{1}{2}} \\
&= \Upsilon^{\frac{1}{2}}(\mathrm{Id} - \Upsilon^{\frac{1}{2}} L^*(L\Upsilon L^* + B^{-1})^{-1} L\Upsilon^{\frac{1}{2}})\Upsilon^{-\frac{1}{2}} \\
&= \mathrm{Id} - \Upsilon L^*(L\Upsilon L^* + B^{-1})^{-1}L,
\end{aligned} \tag{2.2.41}$$

where the second equality follows from [3, Proposition 23.25(ii)] since $(L\Upsilon^{\frac{1}{2}})(L\Upsilon^{\frac{1}{2}})^* = L\Upsilon L^*$ is invertible. Hence, we have

$$\begin{aligned}
z_{n+1} &\overset{(2.2.8)}{=} x_{n+1} - \Upsilon L^* v_{n+1} \\
&\overset{(2.2.40)}{=} x_{n+1} - \Upsilon L^*(L\Upsilon L^* + B^{-1})^{-1}L(2x_{n+1} - x_n + \Upsilon L^* v_n) \\
&\overset{(2.2.8)}{=} \left(\mathrm{Id} - \Upsilon L^*(L\Upsilon L^* + B^{-1})^{-1}L\right)(2J_{\Upsilon A} - \mathrm{Id})z_n + (\mathrm{Id} - J_{\Upsilon A})z_n \\
&\overset{(2.2.41)}{=} J_{\Upsilon L^* B L}(2J_{\Upsilon A} - \mathrm{Id})z_n + (\mathrm{Id} - J_{\Upsilon A})z_n
\end{aligned}$$

and z_0 is obtained from (2.2.8). ☐

Remark 2.2.9. *Note that* $\Sigma = (L\Upsilon L^*)^{-1}$ *is equivalent to* $\Sigma^{-1} - L\Upsilon L^* = 0$ *and, hence,* $\Upsilon^{-1} - L^*\Sigma L$ *is monotone in view of Proposition 2.2.3. Therefore, Proposition 2.2.8 and Theorem 2.2.6 provide the weak convergence of the non-standard metric version of DRS in (2.2.39) when* $\operatorname{ran} L = \mathcal{G}$. *This also extends the convergence result in [54].*

2.2.4 Split ADMM

In this section we study the numerical approximation of the following convex optimization problem.

Problem 2.2.10. *Let $\mathcal{H}$, $\mathcal{G}$, and $\mathcal{K}$ be real Hilbert spaces. Let $g \in \Gamma_0(\mathcal{K})$, let $f \in \Gamma_0(\mathcal{H})$, and let $T : \mathcal{K} \to \mathcal{G}$ and $K : \mathcal{G} \to \mathcal{H}$ be non-zero bounded linear operators such that $\operatorname{ran} T^* \cap \operatorname{dom} g^* \neq \varnothing$. Consider the following optimization problem*

$$\min_{y\in\mathcal{K}} \big(g(y) + f(KTy)\big) \tag{P}$$

together with the associated Fenchel-Rockafellar dual

$$\min_{x\in\mathcal{H}} \big(f^*(x) + g^*(-T^*K^*x)\big). \tag{D}$$

Moreover, consider the following Fenchel-Rockafellar dual problem associated to (D)

$$\min_{u\in\mathcal{G}} \big((g^* \circ -T^*)^*(u) + f(-Ku)\big). \tag{P^*}$$

We denote by S_P, S_D, and S_{P^*} the set of solutions to (P), (D), and (P^*), respectively.

In the particular case when $K = \mathrm{Id}$, Problem 2.2.10 is also considered in [29, 35, 55, 57] and ADMM is derived in [35] by applying DRS to the first order optimality conditions of (D), with $A = \partial f^*$ and $B = \partial(g^* \circ (-T^*K^*))$. We generalize this procedure by applying Algorithm 2.2.2 to (D) with $A = \partial f^*$, $B = \partial(g^* \circ (-T^*))$, and $L = K^*$. We thus obtain the Split-ADMM (SADMM), which splits K from T. We now provide an example in which this new formulation is relevant.

Example 2.2.11. *Let A and M be $n\times N$ and $m\times N$ real matrices, respectively, let $b \in \mathbb{R}^n$, let $\phi \in \Gamma_0(\mathbb{R}^m)$, let $h \in \Gamma_0(\mathbb{R}^n)$, and consider the optimization problem*

$$\min_{y\in\mathbb{R}^N} h(Ay - b) + \phi(My). \tag{2.2.42}$$

This problem arises in image and signal restoration and denoising [19, 22, 27, 43, 47, 51]. If M is symmetric and positive definite, as in graph Laplacian regularization (see, e.g., [43, Section II.B] and [47, 51] for alternative regularizations), there exist P unitary and D diagonal such that $M = PDP^\top$. Therefore, by setting $\eta \in \,]0,1[$, $K = PD^\eta P^\top$, $T = PD^{1-\eta}P^\top$, $g = \phi$, and $f = h(A\cdot -b)$, (2.2.42) is a particular instance of (P). In some instances, the resolvent computation of $\partial(g^* \circ -T^*)$ is simpler to solve than that of $\partial(g^* \circ -T^*K^*)$ when $\eta \sim 1$, since $D^{1-\eta} \sim \mathrm{Id}$. The numerical advantage of this approach is illustrated in an academical example in Section 2.2.5.2.

Other potential applications arise naturally when $y = \Phi z$, where z denotes frequencies or wavelet coefficients of an image y and Φ is a frame or unitary linear operator allowing *to pass from frequencies to images. Therefore,* (2.2.42) *is a particular case of* (P) *when* $f = h(\cdot - b)$, $g = \phi \circ M \circ \Phi$, $K = A$, *and* $T = \Phi$. *The properties of* T *in this case also make preferable to split* T *from* K.

First we provide some existence results and connections between problems (P), (D), and (P^*).

Proposition 2.2.12. *In the context of Problem 2.2.10, consider the inclusion*

$$\text{find} \quad (\hat{x}, \hat{u}) \in \mathcal{H} \times \mathcal{G} \quad \text{such that} \quad \begin{cases} 0 \in \partial f^*(\hat{x}) + K\hat{u} \\ 0 \in \partial(g^* \circ -T^*)^*(\hat{u}) - K^*\hat{x}. \end{cases} \tag{2.2.43}$$

1. *Suppose that there exists* $\hat{y} \in S_P$ *and that one of the following assertions hold:*
   (a) $0 \in \partial g(\hat{y}) + T^*K^*\partial f(KT\hat{y})$.
   (b) $0 \in \operatorname{sri}(\operatorname{dom} f - KT\operatorname{dom} g)$.

   Then, there exists $\hat{x} \in S_D$ *such that* $(\hat{x}, -T\hat{y})$ *is a solution to* (2.2.43).
2. *Suppose that there exists* $\hat{x} \in S_D$ *and that one of the following assertions hold:*
   (a) $0 \in \partial f^*(\hat{x}) - KT\partial g^*(-T^*K^*\hat{x})$.
   (b) $0 \in \operatorname{sri}(\operatorname{dom} g^* - T^*K^*\operatorname{dom} f^*)$.
   (c) $0 \in \operatorname{sri}(\operatorname{dom}(g^* \circ -T^*) - K^*\operatorname{dom} f^*)$ *and* $0 \in \operatorname{sri}(\operatorname{dom} g^* - \operatorname{ran} T^*)$.

   Then, there exists $\hat{y} \in S_P$ *such that* $(\hat{x}, -T\hat{y})$ *is a solution to* (2.2.43).
3. *Suppose that there exists* $(\hat{x}, \hat{u})$ *solution to* (2.2.43) *and that* $0 \in \operatorname{sri}(\operatorname{dom} g^* - \operatorname{ran} T^*)$. *Then,* $(\hat{x}, \hat{u}) \in S_D \times S_{P^*}$ *and there exists* $\hat{y} \in S_P$ *such that* $\hat{u} = -T\hat{y}$.

Proof. 1a: Let $\hat{x} \in \partial f(KT\hat{y})$ be such that $0 \in \partial g(\hat{y}) + T^*K^*\hat{x}$. Hence, it follows from [3, Corollary 16.30] that

$$\begin{cases} 0 \in \partial f^*(\hat{x}) - KT\hat{y} \\ 0 \in \partial g(\hat{y}) + T^*K^*\hat{x}, \end{cases} \tag{2.2.44}$$

and [12, Proposition 2.8(i)] implies $(\hat{y}, \hat{x}) \in S_P \times S_D$. By defining $\hat{u} = -T\hat{y}$, we obtain $0 \in \partial f^*(\hat{x}) + K\hat{u}$. Moreover, $\operatorname{ran} T^* \cap \operatorname{dom} g^* \neq \varnothing$ yields $g^* \circ (-T^*) \in \Gamma_0(\mathcal{K})$ and $-T(\partial g^*)(-T^*) \subset \partial(g^* \circ -T^*)$ in view of [3, Proposition 16.6(ii)]. Hence, we deduce from [3, Corollary 16.30] and (2.2.44) that

$$\begin{aligned} -T^*K^*\hat{x} \in \partial g(\hat{y}) \quad &\Leftrightarrow \quad \hat{y} \in \partial g^*(-T^*K^*\hat{x}) \\ &\Rightarrow \quad \hat{u} = -T\hat{y} \in -T\partial g^*(-T^*K^*\hat{x}) \\ &\Rightarrow \quad \hat{u} \in \partial(g^* \circ -T^*)(K^*\hat{x}) && (2.2.45) \\ &\Leftrightarrow \quad K^*\hat{x} \in \partial(g^* \circ -T^*)^*(\hat{u}) \\ &\Leftrightarrow \quad 0 \in \partial(g^* \circ -T^*)^*(\hat{u}) - K^*\hat{x}. && (2.2.46) \end{aligned}$$

Therefore, $(\hat{x}, -T\hat{y})$ is a solution to (2.2.43).

1b: By [3, Theorem 16.3 & Theorem 16.47(i)], $0 \in \partial(g + f \circ KT)(\hat{y}) = \partial g(\hat{y}) + T^*K^*\partial f(KT\hat{y})$. The result follows from 1a.

2a: Since, by taking $\hat{y} \in \partial g^*(-T^*K^*\hat{x})$ such that $0 \in \partial f^*(\hat{x}) - KT\hat{y}$, we obtain (2.2.44), the argument is analogous to that in 1a.

2b: By [3, Theorem 16.3 & Theorem 16.47(i)], $0 \in \partial(f^* + g^* \circ (-T^*K^*))(\hat{x}) = \partial f^*(\hat{x}) - KT(\partial g^*)(-T^*K^*\hat{x})$. The result hence follows from 2a.

2c: By [3, Theorem 16.3 & Theorem 16.47(i)], $0 \in \partial f^*(\hat{x}) + K\partial(g^* \circ -T^*)(K^*\hat{x})$. Moreover $0 \in \operatorname{sri}(\operatorname{dom} g^* - \operatorname{ran} T^*)$ and [3, Theorem 16.47] imply $0 \in \partial f^*(\hat{x}) - KT(\partial g^*)(-T^*K^*\hat{x})$. The result hence follows from 2a.

3: It follows from the second inclusion of (2.2.43) and [3, Theorem 16.47] that $\hat{u} \in \partial(g^* \circ (-T^*))(K^*\hat{x}) = -T\partial g^*(-T^*K^*\hat{x})$. Hence, there exists $\hat{y} \in \partial g^*(-T^*K^*\hat{x})$ such that $\hat{u} = -T\hat{y}$, which yields $0 \in \partial g(\hat{y}) + T^*K^*\hat{x}$. Therefore, by combining $\hat{u} = -T\hat{y}$ with the first inclusion of (2.2.43), we deduce (2.2.44) and the result follows from [12, Proposition 2.8(i)].

$\square$

Remark 2.2.13. *In the context of Proposition 2.2.12(3) we obtain the existence of* $\hat{y} \in S_P$ *such that* $(\hat{x}, \hat{y})$ *satisfies* (2.2.44). *If we additionally assume that* $\operatorname{ran} T$ *is closed, the second equation in* (2.2.44) *implies that* $\hat{y} \in \arg\min_{Ty=-\hat{u}} g(y)$. *We thus recover the results in [57, Lemma 2], obtained when* $K = -\mathrm{Id}$.

Algorithm 2.2.14 (Split-Alternating Direction Method of Multipliers (SADMM))**.** *In the context of Problem 2.2.10, let* $\Sigma\colon \mathcal{G} \to \mathcal{G}$ *and* $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ *be strongly monotone self-adjoint linear operators such that* $\Sigma^{-1} - K^*\Upsilon K$ *is monotone, let* $p_0 \in \mathcal{K}$, *and let* $(q_0, x_0) \in \mathcal{H} \times \mathcal{H}$. *Consider, the sequences defined by the recurrence*

$$
(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} y_n = x_n + \Upsilon(KTp_n - q_n) \\ p_{n+1} \in \underset{p \in \mathcal{K}}{\arg\min} \left(g(p) + \frac{1}{2}\|Tp - (Tp_n - \Sigma K^* y_n)\|^2_{\Sigma^{-1}} \right) \\ q_{n+1} = \operatorname{prox}_f^{\Upsilon}(\Upsilon^{-1}x_n + KTp_{n+1}) \\ x_{n+1} = x_n + \Upsilon(KTp_{n+1} - q_{n+1}). \end{array} \right. \tag{2.2.47}
$$

Observe that the existence and uniqueness of solutions to the convex optimization problem of the second step of (2.2.47) is not guaranteed without further hypotheses. The following result provides sufficient conditions for the existence of solutions to the optimization problem in (2.2.47), the equivalence between the sequences generated by Algorithm 2.2.2 and Algorithm 2.2.14, and the weak convergence of SADMM.

Theorem 2.2.15. *In the context of Problem 2.2.10, suppose that there exists a solution to* (2.2.43), *set*

$$
A = \partial f^*, \quad B = \partial(g^* \circ (-T^*)), \quad \text{and} \quad L = K^*, \tag{2.2.48}
$$

and assume that $0 \in \operatorname{sri}(\operatorname{dom} g^* - \operatorname{ran} T^*)$. *Then,* $(p_n)_{n\in\mathbb{N}}$ *defined in* (2.2.47) *exists and the following statements hold.*

1. *(SDR reduces to SADMM) Let* $(\tilde{x}_n)_{n\in\mathbb{N}}$, $(\tilde{u}_n)_{n\in\mathbb{N}}$, *and* $(\tilde{v}_n)_{n\in\mathbb{N}}$ *be the sequences generated by Algorithm 2.2.2 and set*

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \tilde{p}_{n+1} \in T^{-1}(-\tilde{v}_n) \\ \tilde{q}_{n+1} = \Upsilon^{-1}(\tilde{x}_n - \tilde{x}_{n+1} - \Upsilon K \tilde{v}_n). \end{cases} \tag{2.2.49}$$

Moreover, set $p_1 \in \mathcal{K}$ *such that* $Tp_1 = T\tilde{p}_1$, *and* $q_1 = \tilde{q}_1$, $x_1 = \tilde{x}_1$. *Then, sequences* $(p_n)_{n\geq 1}$, $(q_n)_{n\geq 1}$, *and* $(x_n)_{n\geq 1}$ *generated by Algorithm 2.2.14 satisfy, for every* $n \geq 1$, $T\tilde{p}_n = Tp_n$, $\tilde{q}_n = q_n$, *and* $\tilde{x}_n = x_n$.

2. *(SADMM reduces to SDR) Let* $(p_n)_{n\geq 1}$, $(q_n)_{n\geq 1}$, *and* $(x_n)_{n\geq 1}$ *be sequences generated by Algorithm 2.2.14 and define*

$$(\forall n \in \mathbb{N}) \quad u_{n+1} = \Sigma K^*(x_{n+1} - x_n) - Tp_{n+1}. \tag{2.2.50}$$

Moreover, set $\tilde{x}_0 = x_1$, $\tilde{u}_0 = u_1$, *and let* $(\tilde{x}_n)_{n\in\mathbb{N}}$ *and* $(\tilde{u}_n)_{n\in\mathbb{N}}$ *be the sequences generated by Algorithm 2.2.2. Then, for all* $n \in \mathbb{N}$, $\tilde{x}_n = x_{n+1}$ *and* $\tilde{u}_n = u_{n+1}$.

3. *Let* $(p_n)_{n\in\mathbb{N}}$, $(q_n)_{n\in\mathbb{N}}$, *and* $(x_n)_{n\in\mathbb{N}}$ *be sequences generated by Algorithm 2.2.14. Then, the following hold:*

   (a) *There exists* $(\hat{y}, \hat{x}, \hat{u}) \in S_P \times S_D \times S_{P^*}$ *such that* $(x_n, -Tp_n, q_n) \rightharpoonup (\hat{x}, \hat{u}, -K\hat{u})$ *and* $\hat{u} = -T\hat{y}$.

   (b) *Suppose that* $\operatorname{ran} T^* = \mathcal{K}$. *Then, there exists* $\hat{y} \in S_P$ *such that* $p_n \rightharpoonup \hat{y}$.

Proof. Note that $g^* \circ -T^* \in \Gamma_0(\mathcal{G})$, that [3, Corollary 16.53] yields $B = -T \circ (\partial g^*) \circ -T^*$, and that $J_{\Sigma^{-1}B} = (\mathrm{Id} - \Sigma^{-1}T(\partial g^*)(-T^*))^{-1}$. Therefore, it follows from [3, Corollary 16.30] that

$$\begin{aligned}
(\forall (u,y) \in \mathcal{G}^2) \quad y = J_{\Sigma^{-1}B} u \quad &\Leftrightarrow \quad (u - y) \in -\Sigma^{-1} T \partial g^*(-T^* y) \\
&\Leftrightarrow \quad (\exists p \in \mathcal{K}) \begin{cases} y = u + \Sigma^{-1}Tp \\ p \in \partial g^*(-T^*y) \end{cases} \\
&\Leftrightarrow \quad (\exists p \in \mathcal{K}) \begin{cases} y = u + \Sigma^{-1}Tp \\ 0 \in \partial g(p) + T^*y \end{cases} \\
&\Leftrightarrow \quad (\exists p \in \mathcal{K}) \begin{cases} y = u + \Sigma^{-1}Tp \\ p \in S(u), \end{cases}
\end{aligned} \tag{2.2.51}$$

where $S\colon u \mapsto \arg\min(g + \frac{1}{2}\|T\cdot + \Sigma u\|^2_{\Sigma^{-1}})$ and last equivalence follows from [3, Theorem 16.3] and simple gradient computations. We conclude $\operatorname{dom} S = \mathcal{G}$, $\operatorname{prox}^{\Sigma}_{g^*\circ -T^*} = \mathrm{Id} + \Sigma^{-1}TS$, and, therefore,

$$\Sigma(\mathrm{Id} - J_{\Sigma^{-1}B}) = \Sigma\big(\mathrm{Id} - \operatorname{prox}^{\Sigma}_{g^*\circ -T^*}\big) = -TS. \tag{2.2.52}$$

Thus, the optimization problem in (2.2.47) is equivalent to

$$(\forall n \in \mathbb{N}) \quad p_{n+1} \in S\big(K^*(x_n + \Upsilon(KTp_n - q_n)) - \Sigma^{-1}Tp_n\big) \tag{2.2.53}$$

and, hence, sequence $(p_n)_{n\in\mathbb{N}}$ exists.

1: It follows from (2.2.49), (2.2.8), (2.2.52), and (2.2.48) that, for every $n \in \mathbb{N}$, $T\tilde{p}_{n+1} = -\tilde{v}_n = TS(K^*\tilde{x}_n + \Sigma^{-1}\tilde{u}_n)$ and, thus, $\tilde{u}_{n+1} = \Sigma K^*\Upsilon(KT\tilde{p}_{n+1} - \tilde{q}_{n+1}) - T\tilde{p}_{n+1}$. Therefore, we have

$$(\forall n \geq 1) \quad T\tilde{p}_{n+1} = TS(K^*(\tilde{x}_n + \Upsilon(KT\tilde{p}_n - \tilde{q}_n)) - \Sigma^{-1}T\tilde{p}_n). \tag{2.2.54}$$

In addition, from (2.2.8), (2.2.48), and (2.2.12) we have, for every $n \in \mathbb{N}$, $\tilde{x}_{n+1} = \tilde{x}_n + \Upsilon KT\tilde{p}_{n+1} - \Upsilon\operatorname{prox}^{\Upsilon}_f(\Upsilon^{-1}\tilde{x}_n + KT\tilde{p}_{n+1})$ and, thus, (2.2.49) yields $\tilde{q}_{n+1} = \operatorname{prox}^{\Upsilon}_f(\Upsilon^{-1}\tilde{x}_n + KT\tilde{p}_{n+1})$. Altogether, we deduce

$$(\forall n \geq 1) \quad \left\lfloor \begin{array}{l} T\tilde{p}_{n+1} = TS(K^*(\tilde{x}_n + \Upsilon(KT\tilde{p}_n - \tilde{q}_n)) - \Sigma^{-1}T\tilde{p}_n) \\ \tilde{q}_{n+1} = \operatorname{prox}^{\Upsilon}_f(\Upsilon^{-1}\tilde{x}_n + KT\tilde{p}_{n+1}) \\ \tilde{x}_{n+1} = \tilde{x}_n + \Upsilon(KT\tilde{p}_{n+1} - \tilde{q}_{n+1}) \end{array}\right. \tag{2.2.55}$$

and the result follows from (2.2.53), $x_1 = \tilde{x}_1$, $q_1 = \tilde{q}_1$, and $Tp_1 = T\tilde{p}_1$.

2: Define

$$(\forall n \in \mathbb{N}) \quad \begin{cases} v_n = -Tp_{n+1} \\ z_n = x_n + \Upsilon KTp_{n+1} \end{cases} \tag{2.2.56}$$

and fix $n \geq 1$. Hence, we have

$$\begin{aligned} q_{n+1} &\overset{(2.2.47)}{=} \operatorname{prox}^{\Upsilon}_f(\Upsilon^{-1}x_n + KTp_{n+1}) \\ &\qquad \Leftrightarrow \quad x_n + \Upsilon(KTp_{n+1} - q_{n+1}) \overset{(2.2.12)}{=} \operatorname{prox}^{\Upsilon^{-1}}_{f^*}(x_n + \Upsilon KTp_{n+1}) \\ &\qquad \Leftrightarrow \quad x_{n+1} \overset{(2.2.48)}{=} J_{\Upsilon A}z_n. \end{aligned} \tag{2.2.57}$$

Moreover, from (2.2.53), (2.2.50), and (2.2.47), we obtain $p_{n+1} \in S(K^*x_n + \Sigma^{-1}u_n)$. Hence, (2.2.52), (2.2.48), and (2.2.56) yield $v_n = \Sigma(\mathrm{Id} - J_{\Sigma^{-1}B})(Lx_n + \Sigma^{-1}u_n)$. Altogether, from (2.2.50) we recover the recurrence in Algorithm 2.2.2 shifted by one iteration and, by setting $\tilde{x}_0 = x_1$ and $\tilde{u}_0 = u_1$ the result follows.

3a. Set $(u_n)_{n\geq 1}$ via (2.2.50) and define, for every $n \in \mathbb{N}$, $\tilde{x}_n = x_{n+1}$ and $\tilde{u}_n = u_{n+1}$. Then, 2 asserts that $(\tilde{x}_n)_{n\in\mathbb{N}}$ and $(\tilde{u}_n)_{n\in\mathbb{N}}$ are the sequences generated by Algorithm 2.2.2

with the operators defined in (2.2.48). Note that $A = \partial g^*$ and $B = \partial(g^* \circ (-T^*))$ are maximally monotone [3, Theorem 20.25] and that the set $\boldsymbol{Z}$ defined in (2.2.1) is the primal-dual solution set to the inclusion (2.2.43), which is non-empty by hypothesis. Then, by Theorem 2.2.6(2), there exists some $(\hat{x}, \hat{u})$ solution to (2.2.43) such that $(\tilde{x}_n, \tilde{u}_n) = (x_{n+1}, u_{n+1}) \rightharpoonup (\hat{x}, \hat{u})$. Moreover, Theorem 2.2.6(1) yields

$$x_{n+1} - x_n \to 0, \tag{2.2.58}$$

and, thus, (2.2.50) yields $-Tp_{n+1} = u_{n+1} - \Sigma K^*(x_{n+1} - x_n) \rightharpoonup \hat{u}$. Hence, since (2.2.47) yields, for every $n \in \mathbb{N}$, $q_{n+1} = \Upsilon^{-1}(x_n - x_{n+1}) + KTp_{n+1}$, the weak continuity of K and (2.2.58) imply $q_n \rightharpoonup -K\hat{u}$. We conclude that $(x_n, -Tp_n, q_n) \rightharpoonup (\hat{x}, \hat{u}, -K\hat{u})$. The result follows from Proposition 2.2.12(3).

3b. By 3a, there exists $\hat{y} \in S_P$ such that $Tp_n \rightharpoonup T\hat{y}$. Thus, for every $z \in \mathcal{K}$, there exists $w \in \mathcal{G}$ such that $z = T^*w$, which yields $\langle z \mid p_n - \hat{y}\rangle = \langle w \mid Tp_n - T\hat{y}\rangle \to 0$ and, hence, $p_n \rightharpoonup \hat{y}$. This concludes the proof. □

Remark 2.2.16. *1. Note that the existence of a sequence $(p_n)_{n\in\mathbb{N}}$ is guaranteed without any further assumption than $0 \in \operatorname{sri}(\operatorname{dom} g^* - \operatorname{ran} T^*)$. This result is weaker than strong monotonicity or full range assumptions made in [9, 35] and improves [32], in which this existence is assumed. Note that, even if there could exist a continuum of solutions to the optimization problem in (2.2.47), the image through T is unique, in view of (2.2.53) and (2.2.52).*

2. In the case when $K = \mathrm{Id}$, Theorem 2.2.15(1) recovers the reduction of DRS when $A = \partial f^*$ and $B = \partial(g^* \circ (-T^*))$ to ADMM and the convergence is guaranteed under weaker conditions than the strong monotonicity and full range assumptions used in [35, Section 5.1]. Under the assumption $\ker T = \{0\}$, this result is obtained in [46, Theorem 3.2].

3. Suppose that $K = \mathrm{Id}$. Observe that, given the sequence $(\tilde{v}_n)_{n\in\mathbb{N}}$ generated by SDR, Theorem 2.2.15(2) asserts that any sequence $(p_n)_{n\in\mathbb{N}}$ satisfying $-Tp_{n+1} = \tilde{v}_n$ allows the convergence of ADMM and its equivalence with DRS applied to the dual problem (D). The equivalence of ADMM with respect to DRS applied to the primal (P) is studied in [55, 57].

4. In the case when $K = \mathrm{Id}$, Theorem 2.2.15(2) provides the reduction of ADMM to DRS. Note that this reduction does not need any further assumption on T than $\operatorname{ran} T^* \cap \operatorname{dom} g^* \neq \varnothing$, which is weaker than $\ker T = \{0\}$, used in [46, Theorem 3.2] (see also [1, Appendix A] and [29, Proposition 3.43] in finite dimensions).

5. Theorem 2.2.15 provides the weak convergence of shadow sequences, improving [35, Theorem 5.1] in the optimization setting. In addition, Theorem 2.2.15 recovers the result in [29, Proposition 3.42] when K has full column rank in the finite dimensional setting.

The following result allows to deal with more general formulations involving two linear operators.

Corollary 2.2.17. *Let $\mathcal{H}$, $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{K}$ be real Hilbert spaces, let $g \in \Gamma_0(\mathcal{K})$, let $h \in \Gamma_0(\mathcal{H})$, and let $T : \mathcal{K} \to \mathcal{G}$, $J : \mathcal{H} \to \mathcal{H}$, and $K : \mathcal{G} \to \mathcal{H}$ be non-zero bounded linear operators such that $0 \in \operatorname{sri}(\operatorname{dom} g^* - \operatorname{ran} T^*)$, $0 \in \operatorname{sri}(\operatorname{dom} h^* - \operatorname{ran} J^*)$, and $0 \in \operatorname{sri}(KT\operatorname{dom} g + J\operatorname{dom} h)$. Consider the convex optimization problem*

$$\begin{aligned} \min_{y\in\mathcal{K}} \min_{v\in\mathcal{H}} \quad & g(y) + h(v) \\ s.t. \quad & KTy + Jv = 0, \end{aligned} \tag{2.2.59}$$

under the assumption that solutions exist. In addition, let $\Sigma : \mathcal{G} \to \mathcal{G}$ and $\Upsilon : \mathcal{H} \to \mathcal{H}$ be strongly monotone self-adjoint linear operators such that $\Sigma^{-1} - K^*\Upsilon K$ is monotone, let $p_0 \in \mathcal{K}$, let $v_0 \in \mathcal{H}$, let $x_0 \in \mathcal{H}$, and consider the routine:

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} y_n = x_n + \Upsilon(KTp_n + Jv_n) \\ p_{n+1} \in \arg\min\limits_{p\in\mathcal{K}} \left(g(p) + \frac{1}{2}\|Tp - (Tp_n - \Sigma K^* y_n)\|^2_{\Sigma^{-1}}\right) \\ v_{n+1} \in \arg\min\limits_{v\in\mathcal{J}} \left(h(v) + \frac{1}{2}\|Jv + KTp_{n+1} + \Upsilon^{-1}x_n\|^2_{\Upsilon}\right) \\ x_{n+1} = x_n + \Upsilon(KTp_{n+1} + Jv_{n+1}). \end{array} \right. \tag{2.2.60}$$

Then, there exists $(\hat{y}, \hat{v})$ solution to (2.2.59) such that the following hold:

1. *$Tp_n \rightharpoonup T\hat{y}$ and $Jv_n \rightharpoonup J\hat{v}$.*
2. *Suppose that $\operatorname{ran} T^* = \mathcal{K}$. Then, $p_n \rightharpoonup \hat{y}$.*
3. *Suppose that $\operatorname{ran} J^* = \mathcal{H}$. Then, $v_n \rightharpoonup \hat{v}$.*

Proof. Note that, by setting $f = (-J)\rhd h : q \mapsto \min_{Jv=-q} h(v)$, (2.2.59) can be equivalently written as

$$\min_{y\in\mathcal{K}} \left(g(y) + \min_{-Jv=KTy} h(v)\right) \equiv \min_{y\in\mathcal{K}} \big(g(y) + f(KTy)\big). \tag{2.2.61}$$

Since $0 \in \operatorname{sri}(\operatorname{dom} h^* - \operatorname{ran} J^*)$, [3, Corollary 15.28] yields $f = (h^* \circ -J^*)^* \in \Gamma_0(\mathcal{H})$. Hence, the problem in (2.2.59) is a particular instance of Problem 2.2.10 and it follows from (2.2.47), (2.2.12), and an argument analogous to that in (2.2.52) that

$$(\forall n \in \mathbb{N}) \quad q_{n+1} = \Upsilon^{-1}\left(\operatorname{Id} - \operatorname{prox}^{\Upsilon^{-1}}_{h^*\circ -J^*}\right)(x_n + \Upsilon KTp_{n+1}) = -Jv_{n+1}, \tag{2.2.62}$$

where v_{n+1} is defined in (2.2.60). Hence, (2.2.60) is a particular instance of Algorithm 2.2.14. Moreover, [3, Proposition 12.36(i)] yields $0 \in \operatorname{sri}(KT\operatorname{dom} g + J\operatorname{dom} h) = \operatorname{sri}(KT\operatorname{dom} g -$

$\operatorname{dom} f$) and Proposition 2.2.12(1b) implies the existence of a solution to (2.2.43). Altogether, Theorem 2.2.15(3) asserts that there exists $(\hat{y},\hat{x}) \in S_P \times S_D$ such that $(x_n, -Tp_n, q_n) \rightharpoonup (\hat{x}, -T\hat{y}, KT\hat{y})$ and $\hat{u} = -T\hat{y} \in S_{P^*}$. Moreover, since $0 \in \operatorname{sri}(\operatorname{dom} h^* - \operatorname{ran} J^*)$, it follows from (2.2.61) and [3, Corollary 15.28(i)] that there exists $\hat{v} \in \mathcal{H}$ such that $(\hat{y},\hat{v})$ is a solution to (2.2.59). In particular, $Tp_n \rightharpoonup T\hat{y}$ and $q_n = -Jv_n \rightharpoonup KT\hat{y} = -J\hat{v}$, which yields 1. Assertions 2 and 3 follow analogously as in the proof of Theorem 2.2.15(3b). □

Remark 2.2.18. *1. In the context of Corollary 2.2.17, let $U = \Upsilon^{-1} - K\Sigma K^*$ and $V = \Sigma^{-1} - K^*\Upsilon K$, which are monotone in view of Proposition 2.2.3. Then, Algorithm 2.2.14 can be written equivalently as*

$$\left\lfloor \begin{array}{l} p_{n+1} \in \underset{p\in\mathcal{K}}{\arg\min}\left(g(p) + \frac{1}{2}\|KTp + Jv_n + \Upsilon^{-1}x_n\|^2_{\Upsilon} + \frac{1}{2}\|p - p_n\|^2_{T^*VT}\right) \\ v_{n+1} \in \underset{v\in\mathcal{H}}{\arg\min}\left(h(v) + \frac{1}{2}\|KTp_{n+1} + Jv + \Upsilon^{-1}x_n\|^2_{\Upsilon}\right) \\ x_{n+1} = x_n + \Upsilon(KTp_{n+1} + Jv_{n+1}), \end{array}\right. \tag{2.2.63}$$

which is a non-standard version of the preconditioned ADMM (PADMM) [9] without proximal quadratic term in the second optimization problem of (2.2.63). It considers the augmented Lagrangian with non-standard metric

$$\mathcal{L}_{\Upsilon}\colon (p,v,x) \mapsto g(p) + h(v) + \langle x \mid KTp + Jv\rangle + \frac{1}{2}\|KTp + Jv\|^2_{\Upsilon}, \tag{2.2.64}$$

which generalizes the classical augmented Lagrangian $\mathcal{L}_{r\mathrm{Id}}$ for some $r > 0$. Without the strong monotonicity assumptions used in [9, Theorem 2.1 & Theorem 3.1], the sequences of algorithm (2.2.60) are well defined and Corollary 2.2.17 provides weak convergence. Moreover, in the case when $J = -\mathrm{Id}$ and $\Upsilon = r\mathrm{Id}$, Corollary 2.2.17 ensures convergence under weaker assumptions than [52, Algorithm 2] (see also [4] for a variant involving a differentiable convex function). In [59], a non-standard metric is included only in the multiplier update step of [52, Algorithm 2], but the convergence of the iterates is not obtained.

2. In the case when $K = \mathrm{Id}$ and $\Sigma = \Upsilon^{-1}$, the algorithm in (2.2.63) reduces to the ADMM algorithm with the augmented Lagrangian with non-standard metric (2.2.64), which, given $(q_0, x_0) \in \mathcal{H}\times\mathcal{H}$, iterates

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} \in \underset{p\in\mathcal{K}}{\arg\min}\left(g(p) + \frac{1}{2}\|Tp + Jv_n + \Upsilon^{-1}x_n\|^2_{\Upsilon}\right) \\ v_{n+1} \in \underset{v\in\mathcal{H}}{\arg\min}\left(h(v) + \frac{1}{2}\|Tp_{n+1} + Jv + \Upsilon^{-1}x_n\|^2_{\Upsilon}\right) \\ x_{n+1} = x_n + \Upsilon(Tp_{n+1} + Jv_{n+1}). \end{array}\right. \tag{2.2.65}$$

In the particular case when $\Upsilon = \tau\mathrm{Id}$, it reduces to ADMM [7] and [32, 35, 36, 38] when $J = -\mathrm{Id}$.

3. *As in Remark 2.2.16(1), sequences* $(Tp_n)_{n\in\mathbb{N}}$ *and* $(Jv_n)_{n\in\mathbb{N}}$ *in* (2.2.65) *are unique even if the solutions to the optimization problems in* (2.2.65) *are not unique. The uniqueness of* $(p_n)_{n\in\mathbb{N}}$ *(resp.* $(v_n)_{n\in\mathbb{N}}$*) is guaranteed, e.g., if* g *(resp.* h*) is strictly convex or if* $\operatorname{ran} T^* = \mathcal{K}$ *(resp.* $\operatorname{ran} J^* = \mathcal{H}$*).*

The following corollary is a direct consequence of Theorem 2.2.15 when $T = \mathrm{Id}$.

Corollary 2.2.19. *In the context of Problem 2.2.10, suppose that* $T = \mathrm{Id}$ *and that there exists a solution to* (2.2.43). *Let* $\Sigma\colon \mathcal{G} \to \mathcal{G}$ *and* $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ *be strongly monotone self-adjoint linear operators such that* $\Sigma^{-1} - K^*\Upsilon K$ *is monotone, let* $p_0 \in \mathcal{K}$*, let* $(q_0, x_0) \in \mathcal{H} \times \mathcal{H}$*, and consider the sequences* $(p_n)_{n\in\mathbb{N}}$ *and* $(x_n)_{n\in\mathbb{N}}$ *generated by the recurrence*

$$
(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} y_n = x_n + \Upsilon(Kp_n - q_n) \\ p_{n+1} = \operatorname{prox}_g^{\Sigma^{-1}}\big(p_n - \Sigma K^* y_n\big) \\ q_{n+1} = \operatorname{prox}_f^{\Upsilon}(\Upsilon^{-1}x_n + Kp_{n+1}) \\ x_{n+1} = x_n + \Upsilon(Kp_{n+1} - q_{n+1}). \end{array} \right. \tag{2.2.66}
$$

Then, there exists $(\hat{y}, \hat{x}) \in S_P \times S_D$ *such that* $(p_n, x_n) \rightharpoonup (\hat{y}, \hat{x})$.

Remark 2.2.20. 1. *Note that the explicit method proposed in Corollary 2.2.19 includes two multiplier updates as the algorithm in [21, Algorithm I]. Our method allows for different step-sizes in primal and dual updates and the main distinction is that the third step in* (2.2.66) *includes the information of its second step, while the algorithm in [21, Algorithm I] uses the information of previous iteration.*

2. *Note that* (2.2.66) *and* (2.2.12) *yield, for every* $n \in \mathbb{N}$,

$$
\begin{aligned} x_{n+1} &= x_n + \Upsilon Kp_{n+1} - \Upsilon q_{n+1} \\ &= \Upsilon(\mathrm{Id} - \operatorname{prox}_f^{\Upsilon})(\Upsilon^{-1}x_n + Kp_{n+1}) \\ &= \operatorname{prox}_{f^*}^{\Upsilon^{-1}}(x_n + \Upsilon Kp_{n+1}) \end{aligned} \tag{2.2.67}
$$

and $y_{n+1} = x_{n+1} + \Upsilon(Kp_{n+1} - q_{n+1}) = 2x_{n+1} - x_n$. *Therefore, when* $\|\Upsilon^{\frac{1}{2}} \circ K^* \circ \Sigma^{\frac{1}{2}}\| < 1$, (2.2.66) *reduces to the algorithm proposed in [48] applied to the dual problem* $\min(f^* + g^* \circ -K^*)(\mathcal{H})$. *Hence, Corollary 2.2.19 is a generalization of [48, Theorem 1] in this context.*

3. *Observe that the second step in* (2.2.66) *is explicit, differing from the first step in ADMM* (2.2.65), *which is implicit. This feature allows for an algorithm with very low computational cost by iteration. However, the number of iterations may be much larger than those of ADMM in some instances, as we verify numerically in Section 2.2.5.2.*

2.2.5 Numerical experiments

In this section we provide two numerical experiments. In the first experiment we compare SDR with several schemes in the literature for solving the total variation image restoration problem. In the second experiment we consider an academic example in which splitting K from T has numerical advantages with respect to ADMM.

2.2.5.1 Total variation image restoration

A classical model in image processing is the total variation image restoration [50], which aims at recovering an image from a blurred and noisy observation under piecewise constant assumption on the solution. The model is formulated via the optimization problem

$$\min_{x\in[0,255]^N} \frac{1}{2}\|Rx-b\|_2^2+\alpha\|\nabla x\|_1 =: F^{TV}(x), \tag{2.2.68}$$

where $x \in [0,255]^N$ is the image of $N = N_1 \times N_2$ pixels to recover from a blurred and noisy observation $b \in \mathbb{R}^m$, $R : \mathbb{R}^N \to \mathbb{R}^m$ is a linear operator representing a Gaussian blur, the discrete gradient $\nabla\colon x \mapsto \nabla x = (D_1x, D_2x)$ includes horizontal and vertical differences through linear operators D_1 and D_2, respectively, its adjoint ∇^* is the discrete divergence (see, e.g., [18]), and $\alpha \in \,]0,+\infty[$. A difficulty in this model is the presence of the non-smooth ℓ^1 norm composed with the discrete gradient operator ∇, which is non-differentiable and its proximity operator has not a closed form.

Note that, by setting $f = \|R\cdot -b\|^2/2$, $g_1 = \alpha\|\cdot\|_1$, and $g_2 = \iota_{[0,255]^N}$, $L_1 = \nabla$, and $L_2 = \mathrm{Id}$, (2.2.68) can be reformulated as $\min(f + g_1 \circ L_1 + g_2 \circ L_2)$ or equivalently as (qualification condition holds)

$$\text{find } \hat{x} \in \mathbb{R}^N \text{ such that } 0 \in \partial f(\hat{x}) + L_1^*\partial g_1(L_1x) + L_2^*\partial g_2(L_2\hat{x}), \tag{2.2.69}$$

which is a particular instance of (2.2.35), in view of [3, Theorem 20.25]. Moreover, for every $\tau > 0$, $J_{\tau\partial f} = (\mathrm{Id} + \tau R^*R)^{-1}(\mathrm{Id} - \tau R^*b)$, for every $i \in \{1,2\}$, $J_{\tau(\partial g_i)^{-1}} = \tau(\mathrm{Id} - \mathrm{prox}_{g_i/\tau})(\mathrm{Id}/\tau)$, $\mathrm{prox}_{g_2/\tau} = P_{[0,255]^N}$, and $\mathrm{prox}_{g_1/\tau} = \mathrm{prox}_{\alpha\|\cdot\|_1/\tau}$ is the component-wise soft thresholder, computed in [3, Example 24.34]. Note that $(\mathrm{Id} + \tau R^*R)^{-1}$ can be computed efficiently via a diagonalization of R using the fast Fourier transform F [40, Section 4.3]. Altogether, Remark 2.2.7(5) allows us to write Algorithm 2.2.2 as Algorithm 1 below, where we set $\Upsilon = \tau\mathrm{Id}$, $\Sigma_1 = \sigma_1\mathrm{Id}$, and $\Sigma_2 = \sigma_2\mathrm{Id}$, for $\tau > 0$, $\sigma_1 > 0$, and $\sigma_2 > 0$. We denote by $\mathcal{R}$ the primal-dual error

$$\mathcal{R} : (x_+, u_+, x, u) \mapsto \sqrt{\frac{\|(x_+,u_+)-(x,u)\|^2}{\|(x,u)\|^2}} \tag{2.2.70}$$

and by $\varepsilon > 0$ the convergence tolerance. The error $\mathcal{R}$ is inspired from (2.2.31) in the proof of Theorem 2.2.6.

Algorithm 1

Fix $x_0 \in \mathbb{R}^N, v_{1,0} \in \mathbb{R}^m, v_{2,0} \in \mathbb{R}^{2N}$, $\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 \leq 1$, and $r_0 > \varepsilon > 0$.
while $r_n > \varepsilon$ **do**
  $x_{n+1} = (\mathrm{Id} + \tau R^* R)^{-1}(x_n - \tau\nabla^* v_{1,n} - \tau v_{2,n} - \tau R^* b)$
  $v_{1,n+1} = \sigma_1(\mathrm{Id} - \mathrm{prox}_{\alpha\|\cdot\|_1/\sigma_1})(v_{1,n}/\sigma_1 + \nabla(2x_{n+1} - x_n))$
  $v_{2,n+1} = \sigma_2\big(\mathrm{Id} - P_{[0,255]^N}\big)\left(v_{2,n}/\sigma_2 + 2x_{n+1} - x_n\right)$
  $r_n = \mathcal{R}\big((x_{n+1}, v_{1,n+1}, v_{2,n+1}), (x_n, v_{1,n}, v_{2,n})\big)$
end while
return $(x_{n+1}, v_{1,n+1}, v_{2,n+1})$

In this case, (2.2.37) reduces to the monotonicity of $(\tau^{-1} - \sigma_2)\mathrm{Id} - \sigma_1\nabla^*\nabla$, which is equivalent to

$$\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 \leq 1, \tag{2.2.71}$$

in view of Proposition 2.2.3. By using the power iteration [44] with tolerance 10^{-9}, we obtain $\|\nabla\|^2 \approx 7.9997$. This is consistent with [17, Theorem 3.1].

Observe that, when $\sigma_1 = \sigma_2 = \sigma$, Algorithm 1 reduces to the algorithm proposed in [20] (when $\sigma\tau(\|\nabla\|^2 + 1) < 1$) or [25, Theorem 3.3], where the case $\sigma\tau(\|\nabla\|^2 + 1) = 1$ is included.

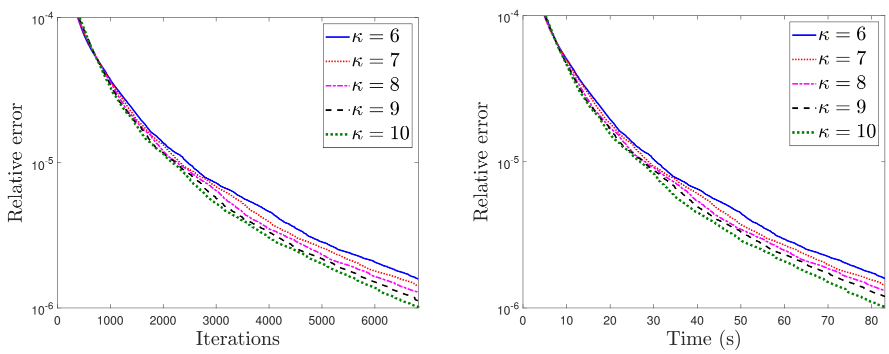
We provide two main numerical experiments in this subsection: we first compare the efficiency of Algorithm 1 when the step-sizes achieve the boundary in (2.2.71), verifying that the efficiency is better when the equality is achieved. Next, we compare the performance of different methods in the literature with optimal step-sizes. For these comparisons, we consider the test image of 256×256 pixels ($N_1 = N_2 = 256$) in Figure 2.4a[2] (denoted by $\overline{x}$). The operator blur R is set as a Gaussian blur of size 9×9 and standard deviation 4 (applied by MATLAB function *fspecial*) and the observation b is obtained by $b = R\overline{x} + e \in \mathbb{R}^{m_1 \times m_2}$, where $m_1 = m_2 = 256$ and e is an additive zero-mean white Gaussian noise with standard deviation 10^{-3} (using *imnoise* function in MATLAB). We generate 20 random realization of random variable e leading to 20 observations $(b_i)_{1\leq i\leq 20}$.

In Table 2.1 we study the efficiency of Algorithm 1, in the simpler case when $\sigma_1 = \sigma_2 = \sigma$, as parameters σ and τ approach the boundary $\sigma\tau(\|\nabla\|^2 + 1) = 1$. In particular, we set $\sigma = \tau = \kappa/(10\sqrt{1 + \|\nabla\|^2})$ for $\kappa \in \{6, 7, 8, 9, 10\}$. We provide the averages of CPU time, number of iterations, and percentage of error between objective values $F^{TV}(\overline{x})$ and $F^{TV}(x_n)$ obtained by applying Algorithm 1 for the 20 observations $(b_i)_{1\leq i\leq 20}$ and for $\kappa \in \{6, 7, 8, 9, 10\}$. The tolerance is set as $\varepsilon = 10^{-6}$. We observe that the algorithm becomes more efficient (in time and iterations) and accurate (in terms of the objective value) as long as parameters approach the boundary. This conclusion is confirmed in Figure 2.1, which shows the performance obtained with the observation b_{13}. Henceforth, we consider only parameters in the boundary of (2.2.71).

[2] Image *Circles* obtained from `http://links.uwaterloo.ca/Repository.html`

Table 2.1: Averages of CPU time, number of iterations, and percentage of error in the objective value obtained from Algorithm 1 with $\tau = \sigma_1 = \sigma_2 = \kappa/(10\sqrt{1+\|\nabla\|^2})$ and tolerance 10^{-6}.

	$\varepsilon = 10^{-6}$		
κ	Av. Time(s)	Av. Iter.	Av.% error o.v.
6	43.22	8729	0.3541
7	40.23	8179	0.3536
8	38.56	7725	0.3533
9	36.43	7340	0.3530
10	34.66	7003	0.3528



(a) (b)

Figure 2.1: Comparison of Algorithm 1 with $\tau = \sigma_1 = \sigma_2 = \kappa/(10\sqrt{1+\|\nabla\|^2})$), for image reconstruction from observation b_{13}.

Next, we compare Algorithm 1 when $\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 = 1$, with alternative algorithms in [25, Theorem 3.3], [25, Theorem 3.1] or [56, Corollary 4.2], [12, Theorem 3.1], and [45], which are called "Condat", "Condat-Vũ", "MS", and "AFBS", respectively. In order to provide a fair comparison in our example, we approximate the best step-sizes by considering a mesh on the feasible set defined by the conditions allowing convergence for each algorithm. In the case when $\varepsilon = 10^{-6}$, the best performance of Condat-Vũ is obtained by setting $\tau = 1.2$ and $\sigma = 0.99 \cdot (2-\tau)/(2\tau\|\nabla\|^2)$ which is next to the boundary of condition $\sigma\tau\|\nabla\|^2 < (1-\tau/2)$. For MS, the performance is better when the only step-size τ is next to the boundary of the condition $\tau < 1/\sqrt{1+\|\nabla\|^2}$, which leads us to set $\tau = 0.99/\sqrt{1+\|\nabla\|^2}$. For AFBS, we found as best parameters $\tau = 0.13$ and $\lambda_n \equiv 1.7/(65n+10)^{0.505}$ (see [45]). In the case of Condat, we consider 34 cases of parameters τ and σ satisfying $\sigma\tau(1+\|\nabla\|^2) = 1$, by setting $\tau_k = \delta^k/(800\sqrt{1+\|\nabla\|^2})$ and $\sigma_k = 800/(\delta^k\sqrt{1+\|\nabla\|^2})$, where $\delta = 800^{1/8}$ and $k \in \{1,\ldots,34\}$. For Algorithm 1 we consider the same parameters $(\tau_k)_{1\le k\le 34}$ than those in Condat, and we set $\sigma_{1,k}^{\ell} = (1-\ell)/(\tau_k\|\nabla\|^2)$ and $\sigma_{2,k}^{\ell} = \ell/\tau_k$, for $\ell \in 10^{-1}\cdot\{5, 0.1, 0.05, 0.01, 0.005, 0.003\}$, in view of (2.2.71). In Table 2.2 we provide the averages of CPU time, number of iterations, and the percentage of error between objective values $F^{TV}(\overline{x})$ and $F^{TV}(x_n)$ obtained by previous algorithms with tolerance $\varepsilon = 10^{-6}$ considering the observations $(b_i)_{1\le i\le 20}$. We show the best 5 cases for Algorithm 1 ($k \in \{20,\ldots,24\}$) and the best case for Condat ($k = 22$). We observe that Algorithm 1 and Condat reduce drastically the computational time and iterations obtained in Table 2.1, which shows the advantage of searching optimal parameters in the boundary of the condition of convergence. We also observe in Table 2.2 that Algorithm 1 ($k = 22$ and $\ell = 0.001$) is the most efficient method for this tolerance, followed closely by Condat ($k = 22$). Both methods outperform drastically the competitors. In Figure 2.2 we show the relative error versus iterations and time for the observation b_{13}, confirming previous results.

In order to make a more precise comparison of Algorithm 1 and Condat, we consider a smaller tolerance $\varepsilon = 10^{-8}$. The obtained results are shown in Table 2.3 and Figure 2.3. We observe that Algorithm 1 ($k = 21$ and $\ell = 0.001$) achieves the tolerance in approximately 11% less CPU time than Condat in its best case ($k = 21$). The efficiency in the case of the observation b_{13} is illustrated in Figure 2.3.

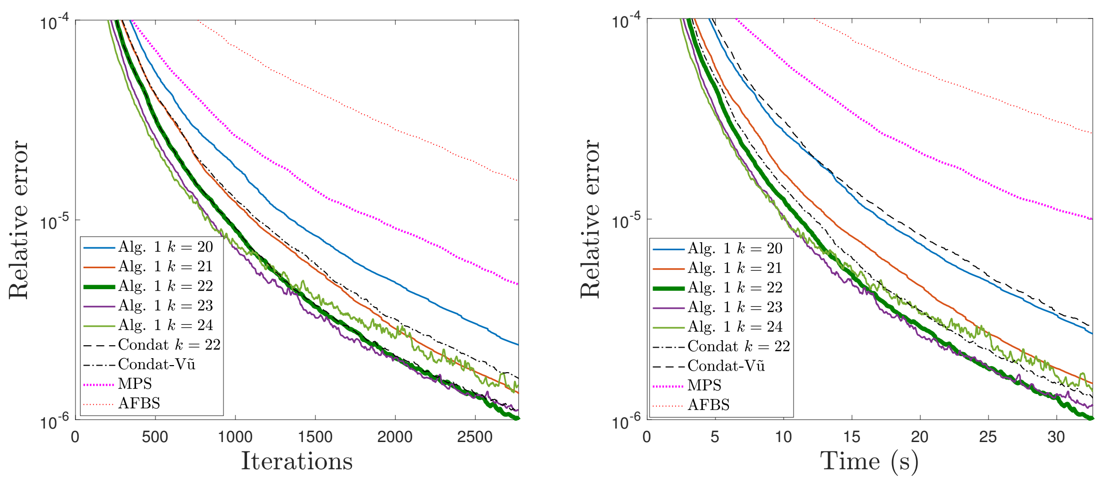
The reconstructed images, after 100 iterations, for the different algorithms are shown in Figure (2.4). The best reconstruction, in terms of objective value F^{TV} and PSNR (Peak signal-to-noise ratio), are obtained by Condat and Algorithm 1.

2.2.5.2 Split-ADMM in an academical example

In this section, we implement Algorithm 2.2.14, Corollary 2.2.19, and ADMM in (2.2.65) for solving an academical example in the context of Example 2.2.11. We compare their

Table 2.2: Averages of CPU time, number of iterations, and percentage of error in the objective value for Algorithm 1 with $\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 = 1$, Condat, Condat-Vũ, AFBS, and MS with tolerance 10^{-6}.

			$\varepsilon = 10^{-6}$		
Algorithm	τ	σ_1	Av. Time(s)	Av. Iter.	Av. % error o.v.
Alg.1	0.77	0.16	21.12	4106	0.3531
	1.17	0.11	15.33	3223	0.3562
	1.77	0.07	13.97	2787	0.3649
	2.69	0.05	14.36	2891	0.3771
	4.09	0.03	16.23	3372	0.3907
Condat	1.77	-	14.89	2853	0.3673
Condat-Vũ	1.2	-	28.19	3539	0.3738
MS	0.33	-	62.48	6193	0.3506
AFBS	0.13	-	85.76	11104	0.6611

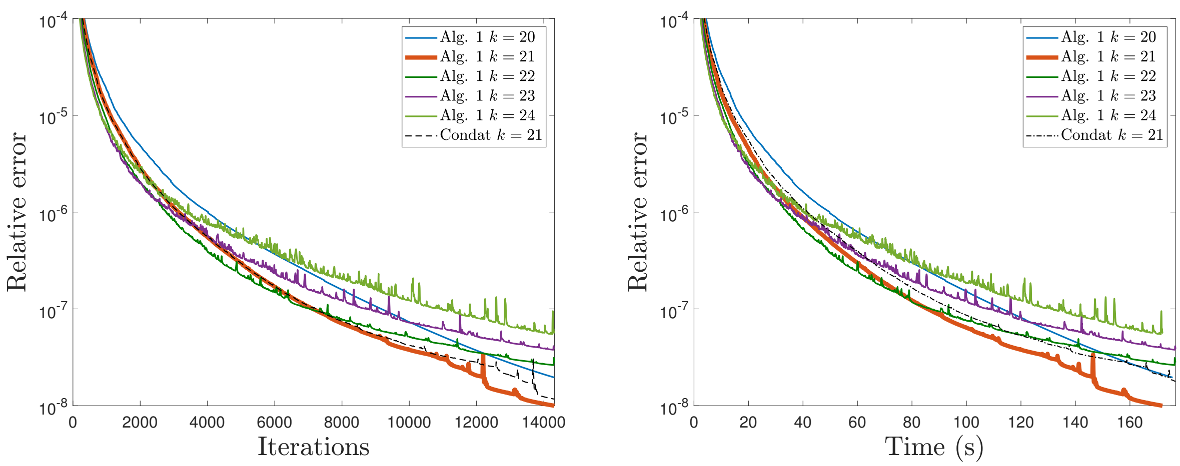


(a) (b)

Figure 2.2: Comparison of Algorithm 1 with $\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 = 1$, Condat, Condat-Vũ, AFBS, and MS (observation b_{13}).

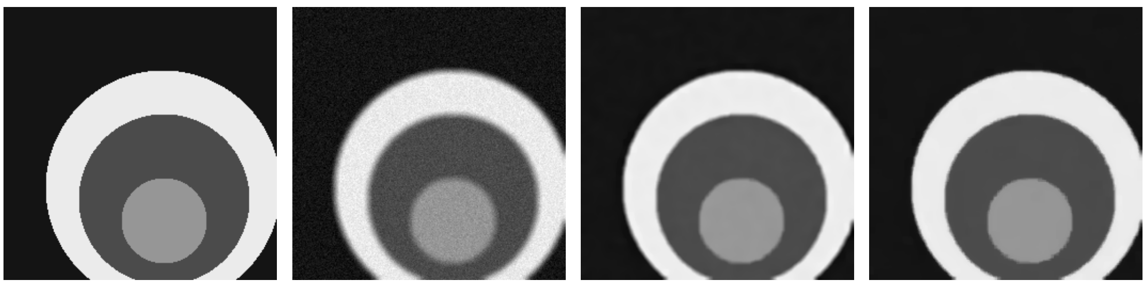
Table 2.3: Averages of CPU time, number of iterations, and percentage of error in the objective value for Algorithm 1 with $\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 = 1$ and Condat with tolerance 10^{-8}.

			$\varepsilon = 10^{-8}$		
Algorithm	τ	σ_1	Av. Time(s)	Av. Iter.	Av. % error o.v.
Alg. 1	0.77	0.16	93.36	19560	0.3514
	1.17	0.11	83.15	17561	0.3515
	1.77	0.07	100.06	20796	0.3515
	2.69	0.05	128.80	26801	0.3516
	4.09	0.03	160.92	33709	0.3517
Condat	1.17	-	93.77	18451	0.3515



(a) (b)

Figure 2.3: Comparison of Algorithm 1 with $\tau\sigma_1\|\nabla\|^2 + \tau\sigma_2 = 1$ and Condat (observation b_{13}).



(a) Original, $F^{TV}(\overline{x}) = 10.32$

(b) Blurry/noisy b_{13}, $F^{TV}(b) = 53.13$, PSNR=22.15

(c) AFBS, $F^{TV}(x_{100}) = 11.28$, PSNR=25.53.

(d) MS, $F^{TV}(x_{100}) = 10.56$, PSNR=26.38.

(e) Condat-Vũ, $F^{TV}(x_{100}) = 10.94$, PSNR=28.38.

(f) Condat ($k = 21$), $F^{TV}(x_{100}) = 10.57$, PSNR=28.80.

(g) Alg. 1 ($k = 21$, $\ell = 0.001$), $F^{TV}(x_{100}) = 10.55$, PSNR=28.80.

Figure 2.4: Reconstructed image, after 100 iterations, from blurred and noisy image using AFBS, MS, Condat-Vũ, Condat and Alg. 1.

performances when solving the following optimization problem

$$\min_{x\in\mathbb{R}^N} F(x) = h(x-z) + \alpha\|Mx\|_1, \tag{2.2.72}$$

where $h : \mathbb{R}^N \to \mathbb{R}$ is defined by

$$h\colon x = (\xi_i)_{1\le i\le n} \mapsto \sum_{i=1}^{N} \phi(\xi_i), \quad \phi : \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} |\xi| - \dfrac{\delta}{2}, & \text{if } |\xi| > \delta; \\ \dfrac{\xi^2}{2\delta}, & \text{if } |\xi| \le \delta, \end{cases} \tag{2.2.73}$$

$\delta > 0$, $z \in \mathbb{R}^N$, $\alpha > 0$, and M is a $N \times N$ symmetric positive definite real matrix. The first term in (2.2.72) is a data fidelity penalization using the Huber distance and the second term imposes sparsity in the solution. This type of problems appears naturally in image and signal denoising (see, e.g., [22, 43, 47, 51]).

Since M is symmetric, there exist $N \times N$ real matrices P and D, such that $P^\top = P^{-1}$, D is diagonal, and $M = PDP^\top$. By setting $g = h(\cdot - z)$, $f = \alpha\|\cdot\|_1$, $K = PD^{1-\eta}P^\top$, and $T = PD^\eta P^\top$, for some $\eta \in [0,1]$, we deduce that $KT = M$ and (2.2.72) is a particular instance of (P). Next, we illustrate the efficiency of Algorithm 2.2.14 for different values of $\eta \in [0,1]$. Observe that, in the case when $\eta = 0$ we have $T = \mathrm{Id}$ and Algorithm 2.2.14 reduces to the algorithm in Corollary 2.2.19. On the other hand, in the case when $\eta = 1$ we have $K = \mathrm{Id}$ and Algorithm 2.2.14 reduces to ADMM in (2.2.65). We have $\mathrm{prox}_f : (\xi_i)_{1\le i\le n} \mapsto \mathrm{prox}_{|\cdot|}(\xi_i)$, where $\mathrm{prox}_{|\cdot|}$ is the scalar soft-thresholder operator [3, Example 24.34(iii)]. Note that, since $\ker T = \{0\}$, for every $\eta \in [0,1]$, the optimization problem in the second step of (2.2.47) admits a unique solution, in view of Remark 2.2.18(3). Therefore, when $\Upsilon = \tau\mathrm{Id}$ and $\Sigma = \sigma\mathrm{Id}$, Algorithm 2.2.14 in this example reads as follows.

Algorithm 2

1: Fix $\tau > 0$, $p_0, q_0, x_0 \in \mathbb{R}^N$, $\varepsilon > 0$, and $r_0 > \varepsilon$.
2: **while** $r_n > \varepsilon$ **do**
3: $\quad y_n = x_n + \tau(KTp_n - q_n)$
4: $\quad p_{n+1} = \mathrm{zer}\big(\sigma\nabla h(\cdot - z) + T^*\big(T\cdot - (Tp_n - \sigma K^* y_n)\big)\big)$
5: $\quad q_{n+1} = \mathrm{prox}_{f/\tau}(x_n/\tau + KTp_{n+1})$
6: $\quad x_{n+1} = x_n + \tau(KTp_{n+1} - q_{n+1})$
7: $\quad u_{n+1} = \sigma K^*(x_{n+1} - x_n) - Tp_{n+1}$
8: $\quad r_{n+1} = \mathcal{R}(x_{n+1}, u_{n+1}, x_n, u_n)$
9: **end while**
10: **return** $(p_{n+1}, q_{n+1}, x_{n+1})$

Note that the step 4 in Algorithm 2 involves the resolution of a non-linear equation when $\eta > 0$. On the other hand, in the case when $\eta = 0$, we have $T = \mathrm{Id}$ and, as noticed in

Remark 2.2.20(3), the step 4 can be computed explicitly by using $\mathrm{prox}_g = z + \mathrm{prox}_h(\cdot - z)$ [3, Proposition 23.17(iii)] and the fact that δh is the real *Huber function* (see [3, Example 8.44 & Example 24.9]). We consider as stopping criterion the primal-dual relative error defined in (2.2.70).

We compare the performance of Algorithm 2 when $\eta \in \{0, 0.8, 0.9, 1\}$ with the standard solver *fmincon* of MATLAB for $N \in \{100, 250, 500\}$ and different values of the minimum and maximum eigenvalues $\lambda_{\max} \geq \lambda_{\min} > 0$ of the matrix M. Since the expected value of $\lambda_{\max}$ (resp. $\lambda_{\min}$) of random matrices generated by a normal distribution increases (resp. decreases) as N increases (see [33, Table 1.2]), we consider three classes of matrices with condition number $\kappa = \lambda_{\max}/\lambda_{\min} = 50$ for each dimension $N \in \{100, 250, 500\}$:

- **Class A:** Class of matrices M with small eigenvalues ($\lambda_{\max} = N/1000$).
- **Class B:** Class of matrices M with average eigenvalues ($\lambda_{\max} = 4N$).
- **Class C:** Class of matrices M with large eigenvalues ($\lambda_{\max} = 100N$).

For each class, we generate 30 random matrices using the *randn* function of MATLAB and the eigenvalues of each randomly generated matrix M is forced to satisfy the conditions of each class after a singular value decomposition $M = PDP^{\top}$. We next generate T and K as described before. Step 4 in Algorithm 2 is computed via *fsolve* function of MATLAB (for $\eta > 0$). We define the *percentage of improvement* of an algorithm with respect to *fmincon* via $I_{\bar{n}} = (\overline{F} - F(p_{\bar{n}})) \cdot 100/\overline{F}$, where $\overline{F}$ stands for the value of the function obtained by *fmincon* with tolerance 10^{-14} and $F(p_{\bar{n}})$ is the value of the function obtained by Algorithm 2 when it stops in iteration $\bar{n}$. Finally, we set the tolerance $\varepsilon = 10^{-6}$ and $\tau = 1$ in Algorithm 2.

Table 2.4 provides the averages of CPU time, iterations, and percentage of improvement with respect to *fmincon* of Algorithm 2 in the cases $\eta \in \{0, 0.8, 0.9, 1\}$ for the 30 random matrices in each class and $N \in \{100, 250, 500\}$. We split our analysis of the results in the three classes of random matrices.

The best performance in the class A (small eigenvalues) is obtained by the case when $\eta = 0$ (Corollary 2.2.19) in each dimension. The function value is very close to the one obtained by *fmincon* (difference of $10^{-5}\%$). For this class, the cases when $\eta \in \{0.8, 0.9\}$ are less accurate and ADMM ($\eta = 1$) is even more precise but much slower than the case when $\eta = 0$ for this class. This is explained by a very low cost per iteration and a comparable average number of iterations of the case when $\eta = 0$.

On the other hand, for matrices belonging to the class B (average eigenvalues), the most efficient method is SADMM with $\eta = 0.9$. The method needs very few number of iterations on average and it is more accurate than *fmincon*, since $I_{\bar{n}}$ is positive. This feature is also verified in $\eta \in \{0.8, 1\}$ but the number of iterations and computational time is larger. We observe that the case when $\eta = 0$ shows a very large number of iterations for achieving convergence and looses precision as the dimension increases. We conclude that SADMM outperforms drastically ADMM and the algorithm of Corollary 2.2.19, for suitable factorizations of matrices M with average eigenvalues.

Finally, ADMM ($\eta = 1$) is the best algorithm for the class C. It needs a very few number of iterations on average for achieving convergence, which nicely scales with the dimension. The computational time is around 1/3 of the closest competitor and the precision is as good as *fmincon*. SADMM algorithms when $\eta \in \{0.8, 0.9\}$ are similarly accurate but much slower. The case when $\eta = 0$ is very far from the solution and extremely slow for this class in all dimensions.

Table 2.4: Performance of Algorithm 2 for $N \in \{100, 250, 500\}$, $\eta \in \{0, 0.8, 0.9, 1\}$ and classes A, B, and C.

N	Class	η	0	0.8	0.9	1
100	A	Av. time	0.019	4.86	4.92	4.37
		Av. iter	688	704	717	656
		Av. $I_{\bar{n}}$ (%)	-$1.8 \cdot 10^{-5}$	-0.47	-0.07	-$1.5 \cdot 10^{-6}$
	B	Av. time	17.52	1.15	0.50	5.41
		Av. iter	798258	118	49	519
		Av. $I_{\bar{n}}$ (%)	0.63	0.36	0.33	0.64
	C	Av. time	31.44	3.77	1.07	0.34
		Av. iter	1410638	395	107	30
		Av. $I_{\bar{n}}$ (%)	-1607	-$8.4 \cdot 10^{-8}$	-$8.1 \cdot 10^{-8}$	-$5.1 \cdot 10^{-8}$
250	A	Av. time	0.036	8.94	9.25	8.88
		Av. iter	380	359	387	393
		Av. $I_{\bar{n}}$ (%)	-$1.6 \cdot 10^{-5}$	-1.03	-0.18	-$8 \cdot 10^{-6}$
	B	Av. time	136.82	5.54	2.61	32.15
		Av. iter	1547593	143	64	886
		Av. $I_{\bar{n}}$ (%)	-0.15	0.18	0.19	0.25
	C	Av. time	85.28	27.14	5.83	1.76
		Av. iter	971230	761	120	39
		Av. $I_{\bar{n}}$ (%)	-18287	-$1.3 \cdot 10^{-7}$	-$9.5 \cdot 10^{-8}$	-$3.3 \cdot 10^{-8}$
500	A	Av. time	0.067	13.41	13.58	13.52
		Av. iter	123	128	129	132
		Av. $I_{\bar{n}}$ (%)	$7.2 \cdot 10^{-5}$	-1.47	-0.30	$8.2 \cdot 10^{-5}$
	B	Av. time	581.25	39.99	23.95	113.24
		Av. iter	1249041	248	162	740
		Av. $I_{\bar{n}}$ (%)	-2.32	0.13	0.13	0.15
	C	Av. time	205.34	193.95	32.09	12.31
		Av. iter	419896	1200	182	46
		Av. $I_{\bar{n}}$(%)	-261808	-$1.8 \cdot 10^{-7}$	-$1.5 \cdot 10^{-7}$	-$9.4 \cdot 10^{-8}$

Acknowledgments

The first author thanks the support of ANID under grant FONDECYT 1190871 and grant Redes 180032. The second author thanks the support of ANID-Subdirección de Capital Humano/Doctorado Nacional/2018-21181024 and of the Dirección de Postgrado y Programas from UTFSM through Programa de Incentivos a la Iniciación Científica (PIIC).

Bibliography

[1] T. Aspelmeier, C. Charitha, and D. R. Luke, *Local linear convergence of the ADMM/Douglas-Rachford algorithms without strong convexity and application to statistical imaging*, SIAM J. Imaging Sci., 9 (2016), pp. 842–868, https://doi.org/10.1137/15M103580X.

[2] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, vol. 2 of Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 1990.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5.

[4] R. I. Boţ and E. R. Csetnek, *ADMM for monotone operators: convergence analysis and rates*, Adv. Comput. Math., 45 (2019), pp. 327–359, https://doi.org/10.1007/s10444-018-9619-3.

[5] R. I. Boţ, E. R. Csetnek, and A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM J. Optim., 23 (2013), pp. 2011–2036, https://doi.org/10.1137/12088255X.

[6] R. I. Boţ and C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565, https://doi.org/10.1137/120901106.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3 (2011), pp. 1–122, https://doi.org/10.1561/2200000016.

[8] K. Bredies and H. Sun, *Preconditioned Douglas-Rachford splitting methods for convex-concave saddle-point problems*, SIAM J. Numer. Anal., 53 (2015), pp. 421–444, https://doi.org/10.1137/140965028.

[9] K. Bredies and H. Sun, *A proximal point analysis of the preconditioned alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 878–907, https://doi.org/10.1007/s10957-017-1112-5.

[10] K. Bredies and H. P. Sun, *Preconditioned Douglas-Rachford algorithms for TV- and TGV-regularized variational imaging problems*, J. Math. Imaging Vision, 52 (2015), pp. 317–344, https://doi.org/10.1007/s10851-015-0564-1.

[11] L. Briceño, R. Cominetti, C. E. Cortés, and F. Martínez, *An integrated behavioral model of land use and transport system: a hyper-network equilibrium approach*, Netw. Spat. Econ., 8 (2008), pp. 201–224, https://doi.org/10.1007/s11067-007-9052-5.

[12] L. M. Briceño-Arias and P. L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250, https://doi.org/10.1137/10081602X.

[13] L. M. Briceño-Arias and P. L. Combettes, *Monotone operator methods for Nash equilibria in non-potential games*, in Computational and analytical mathematics, vol. 50 of Springer Proc. Math. Stat., Springer, New York, 2013, pp. 143–159, https://doi.org/10.1007/978-1-4614-7621-4_9.

[14] L. M. Briceño-Arias and D. Davis, *Forward-backward-half forward algorithm for solving monotone inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871, https://doi.org/10.1137/17M1120099.

[15] L. M. Briceño-Arias and F. Roldán, *Primal-dual splittings as fixed point iterations in the range of linear operators*, 2019, https://arxiv.org/abs/1910.02329.

[16] L. M. Briceño-Arias and F. Roldán, *Split-Douglas-Rachford algorithm for composite monotone inclusions and split-ADMM*, SIAM J. Optim., 31 (2021), pp. 2987–3013, https://doi.org/10.1137/21M1395144.

[17] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97, https://doi.org/10.1023/B:JMIV.0000011320.81911.38.

[18] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, vol. 9 of Radon Ser. Comput. Appl. Math., Walter de Gruyter, Berlin, 2010, pp. 263–340, https://doi.org/10.1515/9783110226157.263.

[19] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188, https://doi.org/10.1007/s002110050258.

[20] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145, https://doi.org/10.1007/s10851-010-0251-1.

[21] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*, Math. Programming, 64 (1994), pp. 81–101, https://doi.org/10.1007/BF01582566.

[22] J. Colas, N. Pustelnik, C. Oliver, P. Abry, J.-C. Géminard, and V. Vidal, *Nonlinear denoising for characterization of solid friction under low confinement pressure*, Physical Review E , 42 (2019), p. 91, https://doi.org/10.1103/PhysRevE.100.032803.

[23] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (Haifa, 2000), vol. 8 of Stud. Comput. Math., North-Holland, Amsterdam, 2001, pp. 115–152, https://doi.org/10.1016/S1570-579X(01)80010-0.

[24] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318, https://doi.org/10.1080/02331934.2012.733883.

[25] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479, https://doi.org/10.1007/s10957-012-0245-9.

[26] D. Dũng and B. C. Vũ, *A splitting algorithm for system of composite monotone inclusions*, Vietnam J. Math., 43 (2015), pp. 323–341, https://doi.org/10.1007/s10013-015-0121-7.

[27] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, https://doi.org/10.1002/cpa.20042.

[28] J. Douglas, Jr. and H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439, https://doi.org/10.2307/1993056.

[29] J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, PhD thesis, Massachusetts Institute of Technology, 1989.

[30] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–318, https://doi.org/10.1007/BF01581204.

[31] J. Eckstein and B. F. Svaiter, *A family of projective splitting methods for the sum of two maximal monotone operators*, Math. Program., 111 (2008), pp. 173–199, https://doi.org/10.1007/s10107-006-0070-8.

[32] J. Eckstein and W. Yao, *Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives*, Pac. J. Optim., 11 (2015), pp. 619–644.

[33] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 543–560, https://doi.org/10.1137/0609045.

[34] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Programming, 72 (1996), pp. 1–15, https://doi.org/10.1016/0025-5610(95)00012-7.

[35] D. Gabay, *Chapter IX applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, New York, 1983, pp. 299 – 331, https://doi.org/10.1016/S0168-2024(08)70034-1.

[36] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, 2 (1976), pp. 17–40, https://doi.org/10.1016/0898-1221(76)90003-1.

[37] E. M. Gafni and D. P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM J. Control Optim., 22 (1984), pp. 936–964, https://doi.org/10.1137/0322061.

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

[39] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Math. Soc., 70 (1964), pp. 709–710, https://doi.org/10.1090/S0002-9904-1964-11178-2.

[40] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring images: Matrices, spectra, and filtering*, vol. 3 of Fundamentals of Algorithms, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006, https://doi.org/10.1137/1.9780898718874.

[41] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149, https://doi.org/10.1137/100814494.

[42] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.

[43] X. Liu, D. Zhai, D. Zhao, G. Zhai, and W. Gao, *Progressive image denoising through hybrid graph Laplacian regularization: a unified framework*, IEEE Trans. Image Process., 23 (2014), pp. 1491–1503, https://doi.org/10.1109/TIP.2014.2303638.

[44] R. V. Mises and H. Pollaczek-Geiringer, *Praktische verfahren der gleichungsauflösung* ., ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik, 9 (1929), pp. 152–164, https://doi.org/10.1002/zamm.19290090206.

[45] C. Molinari, J. Peypouquet, and F. Roldán, *Alternating forward-backward splitting for linearly constrained optimization problems*, Optim. Lett., 14 (2020), pp. 1071–1088, https://doi.org/10.1007/s11590-019-01388-y.

[46] W. M. Moursi and Y. Zinchenko, *A note on the equivalence of operator splitting methods*, in Splitting Algorithms, Modern Operator Theory, and Applications, Springer, Cham, 2019, pp. 331–349.

[47] J. Pang and G. Cheung, *Graph Laplacian regularization for image denoising: analysis in the continuous domain*, IEEE Trans. Image Process., 26 (2017), pp. 1770–1785, https://doi.org/10.1109/TIP.2017.2651400.

[48] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in 2011 International Conference on Computer Vision, 2011, pp. 1762–1769, https://doi.org/10.1109/ICCV.2011.6126441.

[49] F. Riesz and B. Sz.-Nagy, *Functional analysis*, Frederick Ungar Publishing Co., New York, 1955. Translated by Leo F. Boron.

[50] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268, https://doi.org/10.1016/0167-2789(92)90242-F.

[51] L. Sha, D. Schonfeld, and J. Wang, *Graph laplacian regularization with sparse coding for image restoration and representation*, IEEE Transactions on Circuits and Systems for Video Technology, 30 (2020), pp. 2000–2014, https://doi.org/10.1109/TCSVT.2019.2913411.

[52] R. Shefi and M. Teboulle, *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM J. Optim., 24 (2014), pp. 269–297, https://doi.org/10.1137/130910774.

[53] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1997, https://doi.org/10.1090/surv/049.

[54] B. F. Svaiter, *On weak convergence of the Douglas-Rachford method*, SIAM J. Control Optim., 49 (2011), pp. 280–287, https://doi.org/10.1137/100788100.

[55] A. Themelis and P. Patrinos, *Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results*, SIAM J. Optim., 30 (2020), pp. 149–181, https://doi.org/10.1137/18M1163993.

[56] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

[57] M. Yan and W. Yin, *Self equivalence of the alternating direction method of multipliers*, in Splitting Methods in Communication, Imaging, Science, and Engineering, Sci. Comput., Springer, Cham, 2016, pp. 165–194.

[58] Y. Yang, Y. Tang, M. Wen, and T. Zeng, *Preconditioned douglas-rachford type primal-dual method for solving composite monotone inclusion problems with applications*, Inverse Problems & Imaging, 15 (2021), pp. 787–825.

[59] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46, https://doi.org/10.1007/s10915-010-9408-8.

Chapter 3

Primal-Dual Algorithm with Critical Step-Sizes and Relaxation Parameters

3.1 Introduction and Main Results

In this section we provide a theoretical study of the relaxed primal-dual splitting [42] including critical preconditioners for solving the following composite monotone inclusion. We present an alternative formulation of the primal-dual algorithm that differs from the convergence analysis of SDR algorithm made in Theorem 2.2.6, in which, the inclusion of relaxation parameters is not clear.

Problem 3.1.1. *Let $\mathcal{H}$ and $\mathcal{G}$ be a real Hilbert spaces, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone operators, and let $L : \mathcal{H} \to \mathcal{G}$ be a linear bounded operator. The problem is to find $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$, where*

$$\boldsymbol{Z} = \big\{(x, u) \in \mathcal{H} \times \mathcal{G} \;\big|\; 0 \in Ax + L^*u,\ 0 \in B^{-1}u - Lx\big\} \tag{3.1.1}$$

is assumed to be non-empty.

Ir order to include the critical preconditioners and relaxation parameters on the primal-dual algorithm, we study Krasnosel'skiĭ-Mann (KM) iterations defined in the range of monotone self-adjoint linear operators in the following result.

Proposition 3.1.2. *Let $(\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot \rangle)$ be a real Hilbert space, let $\boldsymbol{V} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be a monotone self-adjoint linear bounded operator such that* $\operatorname{ran} \boldsymbol{V}$ *is closed, and let $\boldsymbol{S} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be such that* $\operatorname{Fix} \boldsymbol{S} \neq \varnothing$, *that $\boldsymbol{S} = \boldsymbol{S} \circ P_{\operatorname{ran} \boldsymbol{V}}$, and that $(P_{\operatorname{ran} \boldsymbol{V}} \circ \boldsymbol{S})|_{\operatorname{ran} \boldsymbol{V}}$ is quasinonexpansive in* $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$. *Define*

$$\boldsymbol{T} : \operatorname{ran} \boldsymbol{V} \to \operatorname{ran} \boldsymbol{V} : x \mapsto P_{\operatorname{ran} \boldsymbol{V}} \circ \boldsymbol{S}x \tag{3.1.2}$$

and consider the sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *defined by the recurrence*

$$\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}, \quad (\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = (1-\lambda_n)\,\boldsymbol{x}_n + \lambda_n \boldsymbol{S}\boldsymbol{x}_n. \tag{3.1.3}$$

Moreover, suppose that one of the following holds:

(i) $\boldsymbol{T}$ *is firmly quasinonexpansive,* $\mathbf{Id} - \boldsymbol{T}$ *is demiclosed at* $\mathbf{0}$ *in* $(\operatorname{ran} \boldsymbol{V}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{V}})$*, and* $(\lambda_n)_{n\in\mathbb{N}}$ *is a sequence in* $[\varepsilon, 2-\varepsilon]$ *for some* $\varepsilon \in \,]0,1[$*.*

(ii) $\boldsymbol{T}$ *is* $\alpha-$*averaged nonexpansive in* $(\operatorname{ran} \boldsymbol{V}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{V}})$ *for some* $\alpha \in \,]0,1[$ *and* $(\lambda_n)_{n\in\mathbb{N}}$ *is a sequence in* $[0, 1/\alpha]$ *such that* $\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty$*.*

Then the following hold:

1. $(P_{\operatorname{ran} \boldsymbol{V}} \boldsymbol{x}_n)_{n\in\mathbb{N}}$ *is Féjer monotone in* $(\operatorname{ran} \boldsymbol{V}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{V}})$ *with respect to* $\operatorname{Fix} \boldsymbol{T}$*.*

2. $(P_{\operatorname{ran} \boldsymbol{V}}(\boldsymbol{S}\boldsymbol{x}_n - \boldsymbol{x}_n))_{n\in\mathbb{N}}$ *converges strongly to* $\mathbf{0}$ *in* $(\operatorname{ran} \boldsymbol{V}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{V}})$*.*

3. $(P_{\operatorname{ran} \boldsymbol{V}} \boldsymbol{x}_n)_{n\in\mathbb{N}}$ *converges weakly in* $(\operatorname{ran} \boldsymbol{V}, \langle\cdot \mid \cdot\rangle_{\boldsymbol{V}})$ *to some* $\hat{\boldsymbol{x}} \in \operatorname{Fix} \boldsymbol{T}$ *and* $\boldsymbol{S}\hat{\boldsymbol{x}} \in \operatorname{Fix} \boldsymbol{S}$*.*

To obtain the convergence of the relaxed primal-dual algorithm with critical preconditioners we prove that this method defines KM iterations in the range of the operator $\boldsymbol{V}$ given by

$$\boldsymbol{V} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x,u) \mapsto \left(\Upsilon^{-1}x - L^*u, \Sigma^{-1}u - Lx\right), \tag{3.1.4}$$

where $\Sigma : \mathcal{G} \to \mathcal{G}$ and $\Upsilon : \mathcal{H} \to \mathcal{H}$ are strongly monotone self-adjoint linear operators such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| \leq 1$. This operator is monotone self-adjoint and linear with non-trivial kernel. In addition, to ensure convergence, we prove that $\boldsymbol{V}$ is cocoercive and we assume that $\operatorname{ran} \boldsymbol{V}$ is closed, which is equivalent to the closure of $\operatorname{ran}(\Sigma^{-1} - L\Upsilon L^*)$. The following is the main result of this section.

Theorem 3.1.3. *In the context of Problem 3.1.1, let* $\boldsymbol{V}$ *be the operator defined in (3.1.4), where* $\Sigma : \mathcal{G} \to \mathcal{G}$ *and* $\Upsilon : \mathcal{H} \to \mathcal{H}$ *are self-adjoint linear strongly monotone operators such that* $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| \leq 1$*, and suppose that* $\operatorname{ran} \boldsymbol{V}$ *is closed. Moreover, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $[0,2]$ *satisfying* $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$*, and consider the sequence* $\big((x_n,u_n)\big)_{n\in\mathbb{N}}$ *defined by the recurrence*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon L^* u_n) \\ q_{n+1} = J_{\Sigma B^{-1}}\left(u_n + \Sigma L(2p_{n+1} - x_n)\right) \\ (x_{n+1}, u_{n+1}) = (1-\lambda_n)(x_n,u_n) + \lambda_n(p_{n+1}, q_{n+1}), \end{array} \right. \tag{3.1.5}$$

where $(x_0,u_0) \in \mathcal{H} \times \mathcal{G}$*. Then* $\big(P_{\operatorname{ran} \boldsymbol{V}}(x_n,u_n)\big)_{n\in\mathbb{N}}$ *converges weakly in* $(\operatorname{ran} \boldsymbol{V}, \langle\!\langle\cdot \mid \cdot\rangle\!\rangle_{\boldsymbol{V}})$ *to some* $(\hat{y},\hat{v}) \in \mathcal{H} \times \mathcal{G}$ *such that*

$$\left(J_{\Upsilon A}(\hat{y} - \Upsilon L^*\hat{v}), J_{\Sigma B^{-1}}\left(\hat{v} + \Sigma L(2J_{\Upsilon A}(\hat{y} - \Upsilon L^*\hat{v}) - \hat{y})\right)\right) \tag{3.1.6}$$

is a solution to Problem 3.1.1.

The previous theorem generalizes results in [19, 21, 42] including critical preconditioners, variable metrics, and relaxation parameters. We also recover the convergence of the Douglas-Rachford splitting algorithm.

Proposition 3.1.4. *In the context of Problem 3.1.1, set $L = \mathrm{Id}$, let Υ be a strongly monotone self-adjoint linear operator, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ satisfying $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, and consider the sequence $\big((x_n,u_n)\big)_{n\in\mathbb{N}}$ defined by the recurrence*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon u_n) \\ q_{n+1} = J_{\Upsilon^{-1}B^{-1}}\big(u_n + \Upsilon^{-1}(2p_{n+1} - x_n)\big) \\ (x_{n+1},u_{n+1}) = (1-\lambda_n)(x_n,u_n) + \lambda_n(p_{n+1},q_{n+1}), \end{array} \right. \tag{3.1.7}$$

where $(x_0,u_0) \in \mathcal{H}\times\mathcal{H}$. Then, by setting, for every $n \in \mathbb{N}$, $z_n = x_n - \Upsilon y_n$, $(z_n)_{n\in\mathbb{N}}$ converges weakly in $\mathcal{H}$ to some $\hat{z} \in \mathrm{Fix}\,(J_{\Upsilon B} \circ (2J_{\Upsilon A} - \mathrm{Id}) + (\mathrm{Id} - J_{\Upsilon A}))$ and

$$\big(J_{\Upsilon A}\hat{z}, -\Upsilon^{-1}(\hat{z} - J_{\Upsilon A}\hat{z})\big)$$

is a solution to Problem 3.2.1. Moreover, we have

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = (1-\lambda_n)z_n + \lambda_n J_{\Upsilon B} \circ (2J_{\Upsilon A} - \mathrm{Id})z_n + (\mathrm{Id} - J_{\Upsilon A})z_n. \tag{3.1.8}$$

We finalize the article with numerical experiments in total variation image restoration. From these experiments, we can observe that including relaxation parameters and critical preconditioners the numerical convergence of the method is accelerated.

3.2 Article: Primal-dual Splittings as Fixed Point Iterations in the Range of Linear Operators[1]

Abstract In this paper we study the relaxed primal-dual algorithm for solving composite monotone inclusions in real Hilbert spaces with critical preconditioners. Our approach is based in new results on the asymptotic behaviour of Krasnosel'skiĭ-Mann (KM) iterations defined in the range of monotone self-adjoint linear operators. These results generalize the convergence of classical KM iterations aiming at approximating fixed points. We prove that the relaxed primal-dual algorithm with critical preconditioners define KM iterations in the range of a particular monotone self-adjoint linear operator with non-trivial kernel. We then deduce from our fixed point approach that the shadows of primal-dual iterates on the range of the linear operator converges weakly to some point in this vector subspace from which we obtain a solution. This generalizes [21, Theorem 3.3] to infinite dimensional relaxed primal-dual monotone inclusions involving critical preconditioners. The

[1] [8] Luis M. Briceño-Arias and Fernando Roldán. Primal-dual splitting as fixed point iterations in the range of linear operators, 2019, https://arxiv.org/abs/1910.02329

Douglas-Rachford splitting (DRS) is interpreted as a particular instance of the primal-dual algorithm when the step-sizes are critical and we recover classical results from this new perspective. We implement the relaxed primal-dual algorithm with critical preconditioners in total variation reconstruction and we illustrate its flexibility and efficiency.

3.2.1 Introduction

In this paper we provide a theoretical study of the relaxed primal-dual splitting [42] for solving the following composite monotone inclusion.

Problem 3.2.1. *Let $\mathcal{H}$ and $\mathcal{G}$ be a real Hilbert spaces, let $A: \mathcal{H} \to 2^{\mathcal{H}}$ and $B: \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone operators, and let $L: \mathcal{H} \to \mathcal{G}$ be a linear bounded operator. The problem is to find $(\hat{x}, \hat{u}) \in \boldsymbol{Z}$, where*

$$\boldsymbol{Z} = \left\{(x,u) \in \mathcal{H} \times \mathcal{G} \;\middle|\; 0 \in Ax + L^*u,\ 0 \in B^{-1}u - Lx\right\} \tag{3.2.1}$$

is assumed to be non-empty.

It follows from [9, Proposition 2.8] that any solution $(\hat{x}, \hat{u})$ to Problem 3.2.1 satisfies that $\hat{x}$ is a solution to the primal inclusion

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + L^*BLx \tag{3.2.2}$$

and $\hat{u}$ is solution to the dual inclusion

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad 0 \in B^{-1}u - LA^{-1}(-L^*u). \tag{3.2.3}$$

Conversely, if $\hat{x}$ is a solution to (3.2.2) then there exists $\tilde{u}$ solution to (3.2.3) such that $(\hat{x}, \tilde{u}) \in \boldsymbol{Z}$ and the dual argument also holds. In the particular case when $A = \partial f$ and $B = \partial g$ for lower semicontinuous convex proper functions $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ and $g\colon \mathcal{G} \to \,]-\infty, +\infty]$, we have that $\boldsymbol{Z} \subset \mathcal{P} \times \mathcal{D}$, where $\mathcal{P}$ is the set of solutions to the convex optimization problem

$$\min_{x \in \mathcal{H}} f(x) + g(Lx) \tag{3.2.4}$$

and $\mathcal{D}$ is the set of solutions to its Fenchel-Rockafellar dual

$$\min_{u \in \mathcal{G}} g^*(u) + f^*(-L^*u). \tag{3.2.5}$$

Problem 3.2.1 and its particular optimization case model a wide class of problems in engineering going from from mechanical problems [26, 28, 29], differential inclusions [1, 40], game theory [10], image processing problems as image restoration and denoising [14, 16, 22], traffic theory [7, 25, 27], among other disciplines.

In the last years, several algorithms have been proposed in order to solve Problem 3.2.1 and some generalizations involving cocoercive operators [9, 42, 31]. One of the most used, is the algorithm proposed in [20] (see also [42, 5, 6] and [37, 15] in the context of (3.2.4)), which iterates

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon L^* u_n) \\ q_{n+1} = J_{\Sigma B^{-1}}\big(u_n + \Sigma L(2p_{n+1} - x_n)\big) \\ (x_{n+1}, y_{n+1}) = (1-\lambda_n)(x_n, y_n) + \lambda_n(p_{n+1}, q_{n+1}), \end{array} \right. \tag{3.2.6}$$

where $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$, $(\lambda_n)_{n\in\mathbb{N}} \subset \left]0,2\right[$, and preconditioners $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ and $\Sigma\colon \mathcal{G} \to \mathcal{G}$ are strongly monotone self-adjoint linear bounded operators such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| < 1$. It turns out that (3.2.6) corresponds to the relaxed proximal-point algorithm [38, 34] associated to the operator $\boldsymbol{V}^{-1}\boldsymbol{M}$, where $\boldsymbol{M}\colon (x,u) \mapsto (Ax + L^*u) \times (B^{-1}u - Lx)$ is maximally monotone in $\mathcal{H} \oplus \mathcal{G}$ [9, Proposition 2.7(iii)] and the self-adjoint linear bounded operator

$$\boldsymbol{V} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x,u) \mapsto \big(\Upsilon^{-1}x - L^*u, \Sigma^{-1}u - Lx\big) \tag{3.2.7}$$

is strongly monotone if $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| < 1$. Hence, $\boldsymbol{V}^{-1}\boldsymbol{M}$ is maximally monotone in the real Hilbert space $(\mathcal{H} \times \mathcal{G}, \langle \cdot \mid \boldsymbol{V}\cdot\rangle)$ and the convergence is a consequence of [38, 34]. Note that $J_{\boldsymbol{M}}$ is also firmly nonexpansive in $\mathcal{H} \oplus \mathcal{G}$, however it has no explicit computation. Non-standard metrics are widely used not only to obtain explicit resolvent computations but also to accelerate algorithms [11, 21, 31, 42, 23, 19]. In the presence of critical preconditioners, i.e., $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| = 1$, the non-standard metric approach fails since $\ker \boldsymbol{V} \neq \{0\}$ and, hence, $\langle \cdot \mid \boldsymbol{V}\cdot\rangle$ is not an inner product in $\mathcal{H} \times \mathcal{G}$. Hence, the convergence of (3.2.6) when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| \leq 1$ in the context of Problem 3.2.1 is an open question, which is partially answered in [21, Theorem 3.3] for solving (3.2.4) in a finite dimensional setting when $\Sigma = \sigma\mathrm{Id}$ and $\Upsilon = \tau\mathrm{Id}$. In the particular case when $\lambda_n \equiv 1$, the weak convergence of (3.2.6) when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| \leq 1$ is deduced in [12] from an alternative formulation of (3.2.6). This formulation in the case when $L = \mathrm{Id}$, generates primal-dual iterates in the graph of A and, therefore, the argument does not hold when relaxation steps are included in general.

In this paper we generalize [21, Theorem 3.3] to the monotone inclusion in Problem 3.2.1 in the infinite dimensional setting with critical preconditioners. Our approach is based on a fixed point theory restricted to $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \boldsymbol{V}\cdot\rangle)$, which is a real Hilbert space under the condition $\operatorname{ran} \boldsymbol{V}$ closed. We obtain the weak convergence of Krasnosel'skiĭ-Mann iterations governed by firmly quasinonexpansive and averaged operators in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \boldsymbol{V}\cdot\rangle)$, which generalizes [17, Theorem 5.2(i)] and [3, Proposition 5.16]. This result is interesting in its own right. It is worth to notice that most of known algorithms can be seen as fixed point iterations of operators belonging to the previous classes [5, 6, 9, 21, 31]. Our approach gives new insights on primal-dual algorithms: the convergence of primal-dual iterates in $\boldsymbol{\mathcal{H}}$ follows from the convergence of their *shadows* in $\operatorname{ran} \boldsymbol{V}$.

We also provide a detailed analysis of the case $L = \mathrm{Id}$ and relations of primal-dual algorithms with the relaxed Douglas-Rachford splitting (DRS) [24, 32]. We give a primal-dual version of DRS derived from (3.2.6) when $L = \mathrm{Id}$ and we recover the weak convergence of an auxiliary sequence whose *primal-dual shadow* is a solution to Problem 3.2.1, as in [24, 32].

We finish this paper by providing a numerical experiment on total variation image reconstruction, in which the advantages of using critical preconditioners and relaxation steps are illustrated.

The paper is organized as follows. In Section 3.2.2.1 we set our notation and some preliminaries. In Section 3.2.2.2 we study the fixed point problem on the range of linear operators and we provide conditions for the convergence of fixed point iterations governed by firmly quasinonexpansive or averaged nonexpansive operators. In Section 3.2.3 we apply fixed point results to the particular case of primal-dual monotone inclusions and we provide several connections with other results in the literature. In Section 3.2.4, we study in detail the particular case when $L = \mathrm{Id}$, which is connected with Douglas–Rachford splitting. Finally, in Section 3.2.5 we provide numerical experiments in image processing.

3.2.2 Preliminaries

In this section we first provide our notation and some preliminaries. Next we study a specific fixed point problem involving the range of a self-adjoint monotone linear operator and we provide the convergence of a Krasnosel'skiĭ–Mann iteration in this vector subspace, which is interesting in its own right. The weak convergence of (3.2.6) with critical preconditioners is derived from previous fixed point analysis.

3.2.2.1 Notation

Throughout this paper $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces. We denote the scalar product by $\langle \cdot \mid \cdot \rangle$ and the associated norm by $\|\cdot\|$. The symbols $\rightharpoonup$ and $\to$ denotes the weak and strong convergence, respectively. Given a linear bounded operator $L : \mathcal{H} \to \mathcal{G}$, we denote its adjoint by $L^* \colon \mathcal{G} \to \mathcal{H}$, its kernel by $\ker L$, and its range by $\operatorname{ran} L$. Id denotes the identity operator on $\mathcal{H}$. Let $D \subset \mathcal{H}$ be non-empty and let $T : D \to \mathcal{H}$. The set of fixed points of T is given by $\operatorname{Fix} T = \big\{x \in D \;\big|\; x = Tx\big\}$. Let $\beta \in \left]0, +\infty\right[$. The operator T is $\beta-$cocoercive if

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid Tx - Ty\rangle \geq \beta\|Tx - Ty\|^2, \tag{3.2.8}$$

it is $\beta-$strongly monotone if

$$(\forall x \in D)(\forall y \in D) \;\; \langle x - y \mid Tx - Ty\rangle \geq \beta\|x - y\|^2, \tag{3.2.9}$$

it is nonexpansive if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leq \|x - y\|, \tag{3.2.10}$$

it is quasinonexpansive if

$$(\forall x \in D)(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\| \leq \|x - y\|, \tag{3.2.11}$$

and it is firmly quasinonexpansive (or class $\mathfrak{T}$) if

$$(\forall x \in D)(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\|^2 \leq \|x - y\|^2 - \|Tx - x\|^2. \tag{3.2.12}$$

Let $\alpha \in]0,1[$. The operator T is $\alpha-$averaged nonexpansive if $T = (1-\alpha)\mathrm{Id} + \alpha R$ for some nonexpansive operator $R : \mathcal{H} \to \mathcal{H}$, and T is firmly nonexpansive if it is $\frac{1}{2}-$averaged nonexpansive.

Given a self-adjoint monotone linear bounded operator $V \colon \mathcal{H} \to \mathcal{H}$, we denote $\langle \cdot \mid \cdot \rangle_V = \langle \cdot \mid V \cdot \rangle$, which is bilinear, positive semi-definite, and symmetric. Moreover, there exists a self-adjoint monotone linear bounded operator $\sqrt{V} \colon \mathcal{H} \to \mathcal{H}$ such that

$$V = \sqrt{V}\sqrt{V}, \quad (\forall x \in \mathcal{H}) \quad \langle x \mid Vx \rangle = \|\sqrt{V}x\|^2, \tag{3.2.13}$$

and $\operatorname{ran} V = \operatorname{ran} \sqrt{V}$. In addition, if V is strongly monotone, $\langle \cdot \mid \cdot \rangle_V$ defines an inner product on $\mathcal{H}$ and we denote by $\| \cdot \|_V = \sqrt{\langle \cdot \mid \cdot \rangle_V}$ the induced norm.

Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain, range, and graph of A are $\operatorname{dom} A = \big\{ x \in \mathcal{H} \bigm| Ax \neq \varnothing \big\}$, $\operatorname{ran} A = \big\{ u \in \mathcal{H} \bigm| (\exists x \in \mathcal{H})\, u \in Ax \big\}$, and $\operatorname{gra} A = \big\{ (x,u) \in \mathcal{H} \times \mathcal{H} \bigm| u \in Ax \big\}$, respectively. The set of zeros of A is $\operatorname{zer} A = \big\{ x \in \mathcal{H} \bigm| 0 \in Ax \big\}$, the inverse of A is $A^{-1} \colon \mathcal{H} \to 2^{\mathcal{H}} \colon u \mapsto \big\{ x \in \mathcal{H} \bigm| u \in Ax \big\}$, and the resolvent of A is $J_A = (\mathrm{Id} + A)^{-1}$. We have $\operatorname{zer} A = \operatorname{Fix} J_A$. The operator A is monotone if

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v \rangle \geq 0 \tag{3.2.14}$$

and it is maximally monotone if it is monotone and there exists no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$, i.e., for every $(x,u) \in \mathcal{H} \times \mathcal{H}$,

$$(x,u) \in \operatorname{gra} A \quad \Leftrightarrow \quad (\forall (y,v) \in \operatorname{gra} A) \ \ \langle x - y \mid u - v \rangle \geq 0. \tag{3.2.15}$$

Let C be a non-empty subset of $\mathcal{H}$ and let $\big(x_n\big)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $\big(x_n\big)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C if

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leqslant \|x_n - x\|. \tag{3.2.16}$$

Let D be a non-empty weakly sequentially closed subset of $\mathcal{H}$, let $T : D \to \mathcal{H}$, and let $u \in \mathcal{H}$. Then T is demiclosed at u in $(\mathcal{H}, \langle \cdot \mid \cdot \rangle)$ if, for every sequence $(x_n)_{n\in\mathbb{N}}$ in D and every $x \in D$ such that $x_n \rightharpoonup x$ and $Tx_n \to u$ in $(\mathcal{H}, \langle \cdot \mid \cdot \rangle)$, we have have $Tx = u$. In addition, T is demiclosed if it is demiclosed at every point in D.

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f \colon \mathcal{H} \to]-\infty, +\infty]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is defined by $f^* \colon u \mapsto \sup_{x \in \mathcal{H}}(\langle x \mid u \rangle - f(x))$, which is a function in $\Gamma_0(\mathcal{H})$, the subdifferential of f is the maximally monotone operator

$$\partial f \colon x \mapsto \big\{ u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H}) \;\; f(x) + \langle y - x \mid u \rangle \leq f(y) \big\},$$

$(\partial f)^{-1} = \partial f^*$, and we have that $\operatorname{zer} \partial f$ is the set of minimizers of f, which is denoted by $\arg\min_{x \in \mathcal{H}} f$. Given a strongly monotone self-adjoint linear operator $\Upsilon \colon \mathcal{H} \to \mathcal{H}$, we denote by

$$\operatorname{prox}_f^{\Upsilon} \colon x \mapsto \underset{y \in \mathcal{H}}{\arg\min} \big(f(y) + \frac{1}{2} \|x - y\|_{\Upsilon}^2 \big), \tag{3.2.17}$$

and by $\operatorname{prox}_f = \operatorname{prox}_f^{\mathrm{Id}}$. We have $\operatorname{prox}_f^{\Upsilon} = J_{\Upsilon^{-1}\partial f}$ [3, Proposition 24.24(i)] and it is single valued since the objective function in (3.2.17) is strongly convex. Moreover, it follows from [3, Proposition 24.24] that

$$\operatorname{prox}_f^{\Upsilon} = \mathrm{Id} - \Upsilon^{-1} \operatorname{prox}_{f^*}^{\Upsilon^{-1}} \Upsilon = \Upsilon^{-1} \,(\mathrm{Id} - \operatorname{prox}_{f^*}^{\Upsilon^{-1}})\, \Upsilon. \tag{3.2.18}$$

Given a non-empty closed convex set $C \subset \mathcal{H}$, we denote by P_C the projection onto C and by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise. For further properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [3].

The following result allows us to define algorithms in a real Hilbert space defined by the range of non-invertible self-adjoint linear bounded operators. The result is a direct consequence of [3, Fact 2.26] and (3.2.13).

Proposition 3.2.2. *Let $V : \mathcal{H} \to \mathcal{H}$ be a monotone self-adjoint linear bounded operator. The following statements are equivalent.*

1. $\operatorname{ran} V$ *is closed.*

2. $(\exists \alpha > 0)(\forall x \in \operatorname{ran} V)$
$$\langle Vx \mid x \rangle \geq \alpha \|x\|^2. \tag{3.2.19}$$

Moreover, if 1 or 2 holds, then $(\operatorname{ran} V, \langle \cdot \mid \cdot \rangle_V)$ *is a real Hilbert space.*

The following example exhibits a monotone self-adjoint linear bounded operator whose range is not closed, illustrating that assumption $\operatorname{ran} V$ closed is not redundant in our setting.

Example 3.2.3. *Let $\ell^2(\mathbb{R})$ be the real Hilbert space defined by square summable sequences in $\mathbb{R}$ endowed by the inner product $\langle \cdot \mid \cdot \rangle\colon (x,y) \mapsto \sum_{j\geq 1} x_j\, y_j$ and consider the monotone self-adjoint bounded linear operator*

$$V : \ell^2(\mathbb{R}) \to \ell^2(\mathbb{R})\colon (x_n)_{n\in\mathbb{N}\setminus\{0\}} \mapsto \left(x_1, \frac{x_2}{2}, \frac{x_3}{3}, \ldots\right).$$

By considering the sequence $(x^n)_{n\in\mathbb{N}} \subset \ell^2(\mathbb{R})$ defined by $x^n_j = 1$ for $j \leq n$ and $x^n_j = 0$ for $j > n$... we have $Vx^n \to y = (1/j)_{j\in\mathbb{N}\setminus\{0\}} \in \ell^2(\mathbb{R})$ as $n \to +\infty$, and $y \notin \operatorname{ran} V$, which implies that $\operatorname{ran} V$ is not closed.

3.2.2.2 Fixed points in the range of linear operators

The following fixed point problem is the basis for the analysis of primal-dual algorithms.

Problem 3.2.4. *Let $(\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot \rangle)$ be a real Hilbert space, let $\boldsymbol{V} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be a monotone self-adjoint linear bounded operator such that $\operatorname{ran} \boldsymbol{V}$ is closed, and let $\boldsymbol{S}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be such that $\operatorname{Fix} \boldsymbol{S} \neq \varnothing$, that $\boldsymbol{S} = \boldsymbol{S} \circ P_{\operatorname{ran} \boldsymbol{V}}$, and that $(P_{\operatorname{ran} \boldsymbol{V}} \circ \boldsymbol{S})|_{\operatorname{ran} \boldsymbol{V}}$ is quasinonexpansive in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$. The problem is to*

$$\text{find} \quad \boldsymbol{x} \in \operatorname{Fix} \boldsymbol{S}. \tag{3.2.20}$$

First observe that under the hypotheses on $\boldsymbol{V}$, Proposition 3.2.2 asserts that $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$ is a real Hilbert space. In the particular case when $\boldsymbol{V} = \mathbf{Id}$, we have $\operatorname{ran} \boldsymbol{V} = \boldsymbol{\mathcal{H}}$, $P_{\operatorname{ran} \boldsymbol{V}} = \mathbf{Id}$, and Problem 3.2.4 is solved in [17] when $\boldsymbol{S}$ is firmly quasinonexpansive (or class $\mathfrak{T}$), and $\mathbf{Id} - \boldsymbol{S}$ is demiclosed at $\mathbf{0}$, and in [3, 18] when $\boldsymbol{S}$ is averaged nonexpansive. In the case when $\boldsymbol{V}$ is self-adjoint and strongly monotone, we also have $\operatorname{ran} \boldsymbol{V} = \boldsymbol{\mathcal{H}}$, $P_{\operatorname{ran} \boldsymbol{V}} = \mathbf{Id}$, and several approaches with non-standard metrics are developed for solving Problem 3.2.4 by using contractive assumptions on $\boldsymbol{S}$ (see, e.g., [11, 19, 21, 23, 31, 37, 42]). In all cases, the problem is solved via the Krasnosel'skiĭ–Mann iteration

$$\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}, \quad (\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = (1-\lambda_n)\boldsymbol{x}_n + \lambda_n \boldsymbol{S}\boldsymbol{x}_n, \tag{3.2.21}$$

where $(\lambda_n)_{n\in\mathbb{N}}$ is a strictly positive sequence. The main difference of Problem 3.2.4 with respect to previous literature is that the contractive property is only guaranteed for the *shadow operator* $(P_{\operatorname{ran} \boldsymbol{V}} \circ \boldsymbol{S})|_{\operatorname{ran} \boldsymbol{V}}$ on $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$ without any further assumption on $\boldsymbol{S}$. In this section we obtain conditions for ensuring the convergence of the *shadow sequence* $(P_{\operatorname{ran} \boldsymbol{V}}\boldsymbol{x}_n)_{n\in\mathbb{N}}$ to a solution to Problem 3.2.4 from (3.2.21). We first need the following technical lemma.

Lemma 3.2.5. *Let $\boldsymbol{Q}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ and let $\boldsymbol{S}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be such that $\operatorname{Fix} \boldsymbol{S} \neq \varnothing$ and $\boldsymbol{S} = \boldsymbol{S} \circ \boldsymbol{Q}$. Then $\boldsymbol{S}(\operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S})) = \operatorname{Fix} \boldsymbol{S}$ and, in particular, $\operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S}) \neq \varnothing$.*

Proof. First, let $\boldsymbol{x} \in \operatorname{Fix} \boldsymbol{S}$. Since $\boldsymbol{S} = \boldsymbol{S} \circ \boldsymbol{Q}$ we have

$$\begin{aligned} \boldsymbol{S}\boldsymbol{x} = \boldsymbol{x} \quad &\Leftrightarrow \quad \boldsymbol{S}(\boldsymbol{Q}\boldsymbol{x}) = \boldsymbol{x} \\ &\Rightarrow \quad \boldsymbol{Q} \circ \boldsymbol{S}(\boldsymbol{Q}\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} \end{aligned}$$

hence $\boldsymbol{Q}\boldsymbol{x} \in \operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S})$ and $\boldsymbol{S}(\boldsymbol{Q}\boldsymbol{x}) = \boldsymbol{x}$. Thus, $\boldsymbol{x} \in \boldsymbol{S}(\operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S}))$ and we conclude $\operatorname{Fix} \boldsymbol{S} \subset \boldsymbol{S}(\operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S}))$. Conversely, let $\boldsymbol{x} \in \operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S})$. Since $\boldsymbol{S} = \boldsymbol{S} \circ \boldsymbol{Q}$, we have

$$\begin{aligned} \boldsymbol{Q}(\boldsymbol{S}\boldsymbol{x}) = \boldsymbol{x} \quad &\Rightarrow \quad \boldsymbol{S}(\boldsymbol{Q}(\boldsymbol{S}\boldsymbol{x})) = \boldsymbol{S}\boldsymbol{x} \\ &\Rightarrow \quad \boldsymbol{S}(\boldsymbol{S}\boldsymbol{x}) = \boldsymbol{S}\boldsymbol{x} \\ &\Rightarrow \quad \boldsymbol{S}\boldsymbol{x} \in \operatorname{Fix} \boldsymbol{S}. \end{aligned}$$

Thus $\boldsymbol{S}(\operatorname{Fix}(\boldsymbol{Q} \circ \boldsymbol{S})) \subset \operatorname{Fix} \boldsymbol{S}$ and the result follows. □ □

Now we prove that Krasnosel'skiĭ–Mann iterations defined by $\boldsymbol{S}$ approximate the solutions to Problem 3.2.4 via their *shadows* in $\operatorname{ran} \boldsymbol{V}$. We derive our result for firmly quasinonexpansive (or class $\mathfrak{T}$) operators $\boldsymbol{T}$ such that $\mathbf{Id} - \boldsymbol{T}$ is demiclosed at $\mathbf{0}$, and for $\alpha-$averaged nonexpansive operators, for some $\alpha \in \,]0,1[$.

Proposition 3.2.6. *In the context of Problem 3.2.4, define*

$$\boldsymbol{T} : \operatorname{ran} \boldsymbol{V} \to \operatorname{ran} \boldsymbol{V} : x \mapsto P_{\operatorname{ran} \boldsymbol{V}} \circ \boldsymbol{S}x \tag{3.2.22}$$

and consider the sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ defined by the recurrence

$$\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}, \quad (\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = (1-\lambda_n)\,\boldsymbol{x}_n + \lambda_n \boldsymbol{S}\boldsymbol{x}_n. \tag{3.2.23}$$

Moreover, suppose that one of the following holds:

(i) *$\boldsymbol{T}$ is firmly quasinonexpansive, $\mathbf{Id} - \boldsymbol{T}$ is demiclosed at $\mathbf{0}$ in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$, and $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in $[\varepsilon, 2-\varepsilon]$ for some $\varepsilon \in \,]0,1[$.*

(ii) *$\boldsymbol{T}$ is $\alpha-$averaged nonexpansive in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$ for some $\alpha \in \,]0,1[$ and $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in $[0, 1/\alpha]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty$.*

Then the following hold:

1. *$(P_{\operatorname{ran} \boldsymbol{V}} \boldsymbol{x}_n)_{n\in\mathbb{N}}$ is Féjer monotone in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$ with respect to $\operatorname{Fix} \boldsymbol{T}$.*

2. *$(P_{\operatorname{ran} \boldsymbol{V}}(\boldsymbol{S}\boldsymbol{x}_n - \boldsymbol{x}_n))_{n\in\mathbb{N}}$ converges strongly to $\mathbf{0}$ in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$.*

3. *$(P_{\operatorname{ran} \boldsymbol{V}} \boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly in $(\operatorname{ran} \boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}})$ to some $\hat{\boldsymbol{x}} \in \operatorname{Fix} \boldsymbol{T}$ and $\boldsymbol{S}\hat{\boldsymbol{x}}$ is a solution to Problem 3.2.4.*

Proof. Since $\boldsymbol{V}$ is a monotone bounded self-adjoint linear operator and ran $\boldsymbol{V}$ is closed, it follows from Proposition 3.2.2 that (ran $\boldsymbol{V}, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}}$) is a real Hilbert space. Moreover, since $\boldsymbol{S} = \boldsymbol{S} \circ P_{\operatorname{ran} \boldsymbol{V}}$ and $\operatorname{Fix} \boldsymbol{S} \neq \varnothing$, Lemma 3.2.5 yields

$$\boldsymbol{S}(\operatorname{Fix} \boldsymbol{T}) = \operatorname{Fix} \boldsymbol{S} \tag{3.2.24}$$

and, hence, $\operatorname{Fix} \boldsymbol{T} \neq \varnothing$. Therefore, since $P_{\operatorname{ran} \boldsymbol{V}}$ is linear, by defining, for every $n \in \mathbb{N}$, $\boldsymbol{y}_n = P_{\operatorname{ran} \boldsymbol{V}} \boldsymbol{x}_n$, it follows from (3.2.23) and $\boldsymbol{T} = \boldsymbol{T} \circ P_{\operatorname{ran} \boldsymbol{V}}$ that

$$\boldsymbol{y}_0 \in \operatorname{ran} \boldsymbol{V}, \quad (\forall n \in \mathbb{N}) \quad \boldsymbol{y}_{n+1} = (1 - \lambda_n)\, \boldsymbol{y}_n + \lambda_n \boldsymbol{T} \boldsymbol{y}_n. \tag{3.2.25}$$

If we assume (i), since $\inf_{n \in \mathbb{N}} \lambda_n (2 - \lambda_n) \geq \varepsilon^2$, 1 and 2 follow from [17, Proposition 4.2] in the error free case. Finally, it follows from [17, Theorem 5.2(i)] that $\boldsymbol{y}_n$ converges weakly to some $\hat{\boldsymbol{y}} \in \operatorname{Fix} \boldsymbol{T}$, and 3 is obtained from (3.2.24).

On the other hand, if we assume (ii), the results follow from [3, Proposition 5.16] and (3.2.24). □

Remark 3.2.7.

1. *Previous results does not include summable errors for ease of the presentation, but they can be included effortlessly.*
2. *In the case when $\boldsymbol{V}$ is strongly monotone, we have* $\operatorname{ran} \boldsymbol{V} = \boldsymbol{\mathcal{H}}$, $P_{\operatorname{ran} \boldsymbol{V}} = \mathbf{Id}$, *and Proposition 3.2.6(i) and Proposition 3.2.6(ii) are equivalent to [17, Theorem 5.2(i)] and [3, Proposition 5.16], respectively.*
3. *In [19, 23], a version of [3, Proposition 5.16] allowing for operators $(\boldsymbol{S}_k)_{k \in \mathbb{N}}$ and $(\boldsymbol{V}_k)_{k \in \mathbb{N}}$ varying among iterations is proposed. This modification allows to include variable step-sizes in primal-dual algorithms. In our context, the difficulty of including such generalization lies on the variation of the real Hilbert spaces* $(\operatorname{ran} \boldsymbol{V}_k, \langle \cdot \mid \cdot \rangle_{\boldsymbol{V}_k})_{k \in \mathbb{N}}$, *which complicates the asymptotic analysis.*

3.2.3 Application to Primal-Dual algorithms for monotone inclusions

Now we focus on the asymptotic analysis of the relaxed primal-dual algorithm in (3.2.6) for solving Problem 3.2.1. First, note that $\boldsymbol{Z} = \operatorname{zer} \boldsymbol{M}$, where

$$\boldsymbol{M} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} : (x, u) \mapsto \{(y, v) \in \boldsymbol{\mathcal{H}} \mid y \in Ax + L^* u,\ v \in B^{-1} u - Lx\} \tag{3.2.26}$$

is maximally monotone in $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{G}$ [9, Proposition 2.7(iii)]. We define

$$\boldsymbol{V} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, u) \mapsto \left(\Upsilon^{-1} x - L^* u, \Sigma^{-1} u - Lx \right), \tag{3.2.27}$$

where $\Sigma : \mathcal{G} \to \mathcal{G}$ and $\Upsilon : \mathcal{H} \to \mathcal{H}$ are strongly monotone self-adjoint linear operators such that $\|\sqrt{\Sigma} L \sqrt{\Upsilon}\| \leq 1$. In the case when, $\|\sqrt{\Sigma} L \sqrt{\Upsilon}\| < 1$, $\boldsymbol{V}$ is strongly monotone

[20, eq. (6.15)] and the primal-dual algorithm is obtained by applying the proximal point algorithm (PPA) to the maximally monotone operator $\boldsymbol{V}^{-1}\boldsymbol{M}$ in the space $(\mathcal{H}\times\mathcal{G}, \langle\!\langle\cdot \mid \cdot\rangle\!\rangle_{\boldsymbol{V}})$ [6, 19, 21, 31, 37, 42]. In the case when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| = 1$, $\boldsymbol{V}$ is no longer strongly monotone and $\langle\!\langle\cdot \mid \cdot\rangle\!\rangle_{\boldsymbol{V}}$ does not define an inner product. However, if $\operatorname{ran}\boldsymbol{V}$ is closed, $(\operatorname{ran}\boldsymbol{V}, \langle\!\langle\cdot \mid \cdot\rangle\!\rangle_{\boldsymbol{V}})$ is a real Hilbert space in view of Proposition 3.2.2, and we obtain the convergence of the primal-dual algorithm when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\|\leq 1$ in this Hilbert space using Proposition 3.2.6. The following result provides conditions on Problem 3.2.1 guaranteeing that $\operatorname{ran}\boldsymbol{V}$ is closed.

Proposition 3.2.8. *In the context of Problem 3.2.1, set $\boldsymbol{\mathcal{H}} = \mathcal{H}\oplus\mathcal{G}$, let $\Sigma\colon\mathcal{G}\to\mathcal{G}$ and $\Upsilon\colon\mathcal{H}\to\mathcal{H}$ be strongly monotone self-adjoint linear bounded operators such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\|\leq 1$, and let $\boldsymbol{V}$ be the operator defined in (3.2.27). Then, the following hold:*

1. *$\boldsymbol{V}$ is linear, bounded, self-adjoint, and $\frac{\tau\sigma}{\tau+\sigma}$-cocoercive, where $\sigma>0$ and $\tau>0$ are the strongly monotone constants of Σ and Υ, respectively.*

2. *The followings statements are equivalent.*

   (a) $\operatorname{ran}\boldsymbol{V}$ is closed in $\boldsymbol{\mathcal{H}}$.

   (b) $\operatorname{ran}(\Sigma^{-1}-L\Upsilon L^*)$ is closed in $\mathcal{G}$.

   (c) $\operatorname{ran}(\Upsilon^{-1}-L^*\Sigma L)$ is closed in $\mathcal{H}$.

Proof. 1: It is a direct consequence of [12, Proposition 2.1]. 2: (2a $\Rightarrow$ 2b). Let $(v_n)_{n\in\mathbb{N}}$ be sequence in $\operatorname{ran}(\Sigma^{-1}-L\Upsilon L^*)$ such that $v_n\to v$. Therefore, for each $n\in\mathbb{N}$, there exists $u_n\in\mathcal{G}$ such that $v_n=\Sigma^{-1}u_n-L\Upsilon L^*u_n$. Note that $\boldsymbol{V}(\Upsilon L^*u_n,u_n)=(0,v_n)\to(0,v)$. Since $\operatorname{ran}\boldsymbol{V}$ is closed, there exists some $(x,u)\in\mathcal{H}\times\mathcal{G}$ such that $\boldsymbol{V}(x,u)=(0,v)$, i.e.,

$$
\begin{aligned}
\boldsymbol{V}(x,u)=(0,v) \quad &\Leftrightarrow \quad \begin{cases}\Upsilon^{-1}x-L^*u=0\\ \Sigma^{-1}u-Lx=v\end{cases}\\
&\Rightarrow \quad \Sigma^{-1}u-L\Upsilon L^*u=v.
\end{aligned}
$$

Then $v\in\operatorname{ran}(\Sigma^{-1}-L\Upsilon L^*)$ and, therefore, $\operatorname{ran}(\Sigma^{-1}-L\Upsilon L^*)$ is closed.

(2b $\Rightarrow$ 2a). Let $\big((y_n,v_n)\big)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{ran}\boldsymbol{V}$ such that $(y_n,v_n)\to(y,v)$. Then, for every $n\in\mathbb{N}$, there exists (x_n,u_n) such that $(y_n,u_n)=\boldsymbol{V}(x_n,u_n)$, or equivalently,

$$
\begin{cases} y_n=\Upsilon^{-1}x_n-L^*u_n\\ v_n=\Sigma^{-1}u_n-Lx_n.\end{cases} \tag{3.2.28}
$$

By applying $L\Upsilon$ to the first equation in (3.2.28) and adding it to the second equation, by the continuity of Υ and L, we obtain

$$
(\Sigma^{-1}-L\Upsilon L^*)u_n = L\Upsilon y_n + v_n \ \to \ L\Upsilon y + v. \tag{3.2.29}
$$

Hence, since $\operatorname{ran}(\Sigma^{-1} - L\Upsilon L^*)$ is closed, there exists $u \in \mathcal{G}$ such that $L\Upsilon y + v = (\Sigma^{-1} - L\Upsilon L^*)u$. We deduce $\boldsymbol{V}(\Upsilon(L^*u + y), u) = (y, v)$ and, therefore, $\operatorname{ran}\boldsymbol{V}$ is closed.

(2a $\Leftrightarrow$ 2c). Define $\tilde{\boldsymbol{V}} : \mathcal{G} \oplus \mathcal{H} \to \mathcal{G} \oplus \mathcal{H} : (u, x) \mapsto (\Sigma^{-1}u - Lx, \Upsilon^{-1}x - L^*u)$. By the equivalence 2a $\Leftrightarrow$ 2b $\operatorname{ran}\tilde{\boldsymbol{V}}$ is closed if and only if $\operatorname{ran}(\Upsilon^{-1} - L^*\Sigma L)$ is closed. Consider the isometric map $\boldsymbol{\Lambda} : \mathcal{H} \oplus \mathcal{G} \to \mathcal{G} \oplus \mathcal{H} : (x, u) \mapsto (u, x)$. Since $\boldsymbol{\Lambda} \circ \boldsymbol{V} = \tilde{\boldsymbol{V}}$, $\operatorname{ran}\boldsymbol{V}$ is closed if and only if $\operatorname{ran}\tilde{\boldsymbol{V}}$ is closed and the result follows.

□

Remark 3.2.9. *1. In the case when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| < 1$, we have that $\Upsilon^{-1} - L\Sigma L^*$ is strongly monotone and, thus, invertible. This is indeed an equivalence which follows from [12, eq. (2.7)]. Therefore,* $\operatorname{ran}(\Upsilon^{-1} - L\Sigma L^*) = \mathcal{G}$ *and* $\operatorname{ran}\boldsymbol{V}$ *is closed in view of Proposition 3.2.8.*

2. Assume that $\operatorname{ran} L = \mathcal{G}$*. Note that, for every $u \in \mathcal{G}$, $\langle L\Upsilon L^*u \mid u\rangle \geq \tau\|L^*u\|^2 \geq \tau\alpha^2\|u\|^2$, where $\tau > 0$ is the strong monotonicity parameter of Υ and the existence of $\alpha > 0$ is guaranteed by [3, Fact 2.26]. Hence, by setting $\Sigma = (L\Upsilon L^*)^{-1}$ we have $\Sigma^{-1} - L\Upsilon L^* = 0$. Hence,* $\operatorname{ran}(\Sigma^{-1} - L\Upsilon L^*) = \{0\}$ *which is closed and Proposition 3.2.8 implies that* $\operatorname{ran}\boldsymbol{V}$ *is closed. This case arises in wavelets transformations in image and signal processing (see, e.g., [33]).*

The next theorem is the main result of this section, in which we interpret the primal-dual splitting as a relaxed proximal point algorithm (PPA) applied to the primal-dual operator

$$\boldsymbol{W} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon (x, u) \mapsto \{(y, v) \in \boldsymbol{\mathcal{H}} \mid \boldsymbol{V}(y, v) \in \boldsymbol{M}(x, u)\}, \tag{3.2.30}$$

where $\boldsymbol{M}$ and $\boldsymbol{V}$ are defined in (3.2.26) and (3.2.27), respectively. Note that, in the case when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| < 1$, $\boldsymbol{V}$ is invertible and $\boldsymbol{W} = \boldsymbol{V}^{-1}\boldsymbol{M}$, which is maximally monotone in $(\boldsymbol{\mathcal{H}}, \langle\langle\cdot \mid \boldsymbol{V}\cdot\rangle\rangle)$ in view of [3, Proposition 20.24]. This implies that $J_{\boldsymbol{W}}$ is firmly nonexpansive under the same metric [3, Proposition 23.8(iii)]. These properties do not hold when $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| = 1$, but $P_{\operatorname{ran}\boldsymbol{V}} \circ J_{\boldsymbol{W}}$ is firmly nonexpansive in the real Hilbert space $(\operatorname{ran}\boldsymbol{V}, \langle\langle\cdot \mid \boldsymbol{V}\cdot\rangle\rangle)$, from which the weak convergence of primal-dual algorithm is obtained.

Theorem 3.2.10. *In the context of Problem 3.2.1, let $\boldsymbol{V}$ be the operator defined in (3.2.27), where $\Sigma : \mathcal{G} \to \mathcal{G}$ and $\Upsilon : \mathcal{H} \to \mathcal{H}$ are self-adjoint linear strongly monotone operators such that $\|\sqrt{\Sigma}L\sqrt{\Upsilon}\| \leq 1$, and suppose that* $\operatorname{ran}\boldsymbol{V}$ *is closed. Moreover, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 2]$ satisfying $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, and consider the sequence $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$ defined by the recurrence*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon L^*u_n) \\ q_{n+1} = J_{\Sigma B^{-1}}(u_n + \Sigma L(2p_{n+1} - x_n)) \\ (x_{n+1}, u_{n+1}) = (1 - \lambda_n)(x_n, u_n) + \lambda_n(p_{n+1}, q_{n+1}), \end{array} \right. \tag{3.2.31}$$

where $(x_0, u_0) \in \mathcal{H} \times \mathcal{G}$*. Then* $\big(P_{\operatorname{ran} \boldsymbol{V}}(x_n, u_n)\big)_{n\in\mathbb{N}}$ *converges weakly in* $(\operatorname{ran} \boldsymbol{V}, \langle\!\langle \cdot \mid \cdot \rangle\!\rangle_{\boldsymbol{V}})$ *to some* $(\hat{y}, \hat{v}) \in \operatorname{Fix}(P_{\operatorname{ran} \boldsymbol{V}} \circ J_{\boldsymbol{W}})$*, where* $\boldsymbol{W}$ *is defined in* (3.2.30)*. Moreover,*

$$\big(J_{\Upsilon A}(\hat{y} - \Upsilon L^*\hat{v}), J_{\Sigma B^{-1}}\left(\hat{v} + \Sigma L(2J_{\Upsilon A}(\hat{y} - \Upsilon L^*\hat{v}) - \hat{y})\right)\big) \tag{3.2.32}$$

is a solution to Problem 3.2.1.

Proof. First, it follows from Proposition 3.2.8(1) that $\boldsymbol{V}$ is a monotone self-adjoint linear bounded operator. Note that

$$\operatorname{Fix} J_{\boldsymbol{W}} = \operatorname{zer}\boldsymbol{W} = \operatorname{zer}\boldsymbol{M} = \boldsymbol{Z} \neq \varnothing \tag{3.2.33}$$

and, for every (x, u) and (p, q) in $\boldsymbol{\mathcal{H}}$,

$$\begin{aligned}
(p, q) \in J_{\boldsymbol{W}}(x, u) \quad &\Leftrightarrow \quad (x - p, u - q) \in \boldsymbol{W}(p, q)\\
&\Leftrightarrow \quad \boldsymbol{V}(x - p, u - q) \in \boldsymbol{M}(p, q)\\
&\Leftrightarrow \quad \begin{cases} \Upsilon^{-1}(x - p) - L^*(u - q) \in Ap + L^*q,\\ \Sigma^{-1}(u - q) - L(x - p) \in B^{-1}q - Lp. \end{cases}\\
&\Leftrightarrow \quad \begin{cases} p = J_{\Upsilon A}(x - \Upsilon L^* u),\\ q = J_{\Sigma B^{-1}}\left(u + \Sigma L(2p - x)\right). \end{cases}
\end{aligned} \tag{3.2.34}$$

Hence, $J_{\boldsymbol{W}}$ is single valued and, for every $(x, u) \in \boldsymbol{\mathcal{H}}$,

$$\begin{aligned}
J_{\boldsymbol{W}}(x, u) &= \big(J_{\Upsilon A}\left(x - \Upsilon L^* u\right), J_{\Sigma B^{-1}}(u - \Sigma L x + 2\Sigma L J_{\Upsilon A}\left(x - \Upsilon L^* u\right))\big)\\
&= \boldsymbol{R}\left(\Upsilon^{-1}x - L^*u, \Sigma^{-1}u - Lx\right)\\
&= \boldsymbol{R}(\boldsymbol{V}(x, u)),
\end{aligned}$$

where $\boldsymbol{R}\colon (x, u) \mapsto \big(J_{\Upsilon A}(\Upsilon x), J_{\Sigma B^{-1}}(\Sigma u + 2\Sigma L J_{\Upsilon A}(\Upsilon x))\big)$, which yields

$$J_{\boldsymbol{W}} = \boldsymbol{R} \circ \boldsymbol{V} = \boldsymbol{R} \circ \boldsymbol{V} \circ P_{\operatorname{ran}\boldsymbol{V}} = J_{\boldsymbol{W}} \circ P_{\operatorname{ran}\boldsymbol{V}}. \tag{3.2.35}$$

Moreover, by defining $\boldsymbol{T} = P_{\operatorname{ran} \boldsymbol{V}} \circ J_{\boldsymbol{W}}$, we deduce from $\boldsymbol{V} = \boldsymbol{V} \circ P_{\operatorname{ran} \boldsymbol{V}}$, (3.2.30), $\ker \boldsymbol{V} \oplus \operatorname{ran} \boldsymbol{V} = \boldsymbol{\mathcal{H}}$, and the monotonicity of $\boldsymbol{M}$ that, for every $\boldsymbol{z}$ and $\boldsymbol{w}$ in $\operatorname{ran} \boldsymbol{V}$,

$$\begin{aligned}
&\langle\!\langle \boldsymbol{T}\boldsymbol{z} - \boldsymbol{T}\boldsymbol{w} \mid (\mathbf{Id} - \boldsymbol{T})\boldsymbol{z} - (\mathbf{Id} - \boldsymbol{T})\boldsymbol{w} \rangle\!\rangle_{\boldsymbol{V}}\\
&\qquad = \langle\!\langle J_{\boldsymbol{W}}\boldsymbol{z} - J_{\boldsymbol{W}}\boldsymbol{w} \mid \boldsymbol{V}(\boldsymbol{z} - J_{\boldsymbol{W}}\boldsymbol{z}) - \boldsymbol{V}(\boldsymbol{w} - J_{\boldsymbol{W}}\boldsymbol{w}) \rangle\!\rangle\\
&\qquad \geq 0,
\end{aligned} \tag{3.2.36}$$

which yields the firm nonexpansivity of $\boldsymbol{T}$ in $(\operatorname{ran} \boldsymbol{V}, \langle\!\langle \cdot \mid \boldsymbol{V} \cdot \rangle\!\rangle)$. Therefore, it follows from (3.2.33) and (3.2.35) that Problem 3.2.1 is a particular instance of Problem 3.2.4 with $\boldsymbol{S} = J_{\boldsymbol{W}}$. In addition, (3.2.31) and (3.2.34) yields

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = (1 - \lambda_n)\boldsymbol{x}_n + \lambda_n J_{\boldsymbol{W}}\boldsymbol{x}_n, \tag{3.2.37}$$

where, for every $n \in \mathbb{N}$, $\boldsymbol{x}_n = (x_n, u_n)$. Altogether, we obtain the results by applying Theorem 3.2.6(ii) with $\alpha = 1/2$ and $\boldsymbol{S} = J_{\boldsymbol{W}}$. ☐

Remark 3.2.11. *1. Since $J_{\boldsymbol{W}} \circ P_{\operatorname{ran} \boldsymbol{V}} = J_{\boldsymbol{W}}$, the sequence $\big(P_{\operatorname{ran} \boldsymbol{V}}(x_n, u_n)\big)_{n\in\mathbb{N}}$ is not needed in practice. Indeed, since*

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad \|\boldsymbol{x}\|_{\boldsymbol{V}} = \|P_{\operatorname{ran} \boldsymbol{V}} \boldsymbol{x}\|_{\boldsymbol{V}},$$

we can use a stopping criteria only involving $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$.

2. Suppose that $\mathcal{G} = \oplus_{i=1}^m \mathcal{G}_i$, $B\colon (u_i)_{1\le i\le m} \mapsto \times_{i=1}^m B_i u_i$, $\Sigma\colon (u_i)_{1\le i\le m} \mapsto (\Sigma_i u_i)_{1\le i\le m}$, and $L\colon x \mapsto (L_i x)_{1\le i\le m}$, where, for every $i \in \{1, \ldots, m\}$, $\mathcal{G}_i$ is a real Hilbert space, B_i is maximally monotone, $\Sigma_i\colon \mathcal{G}_i \to \mathcal{G}_i$ is a strongly monotone self-adjoint linear bounded operator, and $L_i\colon \mathcal{H} \to \mathcal{G}_i$ is a linear bounded operator. In this context, the inclusion in (3.2.2) is equivalent to

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + \sum_{i=1}^m L_i^* B_i L_i x. \tag{3.2.38}$$

Then, under the assumptions

$$\sum_{i=1}^m \left\| \sqrt{\Sigma_i} L_i \sqrt{\Upsilon} \right\|^2 \le 1 \quad \text{and} \quad \operatorname{ran} \left(\Upsilon^{-1} - \sum_{i=1}^m L_i^* \Sigma_i L_i \right) \quad \text{is closed,} \tag{3.2.39}$$

Proposition 3.2.8 and Theorem 3.2.10 ensures the convergence of (3.2.31), which reduces to

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon \sum_{i=1}^m L_i^* u_{i,n}) \\ x_{n+1} = (1-\lambda_n) x_n + \lambda_n p_{n+1} \\ \text{for } i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} q_{i,n+1} = J_{\Sigma_i B_i^{-1}}(u_{i,n} + \Sigma_i L_i (2p_{n+1} - x_n)) \\ u_{i,n+1} = (1-\lambda_n) u_{i,n} + \lambda_n q_{i,n+1}. \end{array} \right. \end{array} \right. \tag{3.2.40}$$

Note that (3.2.40) has the same structure than the algorithm in [20, Corolary 6.2] without considering cocoercive operators and the convergence is guaranteed under the weaker assumption (3.2.39) in view of Remark 3.2.9(1).

3. In the context of the optimization problem in (3.2.4), (3.2.31) reduces to

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = \operatorname{prox}_f^{\Upsilon^{-1}}(x_n - \Upsilon L^* u_n) \\ q_{n+1} = \operatorname{prox}_{g^*}^{\Sigma^{-1}}(u_n + \Sigma L(2p_{n+1} - x_n)) \\ (x_{n+1}, u_{n+1}) = (1-\lambda_n)(x_n, u_n) + \lambda_n (p_{n+1}, q_{n+1}), \end{array} \right. \tag{3.2.41}$$

Under the additional condition $\operatorname{ran} \boldsymbol{V}$ closed, Theorem 3.2.10 generalizes [21, Theorem 3.3] to infinite dimensional spaces and allowing preconditioners and a larger choice of parameters $(\lambda_n)_{n\in\mathbb{N}}$. Indeed, in finite dimensional spaces, $\operatorname{ran} \boldsymbol{V}$ is closed, *Theorem 3.2.10 implies that* $P_{\operatorname{ran} \boldsymbol{V}}(x_n, u_n) \to (\hat{y}, \hat{v}) \in \operatorname{ran} \boldsymbol{V}$ *in* $(\operatorname{ran} \boldsymbol{V}, \langle\!\langle \cdot \mid \cdot \rangle\!\rangle_{\boldsymbol{V}})$ *and, since* $J_{\boldsymbol{W}} = J_{\boldsymbol{W}} \circ P_{\operatorname{ran} \boldsymbol{V}}$ *is continuous, we conclude* $(p_{n+1}, q_{n+1}) = J_{\boldsymbol{W}}(x_n, u_n) = J_{\boldsymbol{W}}(P_{\operatorname{ran} \boldsymbol{V}}(x_n, u_n)) \to J_{\boldsymbol{W}}(\hat{y}, \hat{v}) \in \boldsymbol{Z}$. *In order to guarantee the convergence of the relaxed sequence* $((x_n, u_n))_{n\in\mathbb{N}}$, *it is enough to suppose* $(\lambda_n)_{n\in\mathbb{N}} \subset [\epsilon, 2-\epsilon]$ *for some* $\epsilon \in \,]0,1[$, *and use the argument in [21, p.473].*

4. *In the particular case when* $\|\sqrt{\Sigma} L \sqrt{\Upsilon}\| < 1$, *it follows from [19, eq. (6.15)] (see also [37, Lemma 1]) that* $\boldsymbol{V}$ *is strongly monotone, which yields* $\operatorname{ran} \boldsymbol{V} = \boldsymbol{\mathcal{H}}$ *and* $P_{\operatorname{ran} \boldsymbol{V}} = \mathbf{Id}$. *Hence, we recover from Theorem 3.2.10 the weak convergence of* $((x_n, u_n))_{n\in\mathbb{N}}$ *to a solution to Problem 3.2.1 proved in [6, 19, 21, 31, 37, 42].*

5. *In the particular instance when* $\lambda_n \equiv 1$, *the weak convergence of* (3.2.31) *is deduced without any range closedness in [12, Remark 3.4(4)]. The result is obtained from an alternative formulation of the algorithm and the extension to* $\lambda_n \not\equiv 1$ *is not clear. As we will show in Section 3.2.5, the additional relaxation step is relevant in the efficiency of the algorithm.*

3.2.4 Case $L = \mathrm{Id}$: Douglas–Rachford splitting

In this section, we study the particular case of Problem 3.2.1 when $L = \mathrm{Id}$. In this context, the following result is a refinement of Theorem 3.2.10, which relates the primal-dual algorithm in (3.2.31) with Douglas-Rachford splitting (DRS) when

$$\Upsilon = \Sigma^{-1} \text{ is strongly monotone.} \tag{3.2.42}$$

When $\Upsilon = \tau\mathrm{Id}$ and $\Sigma = \sigma\mathrm{Id}$, (3.2.42) reads $\sigma\tau = 1$ and the connection of (3.2.31) with DRS is discovered in [15, Section 4.2] in the optimization context. However, the convergence is guaranteed only if $\tau\sigma < 1$, which is extended to the case $\sigma\tau = 1$ in [21, Section 3.1.3] in the finite dimensional setting. Previous connection allows us to recover the classical convergence results in [24, 32] when $\Upsilon = \tau\mathrm{Id}$ with our approach. Define the operator

$$G_{\Upsilon,B,A} = J_{\Upsilon B} \circ (2J_{\Upsilon A} - \mathrm{Id}) + (\mathrm{Id} - J_{\Upsilon A}), \tag{3.2.43}$$

and we recall that relaxed DRS iterations are defined by the recurrence

$$z_0 \in \mathcal{H}, \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = (1-\lambda_n) z_n + \lambda_n G_{\Upsilon,B,A} z_n, \tag{3.2.44}$$

where $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in $[0,2]$.

Proposition 3.2.12. *In the context of Problem 3.2.1, set* $L = \mathrm{Id}$, *let* Υ *be a strongly monotone self-adjoint linear operator, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $[0,2]$ *satisfying* $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, *and consider the sequence* $\big((x_n, u_n)\big)_{n\in\mathbb{N}}$ *defined by the recurrence*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_{n+1} = J_{\Upsilon A}(x_n - \Upsilon u_n) \\ q_{n+1} = J_{\Upsilon^{-1} B^{-1}}\big(u_n + \Upsilon^{-1}(2p_{n+1} - x_n)\big) \\ (x_{n+1}, u_{n+1}) = (1-\lambda_n)(x_n, u_n) + \lambda_n (p_{n+1}, q_{n+1}), \end{array} \right. \tag{3.2.45}$$

where $(x_0,u_0) \in \mathcal{H}\times\mathcal{H}$. *Then, by setting, for every* $n\in\mathbb{N}$, $z_n = x_n - \Upsilon y_n$, $(z_n)_{n\in\mathbb{N}}$ *converges weakly in* $\mathcal{H}$ *to some* $\hat{z}\in \operatorname{Fix} G_{\Upsilon,B,A}$ *and*

$$\big(J_{\Upsilon A}\hat{z}, -\Upsilon^{-1}(\hat{z}-J_{\Upsilon A}\hat{z})\big)$$

is a solution to Problem 3.2.1. Moreover, we have

$$(\forall n\in\mathbb{N})\quad z_{n+1} = (1-\lambda_n)z_n + \lambda_n G_{\Upsilon,B,A} z_n. \tag{3.2.46}$$

Proof. Note that, since $L=\mathrm{Id}$, (3.2.27) and (3.2.42) yield $\boldsymbol{V}\colon (x,u)\mapsto(\Upsilon^{-1}x-u,\Upsilon u-x)$ and Remark 3.2.9(2) implies that $\operatorname{ran}\boldsymbol{V}$ is closed. Hence, it follows from Theorem 3.2.10 and (3.2.34) in the case $L=\mathrm{Id}$ that $(P_{\operatorname{ran}\boldsymbol{V}}(x_n,u_n))_{n\in\mathbb{N}}$ converges weakly in $(\operatorname{ran}\boldsymbol{V},\langle\!\langle\cdot\mid\cdot\rangle\!\rangle_{\boldsymbol{V}})$ to some $(\hat{y},\hat{v})\in\operatorname{Fix}(P_{\operatorname{ran}\boldsymbol{V}}\circ J_{\boldsymbol{W}})$ and

$$(\hat{x},\hat{u}) = J_{\boldsymbol{W}}(\hat{y},\hat{v}) = \Big(J_{\Upsilon A}(\hat{y}-\Upsilon\hat{v}), J_{\Upsilon^{-1}B^{-1}}\big(\hat{v}+\Upsilon^{-1}(2\hat{x}-\hat{y})\big)\Big)\in\boldsymbol{Z}. \tag{3.2.47}$$

Set $\Lambda\colon (x,u)\mapsto x-\Upsilon u$, and $\boldsymbol{\Upsilon}\colon (x,u)\mapsto(\Upsilon x,\Upsilon u)$. Note that Λ is surjective, that

$$\Lambda^*\Lambda = \boldsymbol{V}\circ\boldsymbol{\Upsilon},\quad \operatorname{ran}\boldsymbol{V} = \operatorname{ran}\Lambda^*, \tag{3.2.48}$$

and, in view of [3, Fact 2.25(iv)], that

$$\mathcal{H}\times\mathcal{H} = \operatorname{ran}\Lambda^*\oplus\ker\Lambda. \tag{3.2.49}$$

Then, for every $(x,u)\in\mathcal{H}\times\mathcal{H}$, it follows from (3.2.30) and (3.2.34) in the case $L=\mathrm{Id}$, (3.2.42), [3, Proposition 23.34 (iii)], and (3.2.43) that

$$\begin{aligned}\Lambda(J_{\boldsymbol{W}}(x,u)) &= J_{\Upsilon A}\left(x-\Upsilon u\right) - \Upsilon J_{\Upsilon^{-1}B^{-1}}\Upsilon^{-1}\big(\Upsilon u - x + 2J_{\Upsilon A}\left(x-\Upsilon u\right)\big)\\ &= -J_{\Upsilon A}(\Lambda(x,u)) + \Lambda(x,u) + J_{\Upsilon B}(2J_{\Upsilon A}(\Lambda(x,u)) - \Lambda(x,u))\\ &= G_{\Upsilon,B,A}(\Lambda(x,u)).\end{aligned} \tag{3.2.50}$$

Moreover, since $(\hat{y},\hat{v})\in\operatorname{Fix}(P_{\operatorname{ran}\boldsymbol{V}}\circ J_{\boldsymbol{W}})$, by using (3.2.50), (3.2.49), and (3.2.48) we deduce

$$\begin{aligned}G_{\Upsilon,B,A}(\Lambda(\hat{y},\hat{v})) &= \Lambda(J_{\boldsymbol{W}}(\hat{y},\hat{v}))\\ &= \Lambda\circ P_{\operatorname{ran}\Lambda^*}(J_{\boldsymbol{W}}(\hat{y},\hat{v}))\\ &= \Lambda(P_{\operatorname{ran}\boldsymbol{V}}\circ J_{\boldsymbol{W}}(\hat{y},\hat{v}))\\ &= \Lambda(\hat{y},\hat{v}),\end{aligned} \tag{3.2.51}$$

and, thus, defining $\hat{z}=\Lambda(\hat{y},\hat{v})$, we obtain $\hat{z}\in\operatorname{Fix} G_{\Upsilon,B,A}$. In addition, it follows from (3.2.47) that $\hat{x}=J_{\Upsilon A}\hat{z}$ and, since $(\hat{y},\hat{v})\in\operatorname{Fix}(P_{\operatorname{ran}\boldsymbol{V}}\circ J_{\boldsymbol{W}})$, we deduce from (3.2.49) and (3.2.34) that

$$\hat{z} = \Lambda(\hat{y},\hat{v}) = \Lambda\left(P_{\operatorname{ran}\boldsymbol{V}}\circ J_{\boldsymbol{W}}(\hat{y},\hat{v})\right) = \Lambda J_{\boldsymbol{W}}(\hat{y},\hat{v}) = \Lambda(\hat{x},\hat{u}) = \hat{x}-\Upsilon\hat{u}, \tag{3.2.52}$$

which yields $\hat{u} = -\Upsilon^{-1}(\hat{z} - J_{\Upsilon A}\hat{z})$. Furthermore, noting that, for every $n \in \mathbb{N}$, $z_n = \Lambda(x_n, u_n)$, we deduce from (3.2.45), (3.2.34), and (3.2.50) that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad z_{n+1} &= \Lambda(x_{n+1}, u_{n+1}) \\ &= (1-\lambda_n)\Lambda(x_n, u_n) + \lambda_n \Lambda\left(J_{\boldsymbol{W}}(x_n, u_n)\right) \\ &= (1-\lambda_n) z_n + \lambda_n G_{\Upsilon,B,A} z_n. \end{aligned} \tag{3.2.53}$$

Finally, in order to prove the weak convergence of $(z_n)_{n\in\mathbb{N}}$ to $\hat{z}$, fix $w \in \mathcal{H}$ and set $(p,q) = ((\mathrm{Id} + \Upsilon^2)^{-1}w, -\Upsilon(\mathrm{Id} + \Upsilon^2)^{-1}w)$. We have $(p,q) \in \operatorname{ran} \Lambda^*$, $\Lambda(p,q) = w$ and it follows from (3.2.49), (3.2.48), $(\hat{y},\hat{v}) \in \operatorname{ran} \boldsymbol{V} = \operatorname{ran} \Lambda^*$, and $(P_{\operatorname{ran} \boldsymbol{V}}(x_n,u_n))_{n\in\mathbb{N}} \rightharpoonup (\hat{y},\hat{v})$ that

$$\begin{aligned}\langle z_n - \hat{z} \mid w\rangle &= \langle \Lambda(x_n - \hat{y}, u_n - \hat{v}) \mid \Lambda(p,q)\rangle \\ &= \langle \Lambda\, P_{\operatorname{ran} \Lambda^*}(x_n - \hat{y}, u_n - \hat{v}) \mid \Lambda(p,q)\rangle \\ &= \langle\!\langle P_{\operatorname{ran} \Lambda^*}(x_n - \hat{y}, u_n - \hat{v}) \mid \boldsymbol{V}(\Upsilon p, \Upsilon q)\rangle\!\rangle \\ &= \langle\!\langle P_{\operatorname{ran} \Lambda^*}(x_n - \hat{y}, u_n - \hat{v}) \mid (\Upsilon p, \Upsilon q)\rangle\!\rangle_{\boldsymbol{V}} \\ &= \langle\!\langle P_{\operatorname{ran} \Lambda^*}(x_n, u_n) - (\hat{y}, \hat{v}) \mid (\Upsilon p, \Upsilon q)\rangle\!\rangle_{\boldsymbol{V}} \to 0 \end{aligned} \tag{3.2.54}$$

and the result follows.

□

Remark 3.2.13. 1. *From the proof of Proposition 3.2.12, we deduce* $\Lambda(\operatorname{Fix}(P_{\operatorname{ran} \boldsymbol{V}} \circ J_{\boldsymbol{W}})) \subset \operatorname{Fix} G_{\Upsilon,B,A}$. *The converse inclusion is also true, as detailed in Proposition 3.2.14 in the Appendix.*

2. *Proposition 3.2.12 provides a connection between classical Douglas–Rachford scheme [24] and the primal-dual version in* (3.2.45)*, and we obtain that the auxiliary sequence* $(z_n)_{n\in\mathbb{N}}$ *converges weakly to a* $\hat{z}$ *whose primal-dual shadow is a primal-dual solution. In [41] (see also [2, 4]) the weak convergence of the primal-dual shadow sequence is proved in the case* $\lambda_n \equiv 1$*, by reformulating DRS as an alternative algorithm with primal-dual iterates in* $\operatorname{gra} A$*. This technique does not allow for relaxation steps, since after relaxation the iterates are no longer in* $\operatorname{gra} A$ *unless it is affine linear.*

3.2.5 Numerical experiments

A classical model in image processing is the total variation image restoration [39], which aims at recovering an image from a blurred and noisy observation under piecewise constant assumption on the solution. The model is formulated via the optimization problem

$$\min_{x\in[0,255]^N} \frac{1}{2}\|Rx - b\|_2^2 + \alpha\|\nabla x\|_1 =: F^{TV}(x), \tag{3.2.55}$$

where $x \in [0,255]^N$ is the image of $N = N_1 \times N_2$ pixels to recover from a blurred and noisy observation $b \in \mathbb{R}^m$, $R : \mathbb{R}^N \to \mathbb{R}^m$ is a linear operator representing a Gaussian blur, the discrete gradient $\nabla\colon x \mapsto (D_1 x, D_2 x)$ includes horizontal and vertical differences through linear operators $D_1\colon \mathbb{R}^N \to \mathbb{R}^N$ and $D_2\colon \mathbb{R}^N \to \mathbb{R}^N$, respectively, its adjoint ∇^* is the discrete divergence (see, e.g., [13]), and $\alpha \in]0,+\infty[$. A difficulty in this model is the presence of the non-smooth ℓ^1 norm composed with the discrete gradient operator ∇, which is non-differentiable and its proximity operator has not a closed form.

Note that, by setting $f = \|R\cdot - b\|^2/2$, $g_1 = \alpha\|\cdot\|_1 = g_2$, and $g_3 = \iota_{[0,255]^N}$, $L_1 = D_1$, $L_2 = D_2$, and $L_3 = \mathrm{Id}$, (3.2.55) can be reformulated as $\min(f + \sum_{i=1}^3 g_i \circ L_i)$ or equivalently as (qualification condition holds)

$$\text{find} \quad x \in \mathbb{R}^N \quad \text{such that} \quad 0 \in \partial f(x) + \sum_{i=1}^{3} L_i^* \partial g_i(L_i x), \tag{3.2.56}$$

which is a particular instance of (3.2.38), in view of [3, Theorem 20.25]. Moreover, for every $\tau > 0$, $J_{\tau\partial f} = (\mathrm{Id} + \tau R^* R)^{-1}(\mathrm{Id} - \tau R^* b)$, for every $i \in \{1,2,3\}$, $J_{\tau(\partial g_i)^{-1}} = \tau(\mathrm{Id} - \mathrm{prox}_{g_i/\tau})(\mathrm{Id}/\tau)$, $\mathrm{prox}_{g_3/\tau} = P_{[0,255]^N}$, and, for $i \in \{1,2\}$, $\mathrm{prox}_{g_i/\tau} = \mathrm{prox}_{\alpha\|\cdot\|_1/\tau}$ is the component-wise soft thresholder, computed in [3, Example 24.34]. Note that $(\mathrm{Id}+\tau R^* R)^{-1}$ can be computed efficiently via a diagonalization of R using the fast Fourier transform F [30, Section 4.3]. Altogether, Remark 3.2.11.(2) allows us to write algorithm in (3.2.40) as Algorithm 3 below, where we set $\Upsilon = \tau\mathrm{Id}$, $\Sigma_1 = \sigma_1\mathrm{Id}$, $\Sigma_2 = \sigma_2\mathrm{Id}$, and $\Sigma_3 = \sigma_3\mathrm{Id}$, for $\tau > 0$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $\sigma_3 > 0$. We denote by $\mathcal{R}$ the primal-dual error

$$\mathcal{R} : (x_+, u_+, x, u) \mapsto \sqrt{\frac{\|(x_+, u_+) - (x, u)\|^2}{\|(x, u)\|^2}} \tag{3.2.57}$$

and by $\varepsilon > 0$ the convergence tolerance.

Algorithm 3

Fix $x_0, u_{1,0}, u_{2,0}$, and $u_{3,0}$ in $\mathbb{R}^N$, let τ, σ_1, σ_2, and σ_3 be in $]0,+\infty[$, let $(\lambda_n)_{n\in\mathbb{N}}$ in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, and fix $r_0 > \varepsilon > 0$.
while $r_n > \varepsilon$ **do**
  $p_{n+1} = (\mathrm{Id} + \tau R^*R)^{-1}(x_n - \tau(D_1^* u_{1,n} + D_2^* u_{2,n} + u_{3,n} + R^* b))$
  $x_{n+1} = (1-\lambda_n)x_n + \lambda_n p_{n+1}$
  $q_{1,n+1} = \sigma_1(\mathrm{Id} - \mathrm{prox}_{\alpha\|\cdot\|_1/\sigma_1})(u_{1,n}/\sigma_1 + D_1(2p_{n+1} - x_n))$
  $q_{2,n+1} = \sigma_2(\mathrm{Id} - \mathrm{prox}_{\alpha\|\cdot\|_1/\sigma_2})(u_{2,n}/\sigma_1 + D_2(2p_{n+1} - x_n))$
  $q_{3,n+1} = \sigma_3\big(\mathrm{Id} - P_{[0,255]^N}\big)\,(u_{3,n}/\sigma_3 + 2p_{n+1} - x_n)$
  for $i = 1,2,3$
    $u_{i,n+1} = (1-\lambda_n)u_{i,n} + \lambda_n q_{i,n+1}$
  $r_n = \mathcal{R}\big((x_{n+1}, u_{1,n+1}, u_{2,n+1}, u_{3,n+1}), (x_n, u_{1,n}, u_{2,n}, u_{3,n})\big)$
end while
return $(x_{n+1}, u_{1,n+1}, u_{2,n+1}, u_{3,n+1})$

In this case, (3.2.39) reduces to

$$\tau(\sigma_1\|D_1\|^2 + \sigma_2\|D_2\|^2 + \sigma_3) \leq 1 \tag{3.2.58}$$

and the closed range condition is trivially satisfied. By using the power iteration [35] with tolerance 10^{-9}, we obtain $\|D_1\|^2 = \|D_2\|^2 \approx 3.9998$.

Observe that, when $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$, Algorithm 3 reduces to the algorithm proposed in [15] (when $\sigma\tau(\|D_1\|^2 + \|D_2\|^2 + 1) < 1$) or [21, Theorem 3.3] (algorithm denoted by `condat`), where the case $\sigma\tau(\|D_1\|^2 + \|D_2\|^2 + 1) = 1$ is included.

Since in [12, Section 5.1], the critical step-sizes achieve the best performance, we provide a numerical experiment which compare the efficiency of Algorithm 3 for different values of the parameters τ, σ_1, σ_2, and σ_3 in the boundary of (3.2.58) and different relaxation parameters λ_n. In particular we compare with the case $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$ (`condat`), which turns out to be more efficient than other methods as AFBS [36], MS [9], Condat-Vũ [21, 42] in this context [12, Section 5.1]. For these comparisons, we consider the test image $\overline{x}$ shown in Figure 3.2a of 256×256 pixels ($N_1 = N_2 = 256$) inspired in [43, Section 5]. The operator blur R is set as a Gaussian blur of size 9×9 and standard deviation 4 (applied by MATLAB function *fspecial*) and the observation b is obtained by $b = R\overline{x} + e \in \mathbb{R}^{m_1\times m_2}$, where $m_1 = m_2 = 256$ and e is an additive zero-mean white Gaussian noise with standard deviation 10^{-3} (using *imnoise* function in MATLAB). We generate 20 random realizations of the random variable e leading to 20 observations $(b_i)_{1\leq i\leq 20}$.

We study the efficiency of Algorithm 3 for different values of τ, σ_1, σ_2, and σ_3 and relaxation steps $\lambda_n \equiv \lambda \in \{1, 1.5, 1.9\}$. In order to approximate the best performance step-sizes in the boundary of (3.2.58), we consider $\tau \in \mathcal{C} := \{0.10 + 0.05\cdot n\}_{n=0,\ldots,10}$ and $\sigma_1 = \sigma_2 = \sigma_3 = \sigma = \tau/(1 + \|D_1\|^2 + \|D_2\|^2)$ in the case of `condat`. In the case

Table 3.1: Averages number of iterations for Algorithm 3 with $\tau(\sigma_1\|D_1\|^2+\sigma_2\|D_1\|^2+\sigma_3) = 1$ and `condat` with tolerance 10^{-8}.

Algorithm	τ	σ_1	σ_2	λ_n	$\varepsilon = 10^{-8}$ Av. Time(s)	Av. Iter.
Alg. 3	0.2	0.7425	0.4950	1	82.6373	8844
	0.2	0.7463	0.4975		82.2817	8827
	0.2	0.7493	0.4995		82.5722	8833
	0.2	0.7425	0.4950	1.5	63.1338	6766
	0.2	0.7463	0.4975		63.0645	6754
	0.2	0.7493	0.4995		63.0996	6758
	0.2	0.8044	0.4331	1.9	53.9059	5770
	0.2	0.8085	0.4353		53.8663	5767
	0.2	0.8117	0.4371		53.8022	5761
`condat`	0.2	-	-	1	92.7997	9326
	0.2	-	-	1.5	67.0886	7131
	0.2	-	-	1.9	57.8523	6121

of Algorithm 3 we consider $\sigma_1 = \gamma_1(1-\gamma_2)/(\tau\|D_1\|^2)$, $\sigma_2 = (1-\gamma_1)(1-\gamma_2)/(\tau\|D_2\|^2)$, $\sigma_3 = \gamma_2/\tau$, where $(\tau,\gamma_1,\gamma_2) \in \mathcal{C} \times \{0.01, 0.005, 0.001\} \times \{0.5, 0.55, 0.6, 0.65\}$.

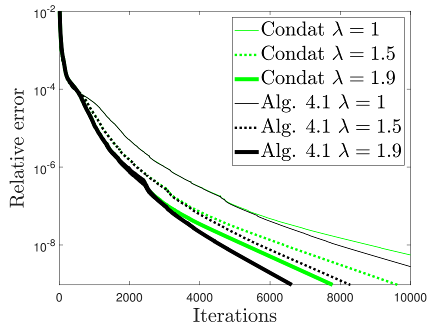
In Table 3.1 we provide the average number of iterations obtained by applying Algorithm 3 for solving (3.2.55) considering the 20 observations $(b_i)_{1\leq i\leq 20}$ and the best set of step-sizes found with the procedure above. The tolerance is set as $\varepsilon = 10^{-8}$. We observe that Algorithm 3 becomes more efficient in iterations as long as the relaxation parameters are larger. The case $\lambda = 1.9$ achieves the tolerance in approximately 35% less iterations than the case $\lambda = 1$. By choosing different parameters σ_1, σ_2, and σ_3, the algorithm achieves the tolerance in approximately 6% less iterations than `condat`.

This conclusion is confirmed in Figure 3.1, which shows the performance obtained with the observation b_4. This figure also shows that both algorithms achieve in less iterations the optimal objective value for higher relaxation parameters, with a slight advantage of Algorithm 3. Note that, since the algorithms under study has the same structure, the CPU time by iteration is very similar.

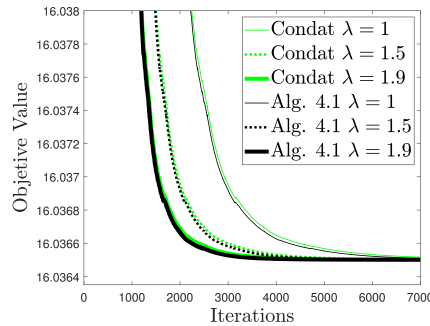
In Figure 3.2 we provide the images reconstructed from observation b_4 by using `condat` and Algorithm 3 after 300 iterations. The best reconstruction, in terms of objective value F^{TV} and PSNR (Peak signal-to-noise ratio) is obtained by Algorithm 3.

Acknowledgments

The first author thanks the support of ANID under grant FONDECYT 1190871 and grant Redes 180032. The second author thanks the support of ANID-Subdirección de Capital Humano/Doctorado Nacional/2018-21181024 and by the Dirección de Postgrado

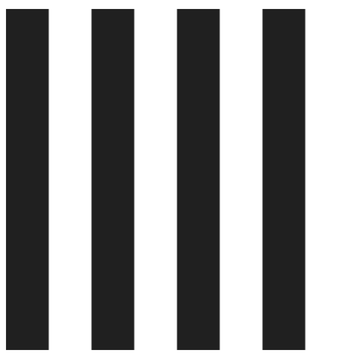


(a)

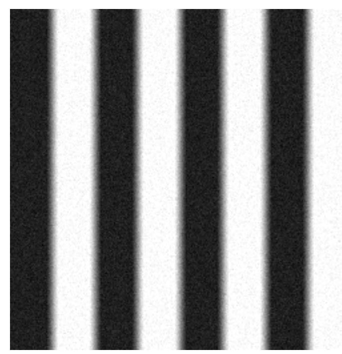


(b)

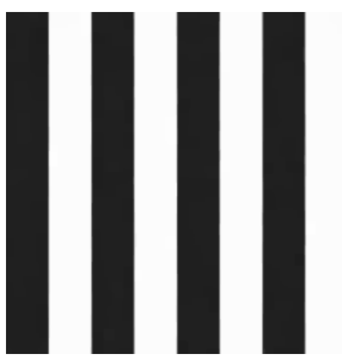
Figure 3.1: Comparison of Algorithm 3 with $\tau(\sigma_1\|D_1\|^2 + \sigma_2\|D_1\|^2 + \sigma_3) = 1$ and `condat` (observation b_4).



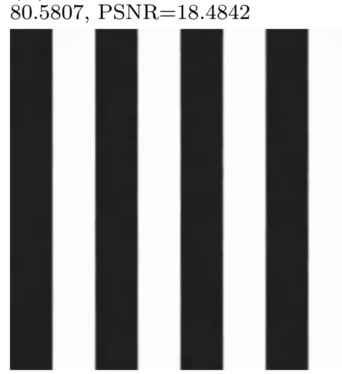
(a) Original, $F^{TV}(\overline{x}) = 17.8430$



(b) Blurry/noisy b_4, $F^{TV}(b) = 80.5807$, PSNR=18.4842



(c) `condat` , $F^{TV}(x_{300}) = 16.3503$, PSNR=28.5957.



(d) Alg. 3, $F^{TV}(x_{300}) = 16.3473$, PSNR=28.6349.

Figure 3.2: Reconstructed image, after 300 iterations, from blurred and noisy image using `condat` and Alg. 3 in their best cases, respectively, cases and $\lambda = 1.9$.

y Programas from UTFSM through Programa de Incentivos a la Iniciación Científica (PIIC).

3.2.6 Appendix

Proposition 3.2.14. *In the context of Problem 3.2.1, set* $L = \mathrm{Id}$*, let* $\Upsilon\colon \mathcal{H} \to \mathcal{H}$ *be a strongly monotone self adjoint linear bounded operator, set* $\Lambda\colon \mathcal{H}\times\mathcal{H} \to \mathcal{H}\colon (x,u) \mapsto x - \Upsilon u$*, let* $\boldsymbol{V}$*,* $\boldsymbol{W}$*, and* $G_{\Upsilon,B,A}$ *be the operators defined in* (3.2.27)*,* (3.2.30)*, and* (3.2.43)*, respectively. Then,* $\Lambda(\mathrm{Fix}\,(P_{\mathrm{ran}\,\boldsymbol{V}} \circ J_{\boldsymbol{W}})) = \mathrm{Fix}\,G_{\Upsilon,B,A}$.

Proof. The inclusion $\subset$ is proved in (3.2.51). Conversely, since $\Lambda^*\colon z \mapsto (z, -\Upsilon z)$, we have $\Lambda\circ\Lambda^* = \mathrm{Id} + \Upsilon^2$ and [3, Proposition 3.30 & Example 3.29] yields $P_{\mathrm{ran}\,\boldsymbol{V}} = P_{\mathrm{ran}\,\Lambda^*} = \Lambda^*(\mathrm{Id}+\Upsilon^2)^{-1}\Lambda$. Therefore, if $\hat{z} \in \mathrm{Fix}\,G_{\Upsilon,B,A}$, by setting $(\hat{x},\hat{u}) := \Lambda^*(\mathrm{Id}+\Upsilon^2)^{-1}\hat{z}$, we have $\hat{z} = \Lambda(\hat{x},\hat{u})$ and we deduce from (3.2.50) that

$$\begin{aligned} P_{\mathrm{ran}\,\boldsymbol{V}} \circ J_{\boldsymbol{W}}(\hat{x},\hat{u}) &= \Lambda^*(\mathrm{Id}+\Upsilon^2)^{-1}\Lambda(J_{\boldsymbol{W}}(\hat{x},\hat{u})) \\ &= \Lambda^*(\mathrm{Id}+\Upsilon^2)^{-1}G_{\Upsilon,B,A}(\Lambda(\hat{x},\hat{u})) \\ &= \Lambda^*(\mathrm{Id}+\Upsilon^2)^{-1}G_{\Upsilon,B,A}\hat{z} \\ &= \Lambda^*(\mathrm{Id}+\Upsilon^2)^{-1}\hat{z} \\ &= (\hat{x},\hat{u}). \end{aligned} \tag{3.2.59}$$

Consequently, $(\hat{x},\hat{u}) \in \mathrm{Fix}\,(P_{\mathrm{ran}\,\boldsymbol{V}} \circ J_{\boldsymbol{W}})$ and $\hat{z} = \Lambda(\hat{x},\hat{u}) \in \Lambda(\mathrm{Fix}\,(P_{\mathrm{ran}\,\boldsymbol{V}} \circ J_{\boldsymbol{W}}))$. ☐

Bibliography

[1] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2009, https://doi.org/10.1007/978-0-8176-4848-0.

[2] H. H. Bauschke, *New demiclosedness principles for (firmly) nonexpansive operators*, in Computational and analytical mathematics, vol. 50 of Springer Proc. Math. Stat., Springer, New York, 2013, pp. 19–28, https://doi.org/10.1007/978-1-4614-7621-4_2.

[3] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5.

[4] H. H. Bauschke and W. M. Moursi, *On the Douglas-Rachford algorithm*, Math. Program., 164 (2017), pp. 263–284, https://doi.org/10.1007/s10107-016-1086-3.

[5] R. I. Boţ, E. R. Csetnek, and A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM J. Optim., 23 (2013), pp. 2011–2036, https://doi.org/10.1137/12088255X.

[6] R. I. Boţ and C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565, https://doi.org/10.1137/120901106.

[7] L. Briceño, R. Cominetti, C. E. Cortés, and F. Martínez, *An integrated behavioral model of land use and transport system: a hyper-network equilibrium approach*, Netw. Spat. Econ., 8 (2008), pp. 201–224, https://doi.org/10.1007/s11067-007-9052-5.

[8] L. Briceño-Arias and F. Roldán, *Primal-dual splittings as fixed point iterations in the range of linear operators*, 2019, https://doi.org/10.48550/ARXIV.1910.02329, https://arxiv.org/abs/1910.02329.

[9] L. M. Briceño-Arias and P. L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250, https://doi.org/10.1137/10081602X.

[10] L. M. Briceño-Arias and P. L. Combettes, *Monotone operator methods for Nash equilibria in non-potential games*, in Computational and analytical mathematics, vol. 50 of Springer Proc. Math. Stat., Springer, New York, 2013, pp. 143–159, https://doi.org/10.1007/978-1-4614-7621-4_9.

[11] L. M. Briceño-Arias and D. Davis, *Forward-backward-half forward algorithm for solving monotone inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871, https://doi.org/10.1137/17M1120099.

[12] L. M. Briceño-Arias and F. Roldán, *Split-Douglas-Rachford algorithm for composite monotone inclusions and split-ADMM*, SIAM J. Optim., 31 (2021), pp. 2987–3013, https://doi.org/10.1137/21M1395144.

[13] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, vol. 9 of Radon Ser. Comput. Appl. Math., Walter de Gruyter, Berlin, 2010, pp. 263–340, https://doi.org/10.1515/9783110226157.263.

[14] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188, https://doi.org/10.1007/s002110050258.

[15] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145, https://doi.org/10.1007/s10851-010-0251-1.

[16] J. Colas, N. Pustelnik, C. Oliver, P. Abry, J.-C. Géminard, and V. Vidal, *Nonlinear denoising for characterization of solid friction under low confinement pressure*, Physical Review E , 42 (2019), p. 91, https://doi.org/10.1103/PhysRevE.100.032803, https://hal.archives-ouvertes.fr/hal-02271333.

[17] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently parallel algorithms in feasibility and optimization and their applications (Haifa, 2000), vol. 8 of Stud. Comput. Math., North-Holland, Amsterdam, 2001, pp. 115–152, https://doi.org/10.1016/S1570-579X(01)80010-0.

[18] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, 53 (2004), pp. 475–504, https://doi.org/10.1080/02331930412331327157.

[19] P. L. Combettes and B. C. Vũ, *Variable metric quasi-Fejér monotonicity*, Nonlinear Anal., 78 (2013), pp. 17–31, https://doi.org/10.1016/j.na.2012.09.008.

[20] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318, https://doi.org/10.1080/02331934.2012.733883.

[21] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479, https://doi.org/10.1007/s10957-012-0245-9.

[22] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, https://doi.org/10.1002/cpa.20042.

[23] D. Davis, *Convergence rate analysis of primal-dual splitting schemes*, SIAM J. Optim., 25 (2015), pp. 1912–1943, https://doi.org/10.1137/151003076.

[24] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318, https://doi.org/10.1007/BF01581204.

[25] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Program., 72 (1996), pp. 1–15, https://doi.org/10.1016/0025-5610(95)00012-7.

[26] D. Gabay, *Chapter ix applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, 1983, pp. 299 – 331, https://doi.org/https://doi.org/10.1016/S0168-2024(08)70034-1.

[27] E. M. Gafni and D. P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM J. Control Optim., 22 (1984), pp. 936–964, https://doi.org/10.1137/0322061.

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

[29] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Math. Soc., 70 (1964), pp. 709–710, https://doi.org/10.1090/S0002-9904-1964-11178-2.

[30] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring images: Matrices, spectra, and filtering*, vol. 3 of Fundamentals of Algorithms, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006, https://doi.org/10.1137/1.9780898718874.

[31] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149, https://doi.org/10.1137/100814494.

[32] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.

[33] S. Mallat, *A wavelet tour of signal processing: The sparse way*, Elsevier/Academic Press, Amsterdam, third ed., 2009.

[34] B. Martinet, *Brève communication. régularisation d'inéquations variationnelles par approximations successives*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, 4 (1970), pp. 154–158, http://www.numdam.org/item/M2AN_1970__4_3_154_0.

[35] R. V. Mises and H. Pollaczek-Geiringer, *Praktische verfahren der gleichungsauflösung* ., ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik, 9 (1929), pp. 152–164, https://doi.org/10.1002/zamm.19290090206.

[36] C. Molinari, J. Peypouquet, and F. Roldan, *Alternating forward-backward splitting for linearly constrained optimization problems*, Optim. Lett., 14 (2020), pp. 1071–1088, https://doi.org/10.1007/s11590-019-01388-y.

[37] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in 2011 International Conference on Computer Vision, Nov 2011, pp. 1762–1769, https://doi.org/10.1109/ICCV.2011.6126441.

[38] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898, https://doi.org/10.1137/0314056.

[39] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268, https://doi.org/10.1016/0167-2789(92)90242-F. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

[40] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1997, https://doi.org/10.1090/surv/049.

[41] B. F. Svaiter, *On weak convergence of the Douglas-Rachford method*, SIAM J. Control Optim., 49 (2011), pp. 280–287, https://doi.org/10.1137/100788100.

[42] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

[43] Y. Yang, Y. Tang, M. Wen, and T. Zeng, *Preconditioned douglas-rachford type primal-dual method for solving composite monotone inclusion problems with applications*, Inverse Probl. Imaging, 15 (2021), pp. 787–825.

Chapter 4

Resolvent of the Parallel Composition and Proximity Operator of the Infimal Postcomposition

4.1 Introduction and Main Results

In this section we aim to calculate the resolvent of *parallel composition* including non-standard metrics under mild assumptions. Given a maximally monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and a linear bounded operator $L\colon \mathcal{H} \to \mathcal{G}$, where $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, the *parallel composition* of A and B, is given by

$$L \rhd A = (LA^{-1}L^*)^{-1}. \tag{4.1.1}$$

This operation arises in primal-dual composite monotone inclusions and related numerical methods available in the literature. For instance, Problem 1.1.2 motivates to derive an explicit computation of the resolvent of $(LA^{-1}-L)^{-1}$ in view of (1.1.4) and the DRS (Algorithm1.1.4). In the particular case when A is the subdifferential of a convex function $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ satisfying dual qualification conditions, we have that $L \rhd A$ is the subdifferential of the *infimal postcomposition* of f by L, defined by

$$L \rhd f\colon \mathcal{G} \to \,]-\infty, +\infty] : u \mapsto \inf_{\substack{x\in\mathcal{H}\\ Lx=u}} f(x). \tag{4.1.2}$$

This operation appears naturally when dealing with the dual of composite optimization problems since we have $(L\rhd f)^* = f^*\circ L^*$ under mild assumptions [1, Proposition 13.24(iv)], see for instance Problem 2.1.5 and Algorithm 2.1.6. Moreover, it is related with the parallel composition via the identities

$$L \rhd (\partial f) = (L(\partial f^*)L^*)^{-1} = (\partial(f^* \circ L^*))^{-1} = \partial(f^* \circ L^*)^* = \partial(L \rhd f), \tag{4.1.3}$$

where the second equality holds if, e.g., $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*)$ [1, Corollary 16.53].

We present a generalization of [1, Proposition 23.25] providing an explicit computation of the resolvent of UM^*BM under mild assumptions. Using this result we obtain the following corollary which allows the computation of the resolvent of the parallel composition $L \rhd A$ under mild assumptions. Recall that J_A^U is defined in [6] and that $U^\dagger$ stands for the Moore-Penrose inverse of U.

Corollary 4.1.1. *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $L\colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator such that $LA^{-1}L^*$ is maximally monotone in $\mathcal{G}$. Moreover, let $U\colon \mathcal{G} \to \mathcal{G}$ be a self-adjoint strongly monotone linear bounded operator. Then, $U(L \rhd A)$ is maximally monotone in $(\mathcal{G}, \langle \cdot \mid \cdot \rangle_{U^{-1}})$ and the following holds:*

1. $J_{U(L \rhd A)} = L(A + L^*U^{-1}L)^{-1}L^*U^{-1}$.
2. *Suppose that* $\operatorname{ran} L$ *is closed. Then,*
$$J_{U(L \rhd A)} = LJ_A^{L^*U^{-1}L}(\sqrt{U}^{-1}L)^\dagger \sqrt{U}^{-1}. \tag{4.1.4}$$
3. *Suppose that* $\operatorname{ran} L^* = \mathcal{H}$. *Then,*
$$J_{U(L \rhd A)} = LJ_{(L^*U^{-1}L)^{-1}A}(L^*U^{-1}L)^{-1}L^*U^{-1}. \tag{4.1.5}$$

Applying previous results in the optimization context, we obtain the following result which provides a formula for the proximity operators of $f^* \circ L^*$ and $L \rhd f$ and generalizes [23, Proposition 5.2(iii)] to non-standard metrics and infinite dimensions. Recall that prox_f^U is defined in (1.3.8).

Proposition 4.1.2. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $f \in \Gamma_0(\mathcal{H})$, let $L\colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator such that*

$$0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*), \tag{4.1.6}$$

and let $U\colon \mathcal{G} \to \mathcal{G}$ be a strongly monotone self-adjoint linear operator. Define

$$\operatorname{prox}_{f,L}^U\colon \mathcal{G} \to 2^{\mathcal{H}}\colon u \mapsto \arg\min_{x \in \mathcal{H}} \left(f(x) + \frac{1}{2}\|Lx - u\|_U^2 \right). \tag{4.1.7}$$

Then, the following hold:

1. $\operatorname{dom} \operatorname{prox}_{f,L}^U = \mathcal{G}$.
2. $\operatorname{prox}_{f^* \circ L^*}^{U^{-1}} = \operatorname{Id} - UL\operatorname{prox}_{f,L}^U U^{-1}$.
3. $L \rhd f = (f^* \circ L^*)^* \in \Gamma_0(\mathcal{H})$ *and* $\operatorname{prox}_{L \rhd f}^U = L\operatorname{prox}_{f,L}^U$.

4.2 Article: Resolvent of the Parallel Composition and Proximity Operator of the Infimal Postcomposition [1]

Abstract In this paper we provide the resolvent computation of the infimal postcomposition of a maximally monotone operator by a linear operator under mild assumptions. Connections with a modification of the warped resolvent are provided. In the context of convex optimization, we obtain the proximity operator of the infimal postcomposition of a convex function by a linear operator and we extend full range conditions on the linear operator to mild qualification conditions. We also introduce a generalization of the proximity operator involving a general linear bounded operator leading to a generalization of Moreau's decomposition for composite convex optimization.

4.2.1 Introduction

In this paper we aim at computing the resolvent of the *parallel composition* of A by L, defined by

$$L \rhd A = (LA^{-1}L^*)^{-1}, \tag{4.2.1}$$

where $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $L\colon \mathcal{H} \to \mathcal{G}$ is linear and bounded. In the case when $\mathcal{H} = H \oplus H$ for some real Hilbert space H, $\mathcal{G} = H$, $A\colon (x, y) \mapsto Bx \times Cx$ for some set-valued operators B and C defined in H and $L\colon (x, y) \mapsto x + y$ we have $L \rhd A = B \,\square\, C$ [1, Example 25.40], where $B \,\square\, C = (B^{-1} + C^{-1})^{-1}$ is the *parallel sum* of B and C, motivating the name of the operation. The parallel composition appears naturally in composite monotone inclusions. Indeed, if $B\colon \mathcal{G} \mapsto 2^{\mathcal{G}}$, the dual inclusion associated to

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + L^*BLx, \tag{4.2.2}$$

is

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad 0 \in B^{-1}u + (-L \rhd A)^{-1}u. \tag{4.2.3}$$

When $L^*L = \alpha\mathrm{Id}$ or when L^* has full range, explicit formulas for the resolvent of $LA^{-1}L^*$ depending on the resolvent of A can be found in [1, Proposition 23.25]. In [13, 25] some variants and fixed point methods to compute the resolvent are proposed under full range condition on L^* and a similar fixed point approach is used in [21] under the maximal monotonicity of $LA^{-1}L^*$. This computation is useful in [22] for the equivalence between the primal-dual [7, 24] and Douglas-Rachford splitting (DRS) [11, 17] algorithms.

In the particular case when A is the subdifferential of a convex function $f\colon \mathcal{H} \to]-\infty, +\infty]$ satisfying dual qualification conditions, we have that $L \rhd A$ is the subdifferential

[1] [4] Luis M. Briceño-Arias and Fernando Roldán. Resolvent of the parallel composition and proximity operator of the infimal postcomposition, 2021, https://arxiv.org/abs/2109.06771.

of the *infimal postcomposition* of f by L, defined by

$$L \rhd f \colon \mathcal{G} \to [-\infty, +\infty] : u \mapsto \inf_{\substack{x \in \mathcal{H} \\ Lx = u}} f(x). \tag{4.2.4}$$

This operation appears naturally when dealing with the dual of composite optimization problems since we have $(L \rhd f)^* = f^* \circ L^*$ under mild assumptions [1, Proposition 13.24(iv)]. Moreover, it is related with the parallel composition via the identities

$$L \rhd (\partial f) = (L(\partial f^*)L^*)^{-1} = (\partial(f^* \circ L^*))^{-1} = \partial(f^* \circ L^*)^* = \partial(L \rhd f), \tag{4.2.5}$$

where the second equality holds if, e.g., $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*)$ [1, Corollary 16.53]. Therefore, the resolvent of $L \rhd (\partial f)$ and the proximity operator of $L \rhd f$ are related and they are useful in the derivation of the *alternating direction method of multipliers* (ADMM) for solving $\inf(f + g \circ L)$ from DRS, since the former is obtained as an application of the latter to the Fenchel-Rockafellar dual $\inf(f^* \circ (-L^*) + g^*)$ [14] (see also [3, 9, 10, 23, 5]). In this context, the single-valuedness of $(\partial f + L^*L)^{-1}L^*$ is assumed in the proof of [9, Theorem 4.7], its full domain is supposed in [23, Proposition 5.2], and the strong monotonicity of $(\partial f + L^*L)$ is assumed in [3, 10] in order obtain both properties. It is worth to notice that some fixed point approaches and algorithms for computing $\operatorname{prox}_{f^* \circ L^*}$ are proposed in [12, 18] in the context of sparse recovery in image processing.

In this paper we derive a formula for the resolvent of the parallel composition and for the proximity operator of the infimal postcomposition in a real Hilbert space with non-standard metric under mild assumptions. This is obtained from a formula of the resolvent of $LA^{-1}L^*$ via the non-standard metric version of Moreau's identity in [1, Proposition 23.34(iii)]. Our computation is related with a modification of the warped resolvent defined in [6] (see [15] for a particular case) and we extend and generalize several results in the literature as [3, 9, 10, 23, 25]). We also derive a generalization of Moreau's decomposition [20] for composite maximally monotone operators and for composite convex optimization under standard assumptions by using a generalization of the proximity operator.

4.2.2 Notation and preliminaries

Throughout this paper $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces with the scalar product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\|$. The identity operator on $\mathcal{H}$ is denoted by Id. Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain of A is $\operatorname{dom} A = \{x \in \mathcal{H} \mid Ax \neq \varnothing\}$, the range of A is $\operatorname{ran} A = \{u \in \mathcal{H} \mid (\exists x \in \mathcal{H})\ u \in Ax\}$, the graph of A is $\operatorname{gra} A = \{(x, u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax\}$, the set of zeros of A is $\operatorname{zer} A = \{x \in \mathcal{H} \mid 0 \in Ax\}$, the inverse of A is $A^{-1} \colon u \mapsto \{x \in \mathcal{H} \mid u \in Ax\}$, and its resolvent is $J_A = (\operatorname{Id} + A)^{-1}$. For every $D \subset \mathcal{H}$, $A|_D$ is the restriction of A to D, which satisfies $\operatorname{dom} A|_D = \operatorname{dom} A \cap D$ and, for every $x \in D$, $A|_D\, x = Ax$. The operator A is *injective on D* if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad Ax \cap Ay \cap D \neq \varnothing \quad \Rightarrow \quad x = y, \tag{4.2.6}$$

and A is injective if it is injective on $\mathcal{H}$. It is clear that injectivity of A on D implies its injectivity on D' when $D' \subset D$. Moreover, the operator A is monotone if

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v \rangle \geq 0, \tag{4.2.7}$$

A is strongly monotone if there exists $\alpha > 0$ such that

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v \rangle \geq \alpha \|x - y\|^2, \tag{4.2.8}$$

and A is maximally monotone if it is monotone and, for every $(x,u) \in \mathcal{H} \times \mathcal{H}$,

$$(x,u) \in \operatorname{gra} A \quad \Leftrightarrow \quad (\forall (y,v) \in \operatorname{gra} A) \;\; \langle x - y \mid u - v \rangle \geq 0. \tag{4.2.9}$$

For every strongly monotone self-adjoint linear bounded operator $U\colon \mathcal{H} \to \mathcal{G}$, we denote $\langle \cdot \mid \cdot \rangle_U = \langle \cdot \mid U \cdot \rangle$ and $\| \cdot \|_U = \sqrt{\langle \cdot \mid \cdot \rangle_U}$, which define an inner product and the associated norm in $\mathcal{H}$, respectively.

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is defined by $f^*\colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$, $f^* \in \Gamma_0(\mathcal{H})$, the subdifferential of f is the maximally monotone operator

$$\partial f\colon x \mapsto \big\{ u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H}) \;\; f(x) + \langle y - x \mid u \rangle \leq f(y) \big\}, \tag{4.2.10}$$

$(\partial f)^{-1} = \partial f^*$, the set of minimizers of f is denoted by $\arg\min_{x\in\mathcal{H}} f(x)$, and we have that $\operatorname{zer}(\partial f) = \arg\min_{x\in\mathcal{H}} f(x)$. Given a strongly monotone self-adjoint linear operator $U\colon \mathcal{H} \to \mathcal{H}$, we denote by

$$\operatorname{prox}_f^U\colon x \mapsto \arg\min_{y\in\mathcal{H}} \big(f(y) + \frac{1}{2}\|x - y\|_U^2 \big), \tag{4.2.11}$$

and by $\operatorname{prox}_f = \operatorname{prox}_f^{\mathrm{Id}}$. We have [1, Proposition 24.24] (see also [8, Section 3])

$$\operatorname{prox}_f^U = U^{-\frac{1}{2}} \operatorname{prox}_{f\circ U^{-\frac{1}{2}}} U^{\frac{1}{2}} = J_{U^{-1}\partial f} \tag{4.2.12}$$

and it is single valued since the objective function in (4.2.11) is strongly convex. Moreover, it follows from [1, Proposition 23.34(iii)] that

$$J_{UA} + U J_{U^{-1}A^{-1}} U^{-1} = \mathrm{Id}, \tag{4.2.13}$$

and, in the case of convex functions, [1, Proposition 24.24] yields

$$\operatorname{prox}_f^U = \mathrm{Id} - U^{-1} \operatorname{prox}_{f^*}^{U^{-1}} U = U^{-1} \big(\mathrm{Id} - \operatorname{prox}_{f^*}^{U^{-1}}\big) U. \tag{4.2.14}$$

Given a non-empty set $C \subset \mathcal{H}$, we denote by $\overline{\operatorname{span}}\, C$ the closed span of C, by $\operatorname{cone} C$ its conical hull. Let C be a non-empty closed convex subset of $\mathcal{H}$. We denote by $\operatorname{sri} C = \big\{x \in C \;\big|\; \operatorname{cone}(C - x) = \overline{\operatorname{span}}\,(C - x)\big\}$ its strong relative interior, by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise, by $P_C^U = \operatorname{prox}_{\iota_C}^U$ the projection onto C with respect to $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_U)$, and we denote $P_C = P_C^{\mathrm{Id}}$. It follows from (4.2.12) that

$$P_C^U = U^{-\frac{1}{2}} \operatorname{prox}_{\iota_C \circ U^{-\frac{1}{2}}} U^{\frac{1}{2}} = U^{-\frac{1}{2}} P_{U^{\frac{1}{2}} C} U^{\frac{1}{2}}. \tag{4.2.15}$$

Given a linear bounded operator $L : \mathcal{H} \to \mathcal{G}$, we denote its adjoint by $L^* \colon \mathcal{G} \to \mathcal{H}$, its kernel (or null space) by $\ker L$, its range by $\operatorname{ran} L$, and, if $\operatorname{ran} L$ is closed, its *Moore-Penrose inverse* by

$$L^\dagger \colon \mathcal{G} \to \mathcal{H} \colon y \mapsto P_{C_y} 0, \tag{4.2.16}$$

where $C_y = \{x \in \mathcal{H} \mid L^* L x = L^* y\}$. If $L^* L$ is invertible, we have [1, Example 3.29]

$$L^\dagger = (L^* L)^{-1} L^*. \tag{4.2.17}$$

For definitions and properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [1].

We now introduce a modification of the warped resolvent introduced in [6] (see also [15] for a particular case and applications). Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator and let $K \colon \mathcal{H} \to \mathcal{H}$. The *warped resolvent of A with kernel K* is defined by $J_A^K = (K + A)^{-1} K$. In the case when K is linear and invertible, we have

$$J_A^K = (K + A)^{-1} K = (K(\mathrm{Id} + K^{-1} A))^{-1} K = J_{K^{-1} A}, \tag{4.2.18}$$

which has full domain and it is single valued if $K^{-1} A$ is maximally monotone. The following result characterizes the full domain and single-valuedness of J_A^K in a general context.

Proposition 4.2.1. *Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator and let $K \colon \mathcal{H} \to \mathcal{H}$. Then the following holds.*

1. $\operatorname{dom} J_A^K = \mathcal{H} \;\Leftrightarrow\; \operatorname{ran} K \subset \operatorname{ran}(K + A)$.
2. *J_A^K is at most single-valued* $\;\Leftrightarrow\;$ *$K + A$ is injective on* $\operatorname{ran} K$.

Proof. 1: For every $x \in \mathcal{H}$ we have

$$\begin{aligned} x \in \operatorname{dom} J_A^K \quad &\Leftrightarrow \quad (\exists u \in \mathcal{H}) \quad u \in (K + A)^{-1} K x \\ &\Leftrightarrow \quad (\exists u \in \mathcal{H}) \quad K x \in (K + A) u \\ &\Leftrightarrow \quad K x \in \operatorname{ran}(K + A), \end{aligned} \tag{4.2.19}$$

and the result follows. 2: First assume that J_A^K is at most single valued. In view of (4.2.6), let x and y in $\mathcal{H}$ and suppose that there exists $z \in \mathcal{H}$ such that $Kz \in (Kx + Ax) \cap (Ky + Ay)$. Then $\{x\} \cup \{y\} \subset J_A^K z$ and single-valuedness of J_A^K implies $x = y$, which yields the injectivity on $\operatorname{ran} K$. Conversely, let $z \in \operatorname{dom} J_A^K$ and let x and y in $J_A^K z$. Then, $Kz \in (Kx + Ax) \cap (Ky + Ay) \cap \operatorname{ran} K$ and injectivity on $\operatorname{ran} K$ implies $x = y$. $\square$

In [6, Definition 1.1] it is assumed that $K + A$ is injective in the whole space in order to guarantee that J_A^K is single-valued, but this is a stronger assumption in general, as the following example illustrates.

Example 4.2.2. *Let $\alpha > 0$, set $\mathcal{H} = \mathbb{R}$, set $K\colon x \mapsto \operatorname{mid}\{-1, x, 1\}$ be the median of real values x, -1, and 1, and set $A = \alpha K$. Note that A and K are maximally monotone, single valued, and $\operatorname{ran} K = [-1, 1] \subset [-1-\alpha, 1+\alpha] = \operatorname{ran}(K + A)$. Moreover, observe that $K + A = (1+\alpha)\operatorname{mid}\{-1, x, 1\}$ is injective on $\operatorname{ran} K$ but it is not injective on $\mathbb{R}$, since $(K + A)1 = (K + A)2 = 1 + \alpha$.*

The *warped proximity operator of f with kernel K* is defined by

$$\operatorname{prox}_f^K = J_{\partial f}^K = (K + \partial f)^{-1} K \tag{4.2.20}$$

and note that it coincides with (4.2.12) when K is strongly monotone, self-adjoint, linear, and bounded, in view of (4.2.18).

4.2.3 Resolvent of parallel composition

The following result is a generalization of [1, Proposition 23.25] and provides an explicit computation of the resolvent of UM^*BM under mild assumptions.

Theorem 4.2.3. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $M\colon \mathcal{G} \to \mathcal{H}$ be a linear bounded operator such that M^*BM is maximally monotone in $\mathcal{G}$, and let $U\colon \mathcal{G} \to \mathcal{G}$ be a $\mu-$strongly monotone self-adjoint linear operator for some $\mu > 0$. Then UM^*BM is maximally monotone in $(\mathcal{G}, \langle \cdot \mid \cdot \rangle_{U^{-1}})$ and the following assertions hold:*

1. *$\operatorname{ran} M \subset \operatorname{dom}(MUM^* + B^{-1})^{-1}$ and*
$$J_{UM^*BM} = \operatorname{Id} - UM^*(MUM^* + B^{-1})^{-1}M. \tag{4.2.21}$$
2. *$\operatorname{ran}(MUM^*) \subset \operatorname{ran}(MUM^* + B^{-1})$.*
3. *$(MUM^* + B^{-1})\,|_{\operatorname{ran} M}$ is injective.*
4. *Suppose that $\operatorname{ran} M$ is closed. Then*
$$J_{UM^*BM} = \operatorname{Id} - UM^* J_{B^{-1}}^{MUM^*}(\sqrt{U}M^*)^{\dagger}\sqrt{U}^{-1}. \tag{4.2.22}$$
5. *Suppose that $\operatorname{ran} M = \mathcal{H}$. Then*
$$\begin{aligned} J_{UM^*BM} &= \operatorname{Id} - UM^* J_{(MUM^*)^{-1}B^{-1}}(MUM^*)^{-1}M & (4.2.23)\\ &= P_{\ker M}^{U^{-1}} + UM^*(MUM^*)^{-1}J_{MUM^*B}M. & (4.2.24)\end{aligned}$$

Proof. The maximal monotonicity of UM^*BM follows from [1, Proposition 20.24]. 1: For every x and p in $\mathcal{G}$, we have

$$
\begin{aligned}
p = J_{UM^*BM}x \quad &\Leftrightarrow \quad x - p \in UM^*BMp && (4.2.25)\\
&\Leftrightarrow \quad (\exists v \in \mathcal{H}) \begin{cases} x - p = UM^*v \\ v \in BMp \end{cases}\\
&\Leftrightarrow \quad (\exists v \in \mathcal{H}) \begin{cases} p = x - UM^*v \\ Mp \in B^{-1}v \end{cases}\\
&\Leftrightarrow \quad (\exists v \in \mathcal{H}) \begin{cases} p = x - UM^*v \\ Mx \in MUM^*v + B^{-1}v. \end{cases}\\
&\Leftrightarrow \quad (\exists v \in (MUM^* + B^{-1})^{-1}Mx) \quad p = x - UM^*v, && (4.2.26)
\end{aligned}
$$

and the result follows. 2: It follows from 1 that

$$
\operatorname{ran}(MUM^*) \subset \operatorname{ran} M \subset \operatorname{dom}(MUM^* + B^{-1})^{-1} = \operatorname{ran}(MUM^* + B^{-1}). \tag{4.2.27}
$$

3: Let x and y in $\operatorname{ran} M$ be such that there exists $u \in (MUM^*x + B^{-1}x) \cap (MUM^*y + B^{-1}y)$. Then, $u - MUM^*x \in B^{-1}x$, $u - MUM^*y \in B^{-1}y$, and the monotonicity of B^{-1} yields

$$
\begin{aligned}
0 &\leq \langle -MUM^*(x-y) \mid x-y \rangle \\
&= -\langle UM^*(x-y) \mid M^*(x-y) \rangle \\
&\leq -\mu \|M^*(x-y)\|^2, && (4.2.28)
\end{aligned}
$$

which implies $x - y \in \ker M^*$. Since $x - y \in \operatorname{ran} M \subset \overline{\operatorname{ran} M}$, it follows from [1, Fact 2.25(iv)] that $x - y \in \ker M^* \cap \overline{\operatorname{ran} M} = \{0\}$, which yields the result.

4: Denote by $\mathcal{G}_U$ the Hilbert space $\mathcal{G}$ endowed with the scalar product $\langle \cdot \mid \cdot \rangle_{U^{-1}}$. Note that $M^{*_U} = UM^*U^{-1}$, where M^{*_U} and M^* are the adjoints of M in $\mathcal{G}_U$ and $\mathcal{G}$, respectively. Moreover, [1, Fact 2.25(iv)] and the closedness of $\operatorname{ran} M$ on $\mathcal{G}_U$ yield $(\ker M)^{\perp_U} = \operatorname{ran} M^{*_U} = \operatorname{ran}(UM^*U^{-1}) = \operatorname{ran}(UM^*)$, where $\perp_U$ stands for the orthogonal complement in $\mathcal{G}_U$. Hence,

$$
\mathcal{G}_U = \ker M \oplus \operatorname{ran}(UM^*) \tag{4.2.29}
$$

is an orthogonal decomposition of $\mathcal{G}_U$. Hence, we have from [1, Proposition 24.24(ii) & Proposition 3.30(iii)] that

$$
\begin{aligned}
P^{U^{-1}}_{\ker M} &= \operatorname{prox}^{U^{-1}}_{\iota_{\ker M}} \\
&= \sqrt{U}\operatorname{prox}_{\iota_{\ker M} \circ \sqrt{U}} \sqrt{U}^{-1} \\
&= \sqrt{U} P_{\ker(M\sqrt{U})} \sqrt{U}^{-1} \\
&= \operatorname{Id} - UM^*(\sqrt{U}M^*)^{\dagger} \sqrt{U}^{-1}, && (4.2.30)
\end{aligned}
$$

and

$$P^{U^{-1}}_{\operatorname{ran}(UM^*)} = UM^*(\sqrt{U}M^*)^\dagger \sqrt{U}^{-1}, \tag{4.2.31}$$

where $(\sqrt{U}M^*)^\dagger$ is the Moore-Penrose inverse of $\sqrt{U}M^*\colon \mathcal{H} \to \mathcal{G}$. Therefore, 1 asserts that

$$\begin{aligned} J_{UM^*BM} &= \mathrm{Id} - UM^*(B^{-1} + MUM^*)^{-1}M \\ &= \mathrm{Id} - UM^*(B^{-1} + MUM^*)^{-1}MP^{U^{-1}}_{\operatorname{ran}(UM^*)} \\ &= \mathrm{Id} - UM^*(B^{-1} + MUM^*)^{-1}MUM^*(\sqrt{U}M^*)^\dagger \sqrt{U}^{-1} \\ &= \mathrm{Id} - UM^* J^{MUM^*}_{B^{-1}}(\sqrt{U}M^*)^\dagger \sqrt{U}^{-1}, \end{aligned} \tag{4.2.32}$$

where in the last equality $J^{MUM^*}_{B^{-1}}$ has full domain in view of 2 and Proposition 4.2.1(1).

5: Since $\operatorname{ran} M = \mathcal{H}$ is closed and U is $\mu-$strongly monotone for some $\mu > 0$, MUM^* is strongly monotone and, thus, invertible. Indeed, for every $v \in \mathcal{H}$, [1, Fact 2.26] implies that there exists $\alpha > 0$ such that

$$\langle MUM^*v \mid v\rangle = \langle UM^*v \mid M^*v\rangle \geq \mu\|M^*v\|^2 \geq \mu\alpha^2\|v\|^2. \tag{4.2.33}$$

Hence, since $(\sqrt{U}M^*)^*(\sqrt{U}M^*) = MUM^*$, (4.2.23) follows from 4, (4.2.18), and (4.2.17). Moreover, since (4.2.30) and (4.2.17) yield $P^{U^{-1}}_{\ker M} = \mathrm{Id} - UM^*(MUM^*)^{-1}M$, (4.2.24) follows from (4.2.23) and (4.2.13). □

Remark 4.2.4. 1. *Note that Theorem 4.2.3(1) provides the existence of zeros of the monotone operator $MUM^* + B^{-1}$ from the maximal monotonicity of M^*BM, which is guaranteed, e.g., if* $\operatorname{cone}(\operatorname{ran} M - \operatorname{dom} B) = \overline{\operatorname{span}}(\operatorname{ran} M - \operatorname{dom} B)$ *[1, Corollary 25.6] (see [2] for a weaker assumption involving the domain of the Fitzpatrick function).*

2. *Note that, from Theorem 4.2.3(1), $M^*(MUM^* + B^{-1})^{-1}M\colon \mathcal{G} \to \mathcal{G}$ is single valued, even if $(MUM^*+B^{-1})^{-1}$ can be a set-valued mapping. Indeed, for every $x \in \mathcal{G}$, let v and w in $(MUM^*+B^{-1})^{-1}Mx$. Then, $M(x-UM^*v) \in B^{-1}v$ and $M(x-UM^*w) \in B^{-1}w$ and the monotonicity of B^{-1} yields*

$$0 \leq \langle -MU(M^*v - M^*w) \mid v - w\rangle = -\|M^*v - M^*w\|^2_U, \tag{4.2.34}$$

which implies $M^*v = M^*w$. This computation is consistent with the fact that the resolvent of the monotone operator UM^*BM is single-valued.

3. *Observe that Theorem 4.2.3(2) and Proposition 4.2.1(1) imply that* $\operatorname{dom} J^{MUM^*}_{B^{-1}} = \mathcal{H}$. *On the other hand, the single-valuedness of $J^{MUM^*}_{B^{-1}}$ is not guaranteed since MUM^*+B^{-1} is not necessarily injective on* $\operatorname{ran}(MUM^*)$ *(see Proposition 4.2.1(2)). Indeed, suppose that* $\ker M^* \neq \{0\}$ *and that $B^{-1} = N_C$, where C is the closed ball*

centered at 0 *with radius* 1. *By taking* $x = 0$ *and* $y \in \ker M^* \smallsetminus \{0\} \cap \operatorname{int} C$, *we have* $\{0\} = N_C x \cap N_C y = (MUM^*x + B^{-1}x) \cap (MUM^*y + B^{-1}y)$, $0 \in \operatorname{ran} MUM^*$, *and* $x \neq y$. *However, when* $\operatorname{ran} M$ *is closed, it follows from Theorem 4.2.3(3) and* $\operatorname{ran}(\sqrt{U}M^*)^\dagger = \operatorname{ran}(M\sqrt{U}) = \operatorname{ran} M$ *[1, Proposition 3.30(v)] that* $J_{B^{-1}}^{MUM^*}(\sqrt{U}M^*)^\dagger$ *is single valued.*

4. *In the particular case when* $U = \operatorname{Id}$, *Theorem 4.2.3(5) coincides with [1, Proposition 23.25]. On the other hand, when* $M = \operatorname{Id}$ *and* $\mathcal{H} = \mathcal{G}$, *we recover from Theorem 4.2.3(5) the Moreau's decomposition with non-standard metric in [1, Proposition 23.34(iii)] recalled in (4.2.13).*

We conclude this section with the computation of the resolvent of the parallel composition $L \rhd A$.

Corollary 4.2.5. *Let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be a maximally monotone operator, let* $L\colon \mathcal{H} \to \mathcal{G}$ *be a linear bounded operator such that* $LA^{-1}L^*$ *is maximally monotone in* $\mathcal{G}$. *Moreover, let* $U\colon \mathcal{G} \to \mathcal{G}$ *be a self-adjoint strongly monotone linear bounded operator. Then,* $U(L \rhd A)$ *is maximally monotone in* $(\mathcal{G}, \langle \cdot \mid \cdot \rangle_{U^{-1}})$ *and the following holds:*

1. $J_{U(L\rhd A)} = L(A + L^*U^{-1}L)^{-1}L^*U^{-1}$.

2. *Suppose that* $\operatorname{ran} L$ *is closed. Then,*
$$J_{U(L\rhd A)} = LJ_A^{L^*U^{-1}L}(\sqrt{U}^{-1}L)^\dagger\sqrt{U}^{-1}. \tag{4.2.35}$$

3. *Suppose that* $\operatorname{ran} L^* = \mathcal{H}$. *Then,*
$$J_{U(L\rhd A)} = LJ_{(L^*U^{-1}L)^{-1}A}(L^*U^{-1}L)^{-1}L^*U^{-1}. \tag{4.2.36}$$

Proof. Since $L \rhd A = (LA^{-1}L^*)^{-1}$, the maximal monotonicity of $U(L \rhd A)$ follows from [1, Propositions 20.22 & 20.24]. 1: By applying Theorem 4.2.3(1) to $B = A^{-1}$ and $M = L^*$, it follows from (4.2.13) that
$$\begin{aligned} J_{U(L\rhd A)} &= \operatorname{Id} - UJ_{U^{-1}LA^{-1}L^*}U^{-1} \\ &= \operatorname{Id} - U(\operatorname{Id} - U^{-1}L(A + L^*U^{-1}L)^{-1}L^*)U^{-1} \\ &= L(A + L^*U^{-1}L)^{-1}L^*U^{-1}. \end{aligned} \tag{4.2.37}$$
2: By applying Theorem 4.2.3(4) to $B = A^{-1}$ and $M = L^*$, we obtain
$$\begin{aligned} J_{U(L\rhd A)} &= \operatorname{Id} - UJ_{U^{-1}LA^{-1}L^*}U^{-1} \\ &= \operatorname{Id} - U\big(\operatorname{Id} - U^{-1}LJ_A^{L^*U^{-1}L}(\sqrt{U}^{-1}L)^\dagger\sqrt{U}\big)U^{-1} \\ &= LJ_A^{L^*U^{-1}L}(\sqrt{U}^{-1}L)^\dagger\sqrt{U}^{-1}. \end{aligned} \tag{4.2.38}$$
3: As in the proof of Theorem 4.2.3(4), $L^*U^{-1}L$ is strongly monotone and, hence, invertible, and the result follows from 2, (4.2.18), and (4.2.17). ☐

4.2.4 Proximity operator of the infimal postcomposition

For every $f \in \Gamma_0(\mathcal{H})$, every linear bounded operator $L\colon \mathcal{H} \to \mathcal{G}$, and every strongly monotone self-adjoint linear bounded operator $U\colon \mathcal{G} \to \mathcal{G}$, define

$$\operatorname{prox}^U_{f,L}\colon \mathcal{G} \to 2^{\mathcal{H}}\colon u \mapsto \arg\min_{x\in\mathcal{H}} \left(f(x) + \frac{1}{2}\|Lx-u\|_U^2 \right). \tag{4.2.39}$$

Note that [1, Theorem 16.3 & Theorem 16.47(i)] yield

$$\begin{aligned}(\forall u \in \mathcal{G})(\forall x \in \mathcal{H}) \quad x \in \operatorname{prox}^U_{f,L}u \quad &\Leftrightarrow \quad 0 \in \partial f(x) + L^*U(Lx-u)\\ &\Leftrightarrow \quad x \in (\partial f + L^*UL)^{-1}L^*Uu. \end{aligned}\tag{4.2.40}$$

When $L = \mathrm{Id}$, we have $\operatorname{prox}^U_{f,\mathrm{Id}} = \operatorname{prox}^U_f$ and it is single valued with full domain. In [16] an extension of definition of the classical proximity operator is studied by considering a Bregman distance instead of $\|\cdot\|_U^2$, under the assumption of uniqueness of the solution to the optimization problem in (4.2.39). In our context, the single-valuedness of $\operatorname{prox}^U_{f,L}$ is not needed. The following result provides some properties of $\operatorname{prox}^U_{f,L}$ in more general contexts.

Proposition 4.2.6. *Let $f \in \Gamma_0(\mathcal{H})$, let $L\colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator, and let $U\colon \mathcal{G} \to \mathcal{G}$ be a $\mu-$strongly monotone self-adjoint linear bounded operator. Then, the following hold:*

1. *For every $u \in \operatorname{dom}\operatorname{prox}^U_{f,L}$, $L(\operatorname{prox}^U_{f,L}u)$, and $P_{(\ker L)^\perp}(\operatorname{prox}^U_{f,L}u)$ are singletons.*
2. *Suppose that $\ker L = \{0\}$. Then, for every $u \in \operatorname{dom}\operatorname{prox}^U_{f,L}$, $\operatorname{prox}^U_{f,L}u$ is a singleton.*
3. *Suppose that $\operatorname{ran} L$ is closed. Then*
$$(\forall u \in \operatorname{dom}\operatorname{prox}^U_{f,L}) \quad \operatorname{prox}^U_{f,L}u = \operatorname{prox}^{L^*UL}_f(\sqrt{U}L)^\dagger\sqrt{U}u. \tag{4.2.41}$$
4. *Suppose that $\operatorname{ran} L^* = \mathcal{H}$. Then $\operatorname{prox}^U_{f,L}$ is single valued, $\operatorname{dom}\operatorname{prox}^U_{f,L} = \mathcal{G}$, and*
$$(\forall u \in \mathcal{G}) \quad \operatorname{prox}^U_{f,L}u = \big\{\operatorname{prox}^{L^*UL}_f(L^*UL)^{-1}L^*Uu\big\}. \tag{4.2.42}$$

Proof. 1: Let x_1 and x_2 in $\operatorname{prox}^U_{f,L}u$. It follows from (4.2.40) applied to x_1 and x_2, the monotonicity of ∂f, and strong monotonicity of U that

$$\begin{aligned}0 &\le \langle -L^*UL(x_1-x_2) \mid x_1-x_2\rangle\\ &= -\langle UL(x_1-x_2) \mid L(x_1-x_2)\rangle\\ &\le -\mu\|L(x_1-x_2)\|^2. \end{aligned}\tag{4.2.43}$$

Therefore, $L(x_1 - x_2) = 0$ which leads to $x_1 - x_2 \in \ker L$ and, hence, $P_{(\ker L)^\perp} x_1 = P_{(\ker L)^\perp} x_2$.

2: In this case $(\ker L)^\perp = \mathcal{H}$, which yields $P_{(\ker L)^\perp} = \mathrm{Id}$ and the result follows from 1.

3: It follows from (4.2.40), the orthogonal decomposition in $(\mathcal{G}, \langle \cdot \mid \cdot \rangle_{U^{-1}})$ in (4.2.29) and (4.2.31) with $M = L^*$ that, for every $u \in \mathcal{G}$ and $x \in \mathcal{H}$,

$$\begin{aligned} x \in \operatorname{prox}^U_{f,L} u \quad &\Leftrightarrow \quad x \in (\partial f + L^* U L)^{-1} L^* P^{U^{-1}}_{\operatorname{ran}(UL)} U u \\ &\Leftrightarrow \quad x \in (\partial f + L^* U L)^{-1} L^* \big(UL(\sqrt{U}L)^\dagger \sqrt{U}^{-1} \big) U u \\ &\Leftrightarrow \quad x \in \operatorname{prox}^{L^* U L}_f \big((\sqrt{U}L)^\dagger \sqrt{U} u \big), \end{aligned} \tag{4.2.44}$$

where the last equivalence follows from (4.2.20).

4: Note that $\operatorname{ran} L^* = \mathcal{H}$ yields, for every $x \in \mathcal{G}$, $\langle L^* U L x \mid x \rangle \geq \mu \|Lx\|^2 \geq \mu \alpha^2 \|x\|^2$, where the existence of $\alpha > 0$ is guaranteed by [1, Fact 2.26]. Therefore, $L^* U L$ is strongly monotone and, hence, invertible. Hence, the result follows from 3 and $(\sqrt{U}L)^\dagger = (L^* U L)^{-1} L^* \sqrt{U}$ in view of (4.2.17). ☐

Note that in Proposition 4.2.6(2), $\operatorname{prox}^U_{f,L} u$ may be empty for some $u \in \mathcal{G}$, as the following examples illustrate.

Example 4.2.7. *Suppose that $U = \mathrm{Id}$, that $\operatorname{ran} L$ is not closed, set $f = 0$, and let $u \in \overline{\operatorname{ran} L} \setminus \operatorname{ran} L$. Then, $\inf_{x \in \mathcal{H}} \|Lx - u\| = 0$ but the minimum is not attained. Observe that, since $f^* = \iota_{\{0\}}$, we have $\operatorname{dom} f^* = \{0\}$ which yields $\operatorname{cone}(\operatorname{dom} f^* - \operatorname{ran} L^*) = \operatorname{cone}(\operatorname{ran} L^*) = \operatorname{ran} L^* \neq \overline{\operatorname{ran} L^*} = \overline{\operatorname{span}} \operatorname{ran} L^*$ and, thus, $0 \notin \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*)$.*

Example 4.2.8. *Suppose that $\mathcal{H} = \mathbb{R}^2$, $\mathcal{G} = \mathbb{R}$, $f \colon (x,y) \mapsto \exp(y)$, and $L \colon (x,y) \mapsto x$. Then $L^* \colon z \mapsto (z, 0)$, $\operatorname{ran} L^* = \mathbb{R} \times \{0\}$, and $f^* \colon (u,v) \mapsto \iota_{\{0\}}(u) + \exp^*(v)$, where*

$$\exp^* \colon v \mapsto \begin{cases} v(\ln v - 1), & \text{if } v > 0; \\ 0, & \text{if } v = 0; \\ +\infty, & \text{if } v < 0. \end{cases} \tag{4.2.45}$$

Then, $\operatorname{dom} f^* = \{0\} \times [0, +\infty[$ and $\operatorname{cone}(\operatorname{dom} f^* - \operatorname{ran} L^*) = \mathbb{R} \times [0, +\infty[\neq \mathbb{R}^2 = \overline{\operatorname{span}}(\operatorname{dom} f^* - \operatorname{ran} L^*)$, which yields $0 \notin \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*)$.

The following result provides sufficient conditions ensuring full domain of $\operatorname{prox}^U_{f,L}$. This is a a consequence of Theorem 4.2.3 in the optimization context and we connect the existence result with the computation of the proximity operators of $f^* \circ L^*$ and $L \rhd f$. Our result generalizes [23, Proposition 5.2(iii)] to non-standard metrics and infinite dimensions.

Proposition 4.2.9. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $f \in \Gamma_0(\mathcal{H})$, let $L \colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator such that*

$$0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*), \tag{4.2.46}$$

and let $U\colon \mathcal{G} \to \mathcal{G}$ be a strongly monotone self-adjoint linear operator. Then, the following hold:

1. $\operatorname{dom} \operatorname{prox}^U_{f,L} = \mathcal{G}$.
2. $\operatorname{prox}^{U^{-1}}_{f^*\circ L^*} = \mathrm{Id} - UL \operatorname{prox}^U_{f,L} U^{-1}$.
3. $L \rhd f = (f^* \circ L^*)^* \in \Gamma_0(\mathcal{H})$ *and* $\operatorname{prox}^U_{L\rhd f} = L \operatorname{prox}^U_{f,L}$.

Proof. 1: Since $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*)$, [1, Corollary 16.53(i)] yields $\partial(f^* \circ L^*) = L(\partial f^*)L^*$, which is maximally monotone in $\mathcal{H}$ because $f^* \circ L^* \in \Gamma_0(\mathcal{H})$ [1, Theorem 20.25]. Hence, by applying Theorem 4.2.3(1) to $B = \partial f^*$ and $M = L^*$, it follows from (4.2.40) that

$$
\begin{aligned}
(\forall x \in \mathcal{H}) \quad \varnothing \neq ((\partial f^*)^{-1} + L^*UL)^{-1}L^*Ux &= (\partial f + L^*UL)^{-1}L^*Ux \\
&= \operatorname{prox}^U_{f,L} x. \qquad (4.2.47)
\end{aligned}
$$

2: We deduce from (4.2.12), Theorem 4.2.3(1), and (4.2.47) that

$$
\operatorname{prox}^{U^{-1}}_{f^*\circ L^*} = J_{U\partial(f^*\circ L^*)} = J_{UL(\partial f^*)L^*} = \mathrm{Id} - UL((\partial f^*)^{-1} + L^*UL)^{-1}L^* = \mathrm{Id} - UL \operatorname{prox}^U_{f,L} U^{-1}.
$$

3: Since $f^* \circ L^* \in \Gamma_0(\mathcal{H})$, (4.2.46) and [1, Corollary 15.28] yield $L \rhd f = (f^* \circ L^*)^* \in \Gamma_0(\mathcal{H})$. Hence, it follows from (4.2.14) and 2 that

$$
\begin{aligned}
\operatorname{prox}^U_{L\rhd f} &= \mathrm{Id} - U^{-1} \operatorname{prox}^{U^{-1}}_{f^*\circ L^*} U \\
&= \mathrm{Id} - U^{-1}\left(\mathrm{Id} - UL \operatorname{prox}^U_{f,L} U^{-1}\right) U \qquad (4.2.48) \\
&= L \operatorname{prox}^U_{f,L} \qquad (4.2.49)
\end{aligned}
$$

and the proof is complete. □

Remark 4.2.10. 1. *In the case when $U = \mu\mathrm{Id}$, the existence of solutions to (4.2.39) is assumed in [23, Proposition 5.2(iii)] and its uniqueness is supposed in [9, Theorem 4.7]. On the other hand, the strong monotonicity of $(L^*L + \partial f)$ is assumed in [3, 10] in order to guarantee the existence and uniqueness of solutions to the optimization problem in (4.2.39). Previous approaches are needed in order to guarantee that sequences of ADMM are well defined. Proposition 4.2.9(1) is more general, since it is obtained from Theorem 4.2.9 and the maximal monotonicity of $L(\partial f^*)L^*$, which is obtained from the qualification condition $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*)$ and $f^* \circ L^* \in \Gamma_0(\mathcal{H})$ in view of [1, Corollary 16.53(i) & Theorem 20.25].*

2. *In [12, 18] fixed point approaches are used in order to compute $\operatorname{prox}_{f\circ L}$ in the context of the sparse recovery in image processing. This approach leads to sub-iterations in optimization algorithms needing to compute $\operatorname{prox}_{f\circ L}$. Our computation is direct once $\operatorname{prox}^U_{f,L}$ is easily computable.*

3. *Note that Proposition 4.2.9(1) yields*

$$\operatorname{prox}^{U^{-1}}_{f^*\circ L^*} + UL \operatorname{prox}^{U}_{f,L} U^{-1} = \mathrm{Id}. \tag{4.2.50}$$

In the case when $L = \mathrm{Id}$*, since* $\operatorname{prox}^U_{f,\mathrm{Id}} = \operatorname{prox}^U_f$*, (4.2.50) reduces to [1, Proposition 24.24(ii)], which is a non-standard metric version of Moreau's decomposition [20] first derived for mutually polar cones [19].*

Acknowledgments

The first author thanks the support of ANID under grants FONDECYT 1190871, programa de financiamento basal of the Center of Mathematical Modelling (CMM) of the Universidad de Chile, and grant Redes 180032. The second author thanks the support of ANID-Subdirección de Capital Humano/Doctorado Nacional/2018-21181024 and of the Dirección de Postgrado y Programas from UTFSM through Programa de Incentivos a la Iniciación Científica (PIIC).

Bibliography

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5.

[2] R. I. Boţ, S.-M. Grad, and G. Wanka, *Maximal monotonicity for the precomposition with a linear operator*, SIAM J. Optim., 17 (2006), pp. 1239–1252, https://doi.org/10.1137/050641491.

[3] K. Bredies and H. Sun, *A proximal point analysis of the preconditioned alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 878–907, https://doi.org/10.1007/s10957-017-1112-5.

[4] L. Briceño-Arias and F. Roldán, *Resolvent of the parallel composition and proximity operator of the infimal postcomposition*, 2021, https://doi.org/10.48550/ARXIV.2109.06771, https://arxiv.org/abs/2109.06771.

[5] L. M. Briceño-Arias and F. Roldán, *Split-Douglas-Rachford algorithm for composite monotone inclusions and split-ADMM*, SIAM J. Optim., 31 (2021), pp. 2987–3013, https://doi.org/10.1137/21M1395144.

[6] M. N. Bùi and P. L. Combettes, *Warped proximal iterations for monotone inclusions*, J. Math. Anal. Appl., 491 (2020), pp. 124315, 21, https://doi.org/10.1016/j.jmaa.2020.124315.

[7] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145, https://doi.org/10.1007/s10851-010-0251-1.

[8] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318, https://doi.org/10.1080/02331934.2012.733883.

[9] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, *Proximal splitting algorithms: A tour of recent advances, with new twists*, 2020, https://arxiv.org/abs/1912.00137. https://arxiv.org/pdf/1912.00137.pdf.

[10] F. D. Côté, I. N. Psaromiligkos, and W. J. Gross, *A theory of generalized proximity for ADMM*, in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2017, pp. 578–582, https://doi.org/10.1109/GlobalSIP.2017.8309025.

[11] J. Douglas, Jr. and H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439, https://doi.org/10.2307/1993056.

[12] M.-J. Fadili and J.-L. Starck, *Monotone operator splitting for optimization problems in sparse recovery*, in 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 1461–1464.

[13] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Programming, 72 (1996), pp. 1–15, https://doi.org/10.1016/0025-5610(95)00012-7.

[14] D. Gabay, *Chapter IX applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, New York, 1983, pp. 299 – 331, https://doi.org/10.1016/S0168-2024(08)70034-1.

[15] P. Giselsson, *Nonlinear Forward-Backward Splitting with Projection Correction*, SIAM J. Optim., 31 (2021), pp. 2199–2226, https://doi.org/10.1137/20M1345062.

[16] X. Jiang and L. Vandenberghe, *Bregman primal-dual first-order method and application to sparse semidefinite programming*, Comput. Optim. Appl., 81 (2022), pp. 127–159, https://doi.org/10.1007/s10589-021-00339-7.

[17] J.-L. Lions and G. Stampacchia, *Variational inequalities*, Comm. Pure Appl. Math., 20 (1967), pp. 493–519, https://doi.org/10.1002/cpa.3160200302.

[18] C. A. Micchelli, L. Shen, and Y. Xu, *Proximity algorithms for image models: denoising*, Inverse Problems, 27 (2011), pp. 045009, 30, https://doi.org/10.1088/0266-5611/27/4/045009.

[19] J.-J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires*, C. R. Acad. Sci. Paris, 255 (1962), pp. 238–240.

[20] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[21] A. Moudafi, *Computing the resolvent of composite operators*, Cubo, 16 (2014), pp. 87–96, https://doi.org/10.4067/s0719-06462014000300007.

[22] D. O'Connor and L. Vandenberghe, *On the equivalence of the primal-dual hybrid gradient method and Douglas-Rachford splitting*, Math. Program., 179 (2020), pp. 85–108, https://doi.org/10.1007/s10107-018-1321-1.

[23] A. Themelis and P. Patrinos, *Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results*, SIAM J. Optim., 30 (2020), pp. 149–181, https://doi.org/10.1137/18M1163993.

[24] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

[25] Y. Yang, Y. Tang, and C. Zhu, *Iterative methods for computing the resolvent of composed operators in Hilbert spaces*, Mathematics, 7 (2019), https://doi.org/10.3390/math7020131.

Part II

Splitting Algorithms when $L = \mathrm{Id}$ and B is Cocoercive

Preface to Part II

In the second part we aim at solving separately the cases $B_3 = 0$ and $V = \mathcal{H}$ in Problem 1.1.9.

In particular, in Chapter 5 we consider the case $B_3 = 0$ and we provide a splitting algorithm which fully exploits the structure of the problem, generalizing some methods in the literature. Also, we derive a splitting method for solving primal-dual monotone inclusions including normal cones, Lipschitzian-monotone operators and cocoercive operators. We also derive an algorithm for solving convex composite optimization problems under vector subspace constraints and we implement it in a TV-regularized least-squares problems with constraints.

In Chapter 6 we derive a fully split method for solving Problem 1.1.9 in the case $V = \mathcal{H}$, which takes advantage of the intrinsic properties of the operators. We also derive an algorithm for solving optimization problems involving convex in Gâteaux differentiable functions, linear compositions, and Gâteaux differentiable nonlinear constrains. Finally, we provide numerical experiments which illustrate the efficiency of that method.

Chapter 5

Forward-Partial Inverse-Half-Forward Splitting Algorithm for Monotone Inclusions

5.1 Introduction and Main Results

In this chapter we aim at solving numerically the following problem.

Problem 5.1.1. *Let $\mathcal{H}$ be a real Hilbert space and let V be a closed vector subspace of $\mathcal{H}$. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $B_2\colon \mathcal{H} \to \mathcal{H}$ be a monotone and L-Lipschitzian operator for some $L \in \left]0,+\infty\right[$, and let $B\colon \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator for some $\beta \in \left]0,+\infty\right[$. The problem is to*

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx + B_2x + N_V x, \tag{5.1.1}$$

under the assumption that its solutions set Z is nonempty.

Problem 5.1.1 models a wide class of problems in engineering including mechanical problems [35, 37, 38], differential inclusions [2, 48], game theory [1, 14], restoration and denoising in image processing [20, 21, 28], traffic theory [9, 34, 36], among others.

The following is our main algorithm from this section, which allows to solve numerically Problem 5.1.1 in view of Theorem 5.1.3 below.

Algorithm 5.1.2. *In the context of Problem 5.1.1, let $(x_0, y_0) \in V \times V^{\perp}$, let $\gamma \in \left]0,+\infty\right[$,* *and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,+\infty[$. *Consider the recurrence*

$$\left\lfloor\begin{array}{l}\textit{for } n=1,2,\ldots\\ \quad \textit{find } (p_n,q_n)\in\mathcal{H}^2 \textit{ such that } x_n+\gamma y_n-\lambda_n\gamma P_V(B+B_2)x_n=p_n+\gamma q_n\\ \qquad\qquad \textit{and } \dfrac{P_Vq_n}{\lambda_n}+P_{V^\perp}q_n\in A\left(P_Vp_n+\dfrac{P_{V^\perp}p_n}{\lambda_n}\right),\\ x_{n+1}=P_Vp_n+\lambda_n\gamma P_V(B_2x_n-B_2P_Vp_n),\\ y_{n+1}=P_{V^\perp}q_n.\end{array}\right. \tag{5.1.2}$$

Note that (5.1.2) involves only one activation of B, two of B_2, and three projections onto V at each iteration. Now, we present the convergence result of Algorithm 5.1.2.

Theorem 5.1.3. *In the context of Problem 5.1.1, set*

$$\chi=\frac{4\beta}{1+\sqrt{1+16\beta^2L^2}}\in\left]0,\min\left\{2\beta,\frac{1}{L}\right\}\right[, \tag{5.1.3}$$

let $\gamma\in\,]0,+\infty[$, *and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $[\varepsilon,\chi/\gamma-\varepsilon]$ *for some* $\varepsilon\in\,]0,\chi/(2\gamma)[$. *Moreover, let* $(x_0,y_0)\in V\times V^\perp$ *and let* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *be the sequences generated by Algorithm 5.1.2. Then* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *are sequences in* V *and* $V^\perp$, *respectively, and there exist* $\overline{x}\in Z$ *and* $\overline{y}\in V^\perp\cap(A\overline{x}+P_V(B+B_2)\overline{x})$ *such that* $x_n\rightharpoonup\overline{x}$ *and* $y_n\rightharpoonup\overline{y}$.

The inclusion in (5.1.2) is not always easy to solve. Hence, by an adequate choosing of the sequence $(\lambda_n)_{n\in\mathbb{N}}$, we derive the following corollary in which this inclusion can be explicitly computed in terms of the resolvent of A.

Corollary 5.1.4. *In the context of Problem 5.1.1, let* $(x_0,y_0)\in V\times V^\perp$, *let* $\chi\in\,]0,+\infty[$ *be the constant defined in (5.1.3), let* $\gamma\in\,]0,\chi[$, *and let* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *be the sequences generated by the recurrence*

$$\left\lfloor\begin{array}{l}\textit{for } n=1,2,\ldots\\ p_n=J_{\gamma A}\big(x_n+\gamma y_n-\gamma P_V(B+B_2)x_n\big)\\ r_n=P_Vp_n\\ x_{n+1}=r_n+\gamma P_V(B_2x_n-B_2r_n)\\ y_{n+1}=y_n-\dfrac{p_n-r_n}{\gamma}.\end{array}\right. \tag{5.1.4}$$

Then, there exist $\overline{x}\in Z$ *and* $\overline{y}\in V^\perp\cap(A\overline{x}+P_V(B+B_2)\overline{x})$ *such that* $x_n\rightharpoonup\overline{x}$ *and* $y_n\rightharpoonup\overline{y}$.

Note that, in the case when $B=0$, the recurrence in (5.1.4) reduces to FPF (Algorithm 1.1.12) [11], in that case we can take $\beta\to\infty$ which yields $\chi\to1/L$. On the other hand, when $B_2 = 0$, taking $L \to 0$ we obtain $\chi \to 2\beta$ and (5.1.4) reduces to FPS (Algorithm 1.1.11) [10]. Moreover, when $V = \mathcal{H}$, (5.1.4) reduces to FBHF (Algorithm 1.1.13) [15].

In this section, we also provide a method for solving numerically the following primal-dual composite inclusion.

Problem 5.1.5. *Let H be a real Hilbert space, let V be a closed vector subspace of H, let $\mathsf{A}\colon \mathsf{H} \to 2^{\mathsf{H}}$ be maximally monotone, let $\mathsf{M}\colon \mathsf{H} \to \mathsf{H}$ be monotone and μ-Lipschitzian, for some $\mu \in \left]0,+\infty\right[$, let $\mathsf{C}\colon \mathsf{H} \to \mathsf{H}$ be ζ-cocoercive, for some $\zeta \in \left]0,+\infty\right[$, and let m be a strictly positive integer. For every $i \in \{1,\ldots,m\}$, let G_i be a real Hilbert space, let $\mathsf{B}_i\colon \mathsf{G}_i \to 2^{\mathsf{G}_i}$ be maximally monotone, let $\mathsf{N}_i\colon \mathsf{G}_i \to 2^{\mathsf{G}_i}$ be monotone and such that N_i^{-1} is ν_i-Lipschitzian, for some $\nu_i \in \left]0,+\infty\right[$, let D_i be maximally monotone and δ_i-strongly monotone, for some $\delta_i \in \left]0,+\infty\right[$, and let $\mathsf{L}_i\colon \mathsf{H} \to \mathsf{G}_i$ be a nonzero bounded linear operator. The problem is to*

$$\text{find}\quad \overline{\mathsf{x}} \in \mathsf{H}, \overline{\mathsf{u}}_1 \in \mathsf{G}_1, \ldots, \overline{\mathsf{u}}_m \in \mathsf{G}_m \quad \text{such that}$$
$$\begin{cases} 0 & \in \mathsf{A}\overline{\mathsf{x}} + \mathsf{M}\overline{\mathsf{x}} + \mathsf{C}\overline{\mathsf{x}} + \sum_{i=1}^m \mathsf{L}_i^*\overline{\mathsf{u}}_i + N_{\mathsf{V}}\overline{\mathsf{x}} \\ 0 & \in \big(\mathsf{B}_1^{-1} + \mathsf{N}_1^{-1} + \mathsf{D}_1^{-1}\big)\overline{\mathsf{u}}_1 - \mathsf{L}_1\overline{\mathsf{x}} \\ & \vdots \\ 0 & \in \big(\mathsf{B}_m^{-1} + \mathsf{N}_m^{-1} + \mathsf{D}_m^{-1}\big)\overline{\mathsf{u}}_m - \mathsf{L}_m\overline{\mathsf{x}}, \end{cases} \tag{5.1.5}$$

under the assumption that the solution set $\boldsymbol{Z}$ to (5.1.5) is nonempty.

The following proposition is consequence of previous methods and provides an algorithm for solving Problem 5.1.5.

Proposition 5.1.6. *Consider the framework of Problem 5.1.5 and set*

$$L = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^m \|\mathsf{L}_i\|^2} \quad \text{and} \quad \beta = \min\{\zeta, \delta_1, \ldots, \delta_m\}. \tag{5.1.6}$$

Let $\mathsf{x}_0 \in \mathsf{V}$, let $\mathsf{y}_0 \in \mathsf{V}^\top$, for every $i \in \{1,\ldots,m\}$, let $\mathsf{u}_{i,0} \in \mathsf{G}_i$, set $\gamma \in \left]0,\chi\right[$, where χ is *defined in* (5.2.12)*, and consider the routine*

$$\left\lfloor\begin{array}{l} \textit{for } n = 1, 2, \ldots \\ \mathsf{p}_n = J_{\gamma \mathsf{A}}\left(\mathsf{x}_n + \gamma \mathsf{y}_n - \gamma P_{\mathsf{V}}\left((\mathsf{M}+\mathsf{C})\mathsf{x}_n + \sum_{i=1}^{m} \mathsf{L}_i^* \mathsf{u}_{i,n}\right)\right) \\ \mathsf{q}_n = P_{\mathsf{V}} \mathsf{p}_n \\ \left\lfloor\begin{array}{l} \textit{for } i = 1, \ldots, m \\ \mathsf{r}_{i,n} = J_{\gamma \mathsf{B}_i^{-1}}\left(\mathsf{u}_{i,n} - \gamma\big((\mathsf{N}_i^{-1} + \mathsf{D}_i^{-1})\mathsf{u}_{i,n} - \mathsf{L}_i \mathsf{x}_n\big)\right) \\ \mathsf{u}_{i,n+1} = \mathsf{r}_{i,n} - \gamma\big(\mathsf{N}_i^{-1}\mathsf{r}_{i,n} - \mathsf{N}_i^{-1}\mathsf{u}_{i,n} - \mathsf{L}_i(\mathsf{q}_n - \mathsf{x}_n)\big) \end{array}\right. \\ \mathsf{x}_{n+1} = \mathsf{q}_n - \gamma P_{\mathsf{V}}\left(\mathsf{M}\mathsf{q}_n - \mathsf{M}\mathsf{x}_n + \sum_{i=1}^{m} \mathsf{L}_i^*(\mathsf{r}_{i,n} - \mathsf{u}_{i,n})\right) \\ \mathsf{y}_{n+1} = \mathsf{y}_n - \dfrac{\mathsf{p}_n - \mathsf{q}_n}{\gamma}. \end{array}\right. \tag{5.1.7}$$

Then, $(\mathsf{x}_n)_{n\in\mathbb{N}}$ *is a sequence in* V *and there exists* $(\bar{\mathsf{x}}, \bar{\mathsf{u}}_1, \ldots, \bar{\mathsf{u}}_m) \in \boldsymbol{Z}$ *such that* $\mathsf{x}_n \rightharpoonup \bar{\mathsf{x}}$ *and, for every* $i \in \{1, \ldots, m\}$*,* $\mathsf{u}_{i,n} \rightharpoonup \bar{\mathsf{u}}_i$*.*

Additionally, by applying the method in (5.1.7) in the optimization context, under adequate qualification condition, we obtain a method for solving the following optimization problem

$$\min_{\mathsf{x}\in\mathsf{V}} \left(\mathsf{f}(\mathsf{x}) + \mathsf{h}(\mathsf{x}) + \sum_{i=1}^{m} (\mathsf{g}_i \,\square\, \ell_i)(\mathsf{L}_i \mathsf{x})\right), \tag{5.1.8}$$

where $\mathsf{f} \in \Gamma_0(\mathsf{H})$, $\mathsf{h}\colon \mathsf{H} \to \mathbb{R}$ is convex differentiable with ζ^{-1}-Lipschitzian gradient, for every $i \in \{1, \ldots, m\}$, $\ell_i \in \Gamma_0(\mathsf{G}_i)$ is ν_i-strongly convex and $\mathsf{g}_i \in \Gamma_0(\mathsf{G}_i)$.

Finally we implement our method in the context of TV-regularized least-square problems with constraints and we compare its performance with previous methods in the literature.

5.2 Article: Forward-Partial Inverse-Half-Forward Splitting Algorithm for Solving Monotone Inclusions[1]

Abstract In this paper we provide a splitting algorithm for solving coupled monotone inclusions in a real Hilbert space involving the sum of a normal cone to a vector subspace,

[1][12] Luis M. Briceño-Arias, Jinjian Chen, Fernando Roldán, and Yuchao Tang. Forward-partial inverse-half-forward splitting algorithm for solving monotone inclusions, 2021, https://arxiv.org/abs/2104.01516.

a maximally monotone, a monotone-Lipschitzian, and a cocoercive operator. The proposed method takes advantage of the intrinsic properties of each operator and generalizes the method of partial inverses and the forward-backward-half forward splitting, among other methods. At each iteration, our algorithm needs two computations of the Lipschitzian operator while the cocoercive operator is activated only once. By using product space techniques, we derive a method for solving a composite monotone primal-dual inclusions including linear operators and we apply it to solve constrained composite convex optimization problems. Finally, we apply our algorithm to a constrained total variation least-squares problem and we compare its performance with efficient methods in the literature.

5.2.1 Introduction

In this paper we study the numerical resolution of the following inclusion problem. The normal cone to V is denoted by N_V.

Problem 5.2.1. *Let $\mathcal{H}$ be a real Hilbert space and let V be a closed vector subspace of $\mathcal{H}$. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $B\colon \mathcal{H} \to \mathcal{H}$ be a monotone and L-Lipschitzian operator for some $L \in \left]0,+\infty\right[$, and let $C\colon \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator for some $\beta \in \left]0,+\infty\right[$. The problem is to*

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx + Cx + N_V x, \tag{5.2.1}$$

under the assumption that its solutions set Z is nonempty.

Problem 5.2.1 models a wide class of problems in engineering including mechanical problems [35, 37, 38], differential inclusions [2, 48], game theory [1, 14], restoration and denoising in image processing [20, 21, 28], traffic theory [9, 34, 36], among others.

In the case when $V = \mathcal{H}$ and the resolvent of B is available, Problem 5.2.1 can be solved by the algorithms in [29, 30] and, if B is linear, by the algorithm in [41]. Moreover, if the resolvent of B is difficult to compute, Problem 5.2.1 can be solved by the *forward-backward-half forward* algorithm (FBHF) proposed in [11]. FBHF implement explicit activations of B and C and generalizes the classical forward-backward splitting [42] and Tseng's splitting [52] when $B = 0$ and $C = 0$, respectively.

In the case when $V \neq \mathcal{H}$, a splitting algorithm for solving the case $B = C = 0$ is proposed in [49] using the partial inverse of A with respect to V and extensions for the cases $B = 0$ and $C = 0$ are proposed in [10] and [11], respectively. On the other hand, the algorithms proposed in [5, 8, 4, 6, 7, 13, 19, 24, 23, 25, 27, 29, 31, 32, 39, 40, 43, 45, 46, 47, 53] can solve Problem 5.2.1 under additional assumptions or without exploiting the vector subspace structure and the intrinsic properties of the operators involved. Indeed, the algorithms in [8, 6, 7, 13, 23, 32] need to compute the resolvents of B and C, which are not explicit in general or they can be numerically expensive. In addition, previous methods do not take advantage of the vector subspace structure of Problem 5.2.1. The schemes proposed in [4, 24, 31, 39] take advantage of the properties of B, but the cocoercivity of C and the vector subspace structure are not leveraged. In fact, the algorithms in [4, 24, 31, 39] may consider $B + C$ as a monotone and Lipschitzian operator and activates it twice by iteration. In contrast, the algorithms in [19, 27, 43, 46, 47] activates $B + C$ only once by iteration, but they need to store in the memory the two past iterations and the step-size is reduced significantly. In addition, the methods proposed in [5, 25, 29, 40, 45, 53] take advantage of the cocoercivity of C, but they do not exploit neither the properties of B nor the vector subspace structure of the problem.

Furthermore, note that Problem 5.2.1 can be solved by the algorithms proposed in [11, 18] by considering N_V as any maximally monotone operator via product space techniques. These approaches do not exploit the vector subspace structure of the problem and need to update additional auxiliary dual variables at each iteration, which affects their efficiency in large scale problems. Moreover, since $B + C$ is monotone and $(\beta^{-1} + L)$-Lipschitzian, Problem 5.2.1 can be solved by [11]. However, this implementation needs two computations of C by iteration which affects its efficiency when C is computationally expensive and also may increment drastically the number of iterations to achieve the convergence criterion, as perceived in [11, Section 7.1] in the case $V = \mathcal{H}$.

In this paper we propose a splitting algorithm which fully exploits the vector subspace structure, the cocoercivity of C, and the Lipschitzian property of B. In the particular case when $V = \mathcal{H}$, we recover [11], which generalizes the forward-backward splitting and Tseng's splitting [52]. For general vector subspaces, our algorithm also recovers the methods proposed in [10, 11, 49]. By using standard product space techniques, we apply our algorithm to solve composite primal-dual monotone inclusions including a normal cone to a vector subspace, cocoercive, and Lipschitzian-monotone operators and composite convex optimization problems under vector subspace constraints. We implement our method in the context of TV-regularized least-squares problems with constraints and we compare its performance with previous methods in the literature including [26]. We observe that, in the case when the matrix in the data fidelity term has large norm values, our implementation is more efficient.

The paper is organized as follows. In Section 5.2.2 we set our notation. In Section 5.2.3 we provide our main algorithm for solving Problem 5.2.1 and its proof of convergence. In Section 5.2.4 we derive a method for solving a composite monotone primal-dual inclusion, including monotone, Lipschitzian, cocoercive, and bounded linear operators. In this section we also derive an algorithm for solve constrained composite convex optimization problems. Finally, in Section 5.2.5 we provide numerical experiments illustrating the efficiency of our proposed method.

5.2.2 Notations and Preliminaries

Throughout this paper $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces. We denote their scalar products by $\langle \cdot \mid \cdot \rangle$, the associated norms by $\| \cdot \|$, and by $\rightharpoonup$ the weak convergence. Given a linear bounded operator $L : \mathcal{H} \to \mathcal{G}$, we denote its adjoint by $L^*\colon \mathcal{G} \to \mathcal{H}$. Id denotes the identity operator on $\mathcal{H}$. Let $D \subset \mathcal{H}$ be non-empty and let $T : D \to \mathcal{H}$. Let $\beta \in]0, +\infty[$. The operator T is $\beta-$cocoercive if

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid Tx - Ty \rangle \geq \beta \|Tx - Ty\|^2 \tag{5.2.2}$$

and it is $L-$Lipschitzian if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leq L\|x - y\|. \tag{5.2.3}$$

Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain, range, and graph of A are $\operatorname{dom} A = \big\{x \in \mathcal{H} \bigm| Ax \neq \varnothing\big\}$, $\operatorname{ran} A = \big\{u \in \mathcal{H} \bigm| (\exists x \in \mathcal{H})\, u \in Ax\big\}$, and $\operatorname{gra} A = \big\{(x, u) \in \mathcal{H} \times \mathcal{H} \bigm| u \in Ax\big\}$, respectively. The set of zeros of A is $\operatorname{zer} A = \big\{x \in \mathcal{H} \bigm| 0 \in Ax\big\}$, the inverse of A is $A^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon u \mapsto \big\{x \in \mathcal{H} \bigm| u \in Ax\big\}$, and the resolvent of A is $J_A = (\mathrm{Id} + A)^{-1}$. The operator A is monotone if

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v \rangle \geq 0 \tag{5.2.4}$$

and it is maximally monotone if it is monotone and there exists no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$, i.e., for every $(x, u) \in \mathcal{H} \times \mathcal{H}$,

$$(x, u) \in \operatorname{gra} A \quad \Leftrightarrow \quad (\forall (y, v) \in \operatorname{gra} A) \;\; \langle x - y \mid u - v \rangle \geq 0. \tag{5.2.5}$$

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f\colon \mathcal{H} \to]-\infty, +\infty]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is defined by $f^*\colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$, which is a function in $\Gamma_0(\mathcal{H})$, the subdifferential of f is the maximally monotone operator

$$\partial f\colon x \mapsto \big\{u \in \mathcal{H} \bigm| (\forall y \in \mathcal{H}) \;\; f(x) + \langle y - x \mid u\rangle \leq f(y)\big\},$$

we have that $(\partial f)^{-1} = \partial f^*$, and that $\operatorname{zer} \partial f$ is the set of minimizers of f, which is denoted by $\operatorname{arg\,min}_{x\in\mathcal{H}} f$. We denote by

$$\operatorname{prox}_f\colon x \mapsto \underset{y\in\mathcal{H}}{\operatorname{arg\,min}} \Big(f(y) + \frac{1}{2}\|x - y\|^2\Big). \tag{5.2.6}$$

We have $\operatorname{prox}_f = J_{\partial f}$. Moreover, it follows from [3, Theorem 14.3] that

$$\operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{f^*/\gamma} \circ \mathrm{Id}/\gamma = \mathrm{Id}. \tag{5.2.7}$$

Given a non-empty closed convex set $C \subset \mathcal{H}$, we denote by P_C the projection onto C, by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise, and by $N_C = \partial(\iota_C)$ the normal cone to C. The partial inverse of A with respect to a closed vector subspace V of $\mathcal{H}$, denoted by A_V, is defined by

$$(\forall (x,y) \in \mathcal{H}^2) \quad y \in A_V x \quad \Leftrightarrow \quad (P_V y + P_{V^\perp} x) \in A(P_V x + P_{V^\perp} y). \tag{5.2.8}$$

Note that $A_{\mathcal{H}} = A$ and $A_{\{0\}} = A^{-1}$. For further properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [3].

The following is a simplified version of the algorithm proposed in [15, Theorem 2.3].

Proposition 5.2.2. *[15, Theorem 2.3] Let $\hat{L} \in]0,+\infty[$, let $\hat{\beta} \in]0,+\infty[$, let $\mathcal{A}\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $\mathcal{B}\colon \mathcal{H} \to \mathcal{H}$ be monotone and $\hat{L}$-Lipschitzian, and let $\mathcal{C}\colon \mathcal{H} \to \mathcal{H}$ be a $\hat{\beta}$-cocoercive operator. Suppose that $\operatorname{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C}) \neq \varnothing$ and set*

$$\hat{\chi} = \frac{4\hat{\beta}}{1+\sqrt{1+16\hat{\beta}^2\hat{L}^2}} \in \left]0, \min\left\{2\hat{\beta}, \frac{1}{\hat{L}}\right\}\right[, \tag{5.2.9}$$

let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, \hat{\chi}-\varepsilon]$, for some $\varepsilon \in]0, \hat{\chi}/2[$. Moreover, let $z_0 \in \mathcal{H}$ and consider the following recurrence

$$\left\lfloor \begin{array}{l} \text{for } n = 1,2,\ldots \\ s_n = J_{\lambda_n \mathcal{A}}\big(z_n - \lambda_n(\mathcal{B}+\mathcal{C})z_n\big) \\ z_{n+1} = s_n + \lambda_n(\mathcal{B}z_n - \mathcal{B}s_n). \end{array} \right. \tag{5.2.10}$$

Then, $(z_n)_{n\in\mathbb{N}}$ converges weakly to some $\overline{z} \in \operatorname{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C})$.

Observe that (5.2.10) reduces to forward-backward splitting when $\mathcal{B} = 0$ (and $L = 0$), and to a version of Tseng's splitting when $\mathcal{C} = 0$ (and $\beta \to +\infty$) [13, 52].

5.2.3 Main Result

The following is our main algorithm, whose convergence is proved in Theorem 5.2.4 below.

Algorithm 5.2.3. *In the context of Problem 5.2.1, let $(x_0, y_0) \in V \times V^\perp$, let $\gamma \in]0,+\infty[$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$. Consider the recurrence*

$$\left\lfloor \begin{array}{l} \text{for } n = 1,2,\ldots \\ \quad \text{find } (p_n, q_n) \in \mathcal{H}^2 \text{ such that } x_n + \gamma y_n - \lambda_n \gamma P_V(B+C)x_n = p_n + \gamma q_n \\ \qquad\quad \text{and } \dfrac{P_V q_n}{\lambda_n} + P_{V^\perp} q_n \in A\left(P_V p_n + \dfrac{P_{V^\perp} p_n}{\lambda_n}\right), \\ x_{n+1} = P_V p_n + \lambda_n \gamma P_V(Bx_n - BP_V p_n), \\ y_{n+1} = P_{V^\perp} q_n. \end{array} \right. \tag{5.2.11}$$

Note that (5.2.3) involves only one activation of C, two of B, and three projections onto V at each iteration.

Theorem 5.2.4. *In the context of Problem 5.2.1, set*

$$\chi = \frac{4\beta}{1+\sqrt{1+16\beta^2L^2}} \in \left]0, \min\left\{2\beta, \frac{1}{L}\right\}\right[, \tag{5.2.12}$$

let $\gamma \in \left]0,+\infty\right[$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, \chi/\gamma - \varepsilon]$ for some $\varepsilon \in \left]0, \chi/(2\gamma)\right[$. Moreover, let $(x_0, y_0) \in V \times V^\perp$ and let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 5.2.3. Then $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are sequences in V and $V^\perp$, respectively, and there exist $\overline{x} \in Z$ and $\overline{y} \in V^\perp \cap (A\overline{x} + P_V(B+C)\overline{x})$ such that $x_n \rightharpoonup \overline{x}$ and $y_n \rightharpoonup \overline{y}$.

Proof. Define

$$\begin{cases} \mathcal{A}_\gamma = (\gamma A)_V \colon \mathcal{H} \to 2^{\mathcal{H}} \\ \mathcal{B}_\gamma = \gamma P_V \circ B \circ P_V \colon \mathcal{H} \to \mathcal{H} \\ \mathcal{C}_\gamma = \gamma P_V \circ C \circ P_V \colon \mathcal{H} \to \mathcal{H}. \end{cases} \tag{5.2.13}$$

It follows from [11, Proposition 3.1(i)&(ii)] that $\mathcal{A}_\gamma$ is maximally monotone and that $\mathcal{B}_\gamma$ is monotone and γL-Lipschitzian. Moreover, $\mathcal{C}_\gamma$ is β/γ-cocoercive in view of [10, Proposition 5.1(ii)]. Since C is β^{-1}-Lipschitzian, $B + C$ is $(\beta^{-1} + L)$-Lipschitzian, and (5.2.13) and the linearity of P_V yield

$$\mathcal{B}_\gamma + \mathcal{C}_\gamma = \gamma P_V \circ (B + C) \circ P_V. \tag{5.2.14}$$

Therefore, [11, Proposition 3.1(iii)] implies that $\hat{x} \in \mathcal{H}$ is a solution to Problem 5.2.1 if and only if

$$\hat{x} \in V \quad \text{and} \quad \big(\exists \hat{y} \in V^\perp \cap (A\hat{x} + B\hat{x} + C\hat{x})\big) \\ \hat{x} + \gamma\big(\hat{y} - P_{V^\perp}(B+C)\hat{x}\big) \in \operatorname{zer}(\mathcal{A}_\gamma + \mathcal{B}_\gamma + \mathcal{C}_\gamma). \tag{5.2.15}$$

Now, since $x_0 \in V$ and $y_0 \in V^\perp$, it follows from Algorithm 5.2.11 that $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are sequences in V and $V^\perp$, respectively. In addition, from Algorithm 5.2.11 and [11, Proposition 3.1(i)] we deduce that

$$(\forall n \in \mathbb{N}) \quad J_{\lambda_n \mathcal{A}_\gamma}(x_n + \gamma y_n - \lambda_n \gamma P_V(B+C)x_n) = P_V p_n + \gamma P_{V^\perp} q_n. \tag{5.2.16}$$

For every $n \in \mathbb{N}$, set $z_n = x_n + \gamma y_n$ and set $s_n = P_V p_n + \gamma P_{V^\perp} q_n$. Hence, for every $n \in \mathbb{N}$, $P_V s_n = P_V p_n$, $P_{V^\perp} s_n = \gamma P_{V^\perp} q_n$, and (5.2.16) and (5.2.14) yield

$$\begin{aligned} s_n &= J_{\lambda_n \mathcal{A}_\gamma}(x_n + \gamma y_n - \lambda_n \gamma P_V(B+C)x_n) \\ &= J_{\lambda_n \mathcal{A}_\gamma}(z_n - \lambda_n \gamma P_V(B+C)P_V z_n) \\ &= J_{\lambda_n \mathcal{A}_\gamma}\big(z_n - \lambda_n(\mathcal{B}_\gamma + \mathcal{C}_\gamma)z_n\big). \end{aligned} \tag{5.2.17}$$

Thus, from Algorithm 5.2.11 we deduce that, for every $n \in \mathbb{N}$,

$$\begin{aligned} z_{n+1} &= x_{n+1} + \gamma y_{n+1} \\ &= P_V p_n + \lambda_n \gamma P_V(Bx_n - BP_V p_n) + \gamma P_{V^\perp} q_n \\ &= P_V s_n + \lambda_n(\gamma P_V B P_V z_n - \gamma P_V B P_V s_n) + P_{V^\perp} s_n \\ &= s_n + \lambda_n(\mathcal{B}_\gamma z_n - \mathcal{B}_\gamma s_n). \end{aligned} \tag{5.2.18}$$

Therefore, we obtain from (5.2.17) and (5.2.18) that

$$\left\lfloor \begin{array}{l} \text{for } n = 1, 2, \ldots \\ s_n = J_{\lambda_n \mathcal{A}_\gamma}\big(z_n - \lambda_n(\mathcal{B}_\gamma + \mathcal{C}_\gamma) z_n\big) \\ z_{n+1} = s_n + \lambda_n(\mathcal{B}_\gamma z_n - \mathcal{B}_\gamma s_n). \end{array} \right. \tag{5.2.19}$$

Altogether, by setting $\hat{\beta} = \beta/\gamma$ and $\hat{L} = \gamma L$, we have $\hat{\chi} = \chi/\gamma$ and Proposition 5.2.2 asserts that there exists $\overline{z} \in \operatorname{zer}(\mathcal{A}_\gamma + \mathcal{B}_\gamma + \mathcal{C}_\gamma)$ such that $z_n \rightharpoonup \overline{z}$. Furthermore, by setting $\overline{x} = P_V \overline{z}$ and $\overline{y} = P_{V^\perp}\overline{z}/\gamma$, we have $-(\mathcal{B}_\gamma + \mathcal{C}_\gamma)(\overline{x} + \gamma\overline{y}) \in \mathcal{A}_\gamma(\overline{x} + \gamma\overline{y})$, which, in view of (5.2.13), is equivalent to $-P_V(B + C)\overline{x} + \overline{y} \in A\overline{x}$. Therefore, by defining $\hat{y} = \overline{y} + P_{V^\perp}(B + C)\overline{x} \in V^\perp \cap (A\overline{x} + B\overline{x} + C\overline{x})$, we have $\overline{x} + \gamma(\hat{y} - P_{V^\perp}(B + C)\overline{x}) \in \operatorname{zer}(\mathcal{A}_\gamma + \mathcal{B}_\gamma + \mathcal{C}_\gamma)$ and (5.2.15) implies that $\overline{x} \in Z$ and that $\overline{y} \in V^\perp \cap (A\overline{x} + P_V(B + C)\overline{x})$. Moreover, from the weakly continuity of P_V and $P_{V^\perp}$, we obtain $x_n = P_V z_n \rightharpoonup P_V \overline{z} = \overline{x}$ and $y_n = P_{V^\perp} z_n/\gamma \rightharpoonup P_{V^\perp}\overline{z}/\gamma = \overline{y}$, which completes the proof. □

The sequence $(\lambda_n)_{n\in\mathbb{N}}$ in Algorithm 5.2.3 can be manipulated in order to accelerate the convergence. However, as in [10, 11, 50], the inclusion in (5.2.11) is not always easy to solve. The following result provides a particular case of our method, in which this inclusion can be explicitly computed in terms of the resolvent of A.

Corollary 5.2.5. *In the context of Problem 5.2.1, let $(x_0, y_0) \in V \times V^\perp$, let $\chi \in\,]0, +\infty[$ be the constant defined in (5.2.12), let $\gamma \in\,]0, \chi[$, and let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be the sequences generated by the recurrence*

$$\left\lfloor \begin{array}{l} \text{for } n = 1, 2, \ldots \\ p_n = J_{\gamma A}\big(x_n + \gamma y_n - \gamma P_V(B + C)x_n\big) \\ r_n = P_V p_n \\ x_{n+1} = r_n + \gamma P_V(Bx_n - Br_n) \\ y_{n+1} = y_n - \dfrac{p_n - r_n}{\gamma}. \end{array} \right. \tag{5.2.20}$$

Then, there exist $\overline{x} \in Z$ and $\overline{y} \in V^\perp \cap (A\overline{x} + P_V(B + C)\overline{x})$ such that $x_n \rightharpoonup \overline{x}$ and $y_n \rightharpoonup \overline{y}$.

Proof. Note that (5.2.20) implies that $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are sequences in V and $V^\perp$, respectively. Fix $n\in\mathbb{N}$ and set $q_n=(x_n+\gamma y_n-\gamma P_V(B+C)x_n-p_n)/\gamma$. Hence, we obtain from (5.2.20) that $p_n+\gamma q_n=x_n+\gamma y_n-\gamma P_V(B+C)x_n$, that $q_n\in Ap_n$, that $x_{n+1}=P_Vp_n+\gamma P_V(Bx_n-BP_Vp_n)$, and that $y_{n+1}=y_n-(p_n-P_Vp_n)/\gamma=y_n-P_{V^\perp}p_n/\gamma=P_{V^\perp}q_n$. Therefore, (5.2.20) is a particular case of Algorithm 5.2.3 when $\lambda_n\equiv 1\in\,]0,\chi/\gamma[$ and the result hence follows from Theorem 5.2.4. □

Remark 5.2.6.

1. *Note that, in the case when $C=0$, (5.2.20) reduces to the method proposed in [11]. Observe that in this case we can take $\beta\to+\infty$ which yields $\chi\to 1/L$.*
2. *Note that, in the case when $B=0$, (5.2.20) reduces to the method proposed in [10]. In this case, we can take $L\to 0$, which yields $\chi\to 2\beta$.*
3. *In the case when $V=\mathcal{H}$, (5.2.20) reduces to the algorithm proposed in [15] (see also Proposition 5.2.2).*

5.2.4 Applications

In this section we tackle the following composite primal-dual monotone inclusion.

Problem 5.2.7. *Let H be a real Hilbert space, let V be a closed vector subspace of H, let $\mathsf{A}\colon\mathsf{H}\to 2^{\mathsf{H}}$ be maximally monotone, let $\mathsf{M}\colon\mathsf{H}\to\mathsf{H}$ be monotone and μ-Lipschitzian, for some $\mu\in\,]0,+\infty[$, let $\mathsf{C}\colon\mathsf{H}\to\mathsf{H}$ be ζ-cocoercive, for some $\zeta\in\,]0,+\infty[$, and let m be a strictly positive integer. For every $i\in\{1,\ldots,m\}$, let G_i be a real Hilbert space, let $\mathsf{B}_i\colon\mathsf{G}_i\to 2^{\mathsf{G}_i}$ be maximally monotone, let $\mathsf{N}_i\colon\mathsf{G}_i\to 2^{\mathsf{G}_i}$ be monotone and such that N_i^{-1} is ν_i-Lipschitzian, for some $\nu_i\in\,]0,+\infty[$, let D_i be maximally monotone and δ_i-strongly monotone, for some $\delta_i\in\,]0,+\infty[$, and let $\mathsf{L}_i\colon\mathsf{H}\to\mathsf{G}_i$ be a nonzero bounded linear operator. The problem is to*

$$\text{find}\quad \overline{\mathsf{x}}\in\mathsf{H},\overline{\mathsf{u}}_1\in\mathsf{G}_1,\ldots,\overline{\mathsf{u}}_m\in\mathsf{G}_m\quad\text{such that}\quad \begin{cases} 0 & \in \mathsf{A}\overline{\mathsf{x}}+\mathsf{M}\overline{\mathsf{x}}+\mathsf{C}\overline{\mathsf{x}}+\sum_{i=1}^m\mathsf{L}_i^*\overline{\mathsf{u}}_i+N_{\mathsf{V}}\overline{\mathsf{x}}\\ 0 & \in \big(\mathsf{B}_1^{-1}+\mathsf{N}_1^{-1}+\mathsf{D}_1^{-1}\big)\overline{\mathsf{u}}_1-\mathsf{L}_1\overline{\mathsf{x}}\\ & \vdots\\ 0 & \in \big(\mathsf{B}_m^{-1}+\mathsf{N}_m^{-1}+\mathsf{D}_m^{-1}\big)\overline{\mathsf{u}}_m-\mathsf{L}_m\overline{\mathsf{x}}, \end{cases} \tag{5.2.21}$$

under the assumption that the solution set $\boldsymbol{Z}$ to (5.2.21) is nonempty.

Note that, if $(\overline{\mathsf{x}},\overline{\mathsf{u}}_1,\ldots,\overline{\mathsf{u}}_m)\in\boldsymbol{Z}$ then $\overline{\mathsf{x}}$ solves the primal inclusion

$$\text{find}\quad \overline{\mathsf{x}}\in\mathsf{H}\quad\text{such that}\quad 0\in\mathsf{A}\overline{\mathsf{x}}+\mathsf{M}\overline{\mathsf{x}}+\mathsf{C}\overline{\mathsf{x}}+\sum_{i=1}^m\mathsf{L}_i^*\left(\left(\mathsf{B}_i\,\square\,\mathsf{N}_i\,\square\,\mathsf{D}_i\right)\mathsf{L}_i\overline{\mathsf{x}}\right)+N_{\mathsf{V}}\overline{\mathsf{x}} \tag{5.2.22}$$

and $(\overline{\mathsf{u}}_1,\ldots,\overline{\mathsf{u}}_m)$ solves the dual inclusion

$$\text{find } \overline{\mathsf{u}}_1\in\mathsf{G}_1,\ldots,\overline{\mathsf{u}}_m\in\mathsf{G}_m \text{ such that}$$
$$(\exists\mathsf{x}\in\mathsf{H})\quad \begin{cases}-\sum_{i=1}^m \mathsf{L}_i^*\overline{\mathsf{u}}_i\in\mathsf{A}\mathsf{x}+\mathsf{M}\mathsf{x}+\mathsf{C}\mathsf{x}+N_{\mathsf{V}}\mathsf{x}\\ (\forall i\in\{1,\ldots,m\})\quad \overline{\mathsf{u}}_i\in\big(\mathsf{B}_i\,\square\,\mathsf{N}_i\,\square\,\mathsf{D}_i\big)\mathsf{L}_i\mathsf{x}.\end{cases}\tag{5.2.23}$$

In the case when $\mathsf{V}=\mathsf{H}$, $\mathsf{C}=0$, and, for every $i\in\{1,\ldots,m\}$, $\mathsf{D}_i^{-1}=0$, this problem can be solved by algorithms in [22, 24] by using Tseng's splitting [52] in a suitable product space. In the case when $\mathsf{V}=\mathsf{H}$, $\mathsf{M}=0$, and, for every $i\in\{1,\ldots,m\}$, $\mathsf{N}_i^{-1}=0$, this problem can be solved by algorithms in [25, 53] by using forward-backward splitting in a suitable product space. Since $\mathsf{M}+\mathsf{C}$ and $(\mathsf{N}_i^{-1}+\mathsf{D}_i^{-1})_{1\le i\le m}$ are monotone and Lipschitzian and N_{V} is maximally monotone, Problem 5.2.7 can be solved by the algorithms in [22, 24]. However, these methods do not exploit the cocoercivity or the vector subspace structure of Problem 5.2.7. Other algorithms as those in [23, 40, 17] provide alternatives for solving Problem 5.2.7, but any of them exploit its vector subspace and cocoercive structure. In the case when $\mathsf{M}=0$, and, for every $i\in\{1,\ldots,m\}$, $\mathsf{N}_i^{-1}=0$, the algorithm in [16] exploits the vector subspace structure of Problem 5.2.7 by using the partial inverse of A with respect to V. The following result provides a fully split algorithm to solve Problem 5.2.7 in its full generality. It is obtained by using (5.2.20) in a suitable product space, which exploits the vector subspace structure and which activates each cocoercive operator only once by iteration.

Proposition 5.2.8. *Consider the framework of Problem 5.2.7 and set*

$$L=\max\{\mu,\nu_1,\ldots,\nu_m\}+\sqrt{\sum_{i=1}^m\|\mathsf{L}_i\|^2}\quad\textit{and}\quad\beta=\min\{\zeta,\delta_1,\ldots,\delta_m\}.\tag{5.2.24}$$

Let $\mathsf{x}_0\in\mathsf{V}$, *let* $\mathsf{y}_0\in\mathsf{V}^\top$, *for every* $i\in\{1,\ldots,m\}$, *let* $\mathsf{u}_{i,0}\in\mathsf{G}_i$, *set* $\gamma\in\,]0,\chi[$, *where* χ *is* *defined in* (5.2.12), *and consider the routine*

$$
\begin{array}{l}
\textit{for } n = 1, 2, \ldots \\
\mathsf{p}_n = J_{\gamma\mathsf{A}}\Big(\mathsf{x}_n + \gamma\mathsf{y}_n - \gamma P_{\mathsf{V}}\Big((\mathsf{M}+\mathsf{C})\mathsf{x}_n + \sum_{i=1}^m \mathsf{L}_i^*\mathsf{u}_{i,n}\Big)\Big) \\
\mathsf{q}_n = P_{\mathsf{V}}\mathsf{p}_n \\
\quad \textit{for } i = 1, \ldots, m \\
\quad \mathsf{r}_{i,n} = J_{\gamma\mathsf{B}_i^{-1}}\Big(\mathsf{u}_{i,n} - \gamma\big((\mathsf{N}_i^{-1} + \mathsf{D}_i^{-1})\mathsf{u}_{i,n} - \mathsf{L}_i\mathsf{x}_n\big)\Big) \\
\quad \mathsf{u}_{i,n+1} = \mathsf{r}_{i,n} - \gamma\big(\mathsf{N}_i^{-1}\mathsf{r}_{i,n} - \mathsf{N}_i^{-1}\mathsf{u}_{i,n} - \mathsf{L}_i(\mathsf{q}_n - \mathsf{x}_n)\big) \\
\mathsf{x}_{n+1} = \mathsf{q}_n - \gamma P_{\mathsf{V}}\Big(\mathsf{M}\mathsf{q}_n - \mathsf{M}\mathsf{x}_n + \sum_{i=1}^m \mathsf{L}_i^*(\mathsf{r}_{i,n} - \mathsf{u}_{i,n})\Big) \\
\mathsf{y}_{n+1} = \mathsf{y}_n - \dfrac{\mathsf{p}_n - \mathsf{q}_n}{\gamma}.
\end{array}
\tag{5.2.25}
$$

Then, $(\mathsf{x}_n)_{n\in\mathbb{N}}$ *is a sequence in* V *and there exists* $(\overline{\mathsf{x}}, \overline{\mathsf{u}}_1, \ldots, \overline{\mathsf{u}}_m) \in \boldsymbol{Z}$ *such that* $\mathsf{x}_n \rightharpoonup \overline{\mathsf{x}}$ *and, for every* $i \in \{1, \ldots, m\}$, $\mathsf{u}_{i,n} \rightharpoonup \overline{\mathsf{u}}_i$.

Proof. Set $\mathcal{H} = \mathsf{H} \oplus \mathsf{G}_1 \oplus \cdots \oplus \mathsf{G}_m$ and define

$$
\begin{cases}
A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon (\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m) \mapsto \mathsf{A}\mathsf{x} \times \mathsf{B}_1^{-1}\mathsf{u}_1 \times \cdots \times \mathsf{B}_m^{-1}\mathsf{u}_m \\
B\colon \mathcal{H} \to \mathcal{H}\colon (\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m) \mapsto \big(\mathsf{M}\mathsf{x} + \sum_{i=1}^m \mathsf{L}_i^*\mathsf{u}_i, \mathsf{N}_1^{-1}\mathsf{u}_1 - \mathsf{L}_1\mathsf{x}, \ldots, \mathsf{N}_m^{-1}\mathsf{u}_m - \mathsf{L}_m\mathsf{x}\big) \\
C\colon \mathcal{H} \to \mathcal{H}\colon (\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m) \mapsto (\mathsf{C}\mathsf{x}, \mathsf{D}_1^{-1}\mathsf{u}_1, \ldots, \mathsf{D}_m^{-1}\mathsf{u}_m) \\
V = \big\{(\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m) \in \mathcal{H} \;\big|\; \mathsf{x} \in \mathsf{V}\big\}.
\end{cases}
\tag{5.2.26}
$$

Then, A is maximally monotone and B is monotone and L-Lipschitzian [24, eq.(3.11)], C is β-cocoercive [53, eq.(3.12)], and V is a closed vector subspace of $\mathcal{H}$. Therefore, Problem 5.2.7 is a particular instance of Problem 5.2.1. Moreover, we have from [3, Proposition 23.18] that

$$
\begin{cases}
(\forall \gamma > 0) \quad J_{\gamma A}\colon (\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m) \mapsto (J_{\gamma\mathsf{A}}\mathsf{x}, J_{\gamma\mathsf{B}_1^{-1}}\mathsf{u}_1, \ldots, J_{\gamma\mathsf{B}_m^{-1}}\mathsf{u}_m) \\
P_V\colon (\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m) \mapsto (P_{\mathsf{V}}\mathsf{x}, \mathsf{u}_1, \ldots, \mathsf{u}_m).
\end{cases}
\tag{5.2.27}
$$

Altogether, by defining

$$
(\forall n \in \mathbb{N}) \quad
\begin{cases}
x_n = (\mathsf{x}_n, \mathsf{u}_{1,n}, \ldots, \mathsf{u}_{m,n}) \\
y_n = (\mathsf{y}_n, 0, \ldots, 0) \\
p_n = (\mathsf{p}_n, \mathsf{r}_{1,n}, \ldots, \mathsf{r}_{m,n}) \\
q_n = (\mathsf{q}_n, \mathsf{s}_{1,n}, \ldots, \mathsf{s}_{m,n}),
\end{cases}
\tag{5.2.28}
$$

(5.2.25) is a particular case of (5.2.11) and the convergence follows from Corollary 5.2.5. □

Remark 5.2.9. *In the particular case when $\mathsf{V} = \mathsf{H}$ and $\mathsf{C} = \mathsf{D}_1^{-1} = \cdots = \mathsf{D}_m^{-1} = 0$, Proposition 5.2.8 recovers the main result in [24, Theorem 3.1] in the error-free case. By including non-standard metrics in the space $\mathcal{H}$ as in [11], we can also recover [16] when $\mathsf{M} = \mathsf{N}_1^{-1} = \cdots = \mathsf{N}_m^{-1} = 0$ and [53] if we additionally assume that $\mathsf{V} = \mathsf{H}$, but we preferred to avoid this generalization for simplicity.*

We now provide two important examples of Problem 5.2.7 and Proposition 5.2.8 in the context of convex optimization.

Example 5.2.10. *Suppose that $\mathsf{A} = \partial \mathsf{f}$, $\mathsf{M} = \mathsf{N}_1^{-1} = \cdots = \mathsf{N}_m^{-1} = 0$, $\mathsf{C} = \nabla \mathsf{h}$, for every $i \in \{1, \ldots, m\}$, $\mathsf{D}_i = \partial \ell_i$ and $\mathsf{B}_i = \partial \mathsf{g}_i$, where $\mathsf{f} \in \Gamma_0(\mathsf{H})$, $\mathsf{h}\colon \mathsf{H} \to \mathbb{R}$ is convex differentiable with ζ^{-1}-Lipschitzian gradient, for every $i \in \{1, \ldots, m\}$, $\ell_i \in \Gamma_0(\mathsf{G}_i)$ is ν_i-strongly convex and $\mathsf{g}_i \in \Gamma_0(\mathsf{G}_i)$. Then under the qualification condition [24, Proposition 4.3(i)]*

$$(0, \ldots, 0) \in \operatorname{sri}\Big(\times_{i=1}^{m} \big(\mathsf{L}_i(\mathsf{V} \cap \operatorname{dom} \mathsf{f}) - (\operatorname{dom} \mathsf{g}_i + \operatorname{dom} \ell_i)\big)\Big), \tag{5.2.29}$$

Problem 5.2.7 is equivalent to

$$\min_{\mathsf{x} \in \mathsf{V}} \left(\mathsf{f}(\mathsf{x}) + \mathsf{h}(\mathsf{x}) + \sum_{i=1}^{m} (\mathsf{g}_i \,\square\, \ell_i)(\mathsf{L}_i \mathsf{x}) \right), \tag{5.2.30}$$

which, in view of Proposition 5.2.8, can be solved by the algorithm

$$\left\lfloor \begin{array}{l} \text{for } n = 1, 2, \ldots \\ \mathsf{p}_n = \operatorname{prox}_{\gamma \mathsf{f}}\left(\mathsf{x}_n + \gamma \mathsf{y}_n - \gamma P_{\mathsf{V}}\left(\nabla \mathsf{h}(\mathsf{x}_n) + \sum_{i=1}^{m} \mathsf{L}_i^* \mathsf{u}_{i,n} \right)\right) \\ \mathsf{q}_n = P_{\mathsf{V}} \mathsf{p}_n \\ \left\lfloor \begin{array}{l} \text{for } i = 1, \ldots, m \\ \mathsf{r}_{i,n} = \operatorname{prox}_{\gamma \mathsf{g}_i^*}\big(\mathsf{u}_{i,n} - \gamma(\nabla \ell_i^*(\mathsf{u}_{i,n}) - \mathsf{L}_i \mathsf{x}_n)\big) \\ \mathsf{u}_{i,n+1} = \mathsf{r}_{i,n} + \gamma \mathsf{L}_i(\mathsf{q}_n - \mathsf{x}_n) \end{array} \right. \\ \mathsf{x}_{n+1} = \mathsf{q}_n - \gamma P_{\mathsf{V}}\left(\sum_{i=1}^{m} \mathsf{L}_i^*(\mathsf{r}_{i,n} - \mathsf{u}_{i,n}) \right) \\ \mathsf{y}_{n+1} = \mathsf{y}_n - \dfrac{\mathsf{p}_n - \mathsf{q}_n}{\gamma}, \end{array} \right. \tag{5.2.31}$$

where $\mathsf{x}_0 \in V$, $\mathsf{y}_0 \in V^{\perp}$, for every $i \in \{1, \ldots, m\}$, $\mathsf{u}_{i,0} \in \mathsf{G}_i$, $L = \sqrt{\sum_{i=1}^{m} \|\mathsf{L}_i\|^2}$, $\beta = \min\{\zeta, \delta_i, \ldots, \delta_m\}$, χ is defined in (5.2.12), and $\gamma \in \left]0, \chi\right[$. Observe that the algorithm

(5.2.31) *exploits the cocoercivity of* $\nabla \mathsf{h}$ *and* $(\nabla \ell_i^*)_{1\leq i\leq m}$ *by implementing them only once by iteration a difference of [24, Theorem 4.2], which needs to implement them twice by iteration.*

Example 5.2.11. *Consider the convex minimization problem*

$$\min_{\mathcal{x}\in\mathcal{H}} \big(\mathcal{f}(\mathcal{x}) + \mathcal{g}(\mathcal{L}\mathcal{x}) + \mathcal{h}(\mathcal{A}\mathcal{x})\big), \tag{5.2.32}$$

where $\mathcal{H}$, $\mathcal{G}$, *and* $\mathcal{K}$ *are real Hilbert spaces,* $\mathcal{f} \in \Gamma_0(\mathcal{H})$, $\mathcal{g} \in \Gamma_0(\mathcal{G})$, $\mathcal{L}\colon \mathcal{H} \to \mathcal{G}$, $\mathcal{A}\colon \mathcal{H} \to \mathcal{K}$, $\mathcal{h}\colon \mathcal{K} \to \mathbb{R}$ *is convex, differentiable with* β^{-1}*-Lipschitzian gradient, and suppose that*

$$0 \in \operatorname{sri}(\mathcal{L}\operatorname{dom}\mathcal{f} - \operatorname{dom}\mathcal{g}). \tag{5.2.33}$$

Note that $\mathcal{h}\circ\mathcal{A}$ *is convex, differentiable, and* $\nabla(\mathcal{h}\circ\mathcal{A}) = \mathcal{A}^*\circ\nabla\mathcal{h}\circ\mathcal{A}$ *is* $\beta^{-1}\|\mathcal{A}\|^2-$*Lipschitzian. Then, (5.2.32) can be solved by the primal-dual algorithm proposed in [26, 53], whose convergence is guaranteed under the assumption*

$$\sigma\|\mathcal{L}\|^2 \leq \frac{1}{\tau} - \frac{\|\mathcal{A}\|^2}{2\beta}, \tag{5.2.34}$$

where $\tau > 0$ *and* $\sigma > 0$ *are primal and dual step-sizes, respectively. Observe that, when* $\|\mathcal{A}\|$ *is large, this method is forced to choose small primal and dual step-sizes in order to ensure convergence. To overcome this inconvenient, we propose the following formulation*

$$\min_{\mathsf{x}\in\mathsf{V}} \big(\mathsf{f}(\mathsf{x}) + \mathsf{h}(\mathsf{x}) + \mathsf{g}(\mathsf{L}\mathsf{x})\big), \tag{5.2.35}$$

where

$$\begin{cases} \mathsf{H} = \mathcal{H}\oplus\mathcal{K} \\ \mathsf{G} = \mathcal{G} \\ \mathsf{T}\colon \mathsf{x} = (\mathcal{x}, \mathcal{w}) \mapsto \mathcal{A}\mathcal{x} - \mathcal{w} \\ \mathsf{V} = \ker \mathsf{T} \\ \mathsf{f}\colon \mathsf{x} = (\mathcal{x}, \mathcal{w}) \mapsto \mathcal{f}(\mathcal{x}) \\ \mathsf{g} = \mathcal{g} \\ \mathsf{L}\colon \mathsf{x} = (\mathcal{x}, \mathcal{w}) \mapsto \mathcal{L}\mathcal{x} \\ \mathsf{h}\colon \mathsf{x} = (\mathcal{x}, \mathcal{w}) \mapsto \mathcal{h}(\mathcal{w}). \end{cases} \tag{5.2.36}$$

Since in this case (5.2.29) reduces to (5.2.33), (5.2.32) is a particular instance of (5.2.30) when $m = 1$ *and* $\ell_1 = 0$. *Therefore, in view of [3, Example 29.19], (5.2.32) can be solved by the routine in (5.2.31) which, on this setting, reduces to:*

$$\left\lfloor\begin{array}{l} \textit{for } n = 1,2,\ldots \\ p_{1,n} = \mathrm{prox}_{\gamma f}\Big(x_n + \gamma y_{1,n} - \gamma\big(L^* u_n - A^* B\big(AL^* u_n - \nabla h(w_n)\big)\big)\Big) \\ p_{2,n} = w_n + \gamma y_{2,n} - \gamma\big(\nabla h(w_n) + B\big(AL^* u_n - \nabla h(w_n)\big)\big) \\ q_{1,n} = p_{1,n} - A^* B\big(Ap_{1,n} - p_{2,n}\big) \\ q_{2,n} = p_{2,n} + B\big(Ap_{1,n} - p_{2,n}\big) \\ r_n = \mathrm{prox}_{\gamma g^*}\big(u_n + \gamma L x_n\big) \\ u_{n+1} = r_n + \gamma L(q_{1,n} - x_n) \\ x_{n+1} = q_{1,n} - \gamma\left(L^*(r_n - u_n) - A^* BAL^*(r_n - u_n)\right) \\ w_{n+1} = q_{2,n} - \gamma BAL^*(r_n - u_n) \\ y_{1,n+1} = y_{1,n} - \dfrac{p_{1,n+1} - q_{1,n+1}}{\gamma} \\ y_{2,n+1} = y_{2,n} - \dfrac{p_{2,n+1} - q_{2,n+1}}{\gamma}, \end{array}\right. \tag{5.2.37}$$

where $B = (\mathrm{Id} + AA^*)^{-1}$ *can be computed only once before the loop,* $(x_0, w_0) \in V$*,* $(y_{1,0}, y_{2,0}) \in V^\perp$*,* $u_0 \in \mathcal{G}$*,* $L = \|L\|$*,* χ *is defined in* (5.2.12)*, and* $\gamma \in \left]0, \chi\right[$*.*

5.2.5 Numerical Experiments

In this section we consider the following optimization problem

$$\min_{y^0 \le x \le y^1} \left(\frac{\alpha_1}{2}\|Ax - z\|^2 + \alpha_2\|\nabla x\|_1 \right), \tag{5.2.38}$$

where $y^0 = (\eta_i^0)_{1\le i\le N}$, $y^1 = (\eta_i^1)_{1\le i\le N}$ are vectors in $\mathbb{R}^N$, α_1 and α_2 are in $\left]0, +\infty\right[$, $A \in \mathbb{R}^{K\times N}$, $z \in \mathbb{R}^K$, and $\nabla : \mathbb{R}^N \to \mathbb{R}^{N-1} : (\xi_i)_{1\le i\le N} \mapsto (\xi_{i+1} - \xi_i)_{1\le i\le N-1}$ is the discrete gradient. This problem appears when computing the fusion estimator in fused LASSO problems [33, 44, 51].

Note that (5.2.38) can be written equivalently as (5.2.32), where

$$\begin{cases} \mathcal{H} = \mathbb{R}^N \\ f = \iota_C \\ C = \times_{i=1}^N [\eta_i^0, \eta_i^1] \\ g = \alpha_2\|\cdot\|_1 \\ h = \alpha_1\|\cdot - z\|^2/2 \\ L = \nabla \end{cases} \tag{5.2.39}$$

Since $f \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^{N-1})$, h is convex, differentiable, $\nabla h = \alpha_1(\mathrm{Id} - z)$ is α_1-Lipschitzian, $\|\mathcal{L}\| = 2$, and (5.2.33) is trivially satisfied, (5.2.38) is a particular instance of Example 5.2.11. Hence, (5.2.38) can be solved by the algorithm in [26, 53] (called Condat-Vũ), by (5.2.37) (called FPIHF), and by [11] (called FPIF), which are compared in this section. In this context, the Algorithm Condat-Vũ [26, 53] reduces to the following routine.

Algorithm 4 Condat-Vũ [26, 53]

Let $x_0 \in \mathbb{R}^N$ and $u_0 \in \mathbb{R}^{N-1}$, let $(\sigma, \tau, \rho) \in \left]0, +\infty\right[^3$, and fix $\epsilon_0 > \varepsilon > 0$.
while $\epsilon_n > \varepsilon$ **do**
  $p_{n+1} = P_C\big(x_n - \tau(\alpha_1 \mathcal{A}^\top(\mathcal{A}x_n - z) + \nabla^\top u_n)\big)$
  $q_{n+1} = \sigma(\mathrm{Id} - \mathrm{prox}_{\alpha_2\|\cdot\|_1/\sigma})(u_n/\sigma + \nabla(2p_{n+1} - x_n))$
  $x_{n+1} = x_n + \rho(p_{n+1} - x_n)$
  $u_{n+1} = u_n + \rho(q_{n+1} - u_n)$
  $\epsilon_{n+1} = \mathcal{R}\big((x_{n+1}, u_{n+1}), (x_n, u_n)\big)$
end while
return (x_{n+1}, u_{n+1})

Observe that $P_C \colon (\xi_i)_{1 \le i \le N} \mapsto (\max\{\min\{\xi_i, \eta_i^1\}, \eta_i^0\})_{1 \le i \le N}$. The convergence of Algorithm 4 is guaranteed if

$$\sigma\|\mathcal{L}\|^2 \le \frac{1}{\tau} - \frac{\alpha_1\|\mathcal{A}\|^2}{2} \quad \text{and} \quad \rho \in \left]0, \delta\right[, \quad \text{where } \delta = 2 - \frac{\alpha_1\|\mathcal{A}\|^2}{2(\frac{1}{\tau} - \sigma\|\mathcal{L}\|^2)}. \tag{5.2.40}$$

Note that, the larger is $\alpha_1\|\mathcal{A}\|^2$, the smaller should be τ and σ in order to achieve convergence. On the other hand, by considering T defined in (5.2.36), the method in (5.2.37) writes as follows.

Algorithm 5 Forward-partial inverse-half-forward splitting (FPIHF)

Set $\mathcal{B} = (\mathrm{Id} + \mathcal{A}\mathcal{A}^\top)^{-1}$, let $(x_0, w_0) \in (\ker \mathsf{T})^2$, $(y_{1,0}, y_{2,0}) \in (\ker \mathsf{T}^\perp)^2$, $u_0 \in \mathbb{R}^K$, let $\gamma \in \left]0, +\infty\right[$, and fix $\epsilon_0 > \varepsilon > 0$.
while $\epsilon_n > \varepsilon$ **do**
  $p_{1,n} = P_C\big(x_n + \gamma y_{1,n} - \gamma\big(\nabla^\top u_n - \mathcal{A}^\top \mathcal{B}\big(\mathcal{A}\nabla^\top u_n - \alpha_1(w_n - z)\big)\big)\big)$
  $p_{2,n} = w_n + \gamma y_{2,n} - \gamma\big(\alpha_1(w_n - z) + \mathcal{B}\big(\mathcal{A}\nabla^\top u_n - \alpha_1(w_n - z)\big)\big)$
  $q_{1,n} = p_{1,n} - \mathcal{A}^\top \mathcal{B}(\mathcal{A}p_{1,n} - p_{2,n})$
  $q_{2,n} = p_{2,n} + \mathcal{B}\big(\mathcal{A}p_{1,n} - p_{2,n}\big)$
  $r_n = \gamma(\mathrm{Id} - \mathrm{prox}_{\alpha_2\|\cdot\|_1/\gamma})(u_n/\gamma + \nabla x_n)$
  $u_{n+1} = r_n + \gamma\nabla(q_{1,n} - x_n)$
  $x_{n+1} = q_{1,n} - \gamma\left(\nabla^\top(r_n - u_n) - \mathcal{A}^\top\mathcal{B}\mathcal{A}\nabla^\top(r_n - u_n)\right)$
  $w_{n+1} = q_{2,n} - \gamma\mathcal{B}\mathcal{A}\nabla^\top(r_n - u_n)$
  $y_{1,n+1} = y_{1,n} - (p_{1,n+1} - q_{1,n+1})/\gamma$
  $y_{2,n+1} = y_{2,n} - (p_{2,n+1} - q_{2,n+1})/\gamma$
  $\epsilon_{n+1} = \mathcal{R}\big((x_{n+1}, w_{n+1}, y^1_{n+1}, y^2_{n+1}), (x_n, w_n, y^1_n, y^2_n)\big)$
end while
return $(x_{n+1}, w_{n+1}, y^1_{n+1}, y^2_{n+1})$

In view of Example 5.2.11, the algorithm in (5.2.37) reduces to Algorithm 5, whose convergence is guaranteed if the step-size γ satisfies

$$0 < \gamma < \chi = \frac{4}{\alpha_1 + \sqrt{\alpha_1^2 + 64}}. \tag{5.2.41}$$

Observe that the condition for the step-size γ in (5.2.41) does not depend on $\|\mathcal{A}\|$.

The FPIF algorithm proposed in [11] for solving (5.2.38) differs from Algorithm 5 in the fact that the cocoercive gradient $\nabla h\colon x \mapsto x - z$ is implemented twice by iteration. Indeed, the algorithm considers the monotone Lipschitzian operator $(x, u, w) \mapsto (\nabla^\top u, \nabla x, \alpha_1(w - z))$, whose Lipschitz constant follows from

$$\begin{aligned}
|||\big(\nabla^\top u_1, \nabla x_1, \alpha_1(w_1 - z)\big) &- \big(\nabla^\top u_2, \nabla x_2, \alpha_1(w_2 - z)\big)|||^2 \\
&= \|\nabla^\top(u_1 - u_2)\|^2 + \|\nabla(x_1 - x_2)\|^2 + \alpha_1^2\|w_1 - w_2\|^2 \\
&\leq \|\nabla^\top\|^2\|u_1 - u_2\|^2 + \|\nabla\|^2\|x_1 - x_2\|^2 + \alpha_1^2\|w_1 - w_2\|^2 \\
&\leq \max\{\|\nabla\|^2, \alpha_1^2\}|||(x_1 - x_2, u_1 - u_2, w_1 - w_2)|||^2.
\end{aligned}$$

Therefore, the convergence of FPIF is guaranteed if $\gamma \in \left]0, 1/\max\{\|\nabla\|, \alpha_1\}\right[$, and, as in Algorithm 5, this condition does not depend on $\|\mathcal{A}\|$. In order to compare Condat-Vũ, FPIHF, and FPIF, we set $\alpha_1 = 5$ and $\alpha_2 = 0.5$ and we consider $\mathcal{A} = \kappa \cdot \mathrm{rand}(N, K)$, $y^0 = -1.5 \cdot \mathrm{rand}(N)$, $y^1 = 1.5 \cdot \mathrm{rand}(N)$, and $z = \mathrm{randn}(N)$, where $\kappa \in \{1/5, 1/10, 1/20, 1/30\}$, $N \in \{600, 1200, 2400\}$, $K \in \{N/3, N/2, 2N/3\}$, and $\mathrm{rand}(\cdot,\cdot)$ and $\mathrm{randn}(\cdot,\cdot)$ are functions in MATLAB generating matrices/vectors with uniformly and normal distributed entries, respectively. For each value of κ, N, and K, we generate 20 random realizations for $\mathcal{A}$, z, y^0, and y^1. Note that the average value of $\|\mathcal{A}\|$ increases as κ increase (see Figure 5.1 for $K = N/2$), which affects Algorithm 4 in view of (5.2.40). We also set $\rho = 0.99\cdot\delta$, where δ is defined in (5.2.40). In this setting, from (5.2.41) we deduce that the convergence of FPIHF is guaranteed for $\gamma < \chi \approx 0.2771$. On the other hand, since $\max\{\|\nabla\|, \alpha_1\} = \alpha_1 = 5$, the convergence of FPIF is guaranteed for $\gamma < 0.2$.

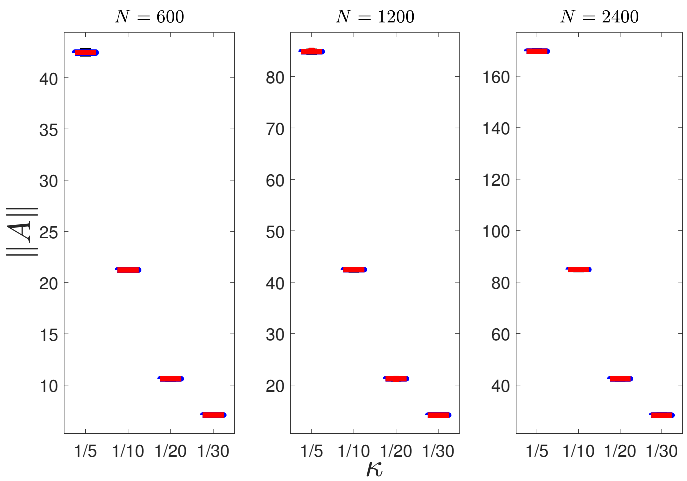


Figure 5.1: Box plot for the norm of the 20 random realizations of $\mathcal{A}$, $N \in \{600, 1200, 2400\}$, $K = N/2$.

In Tables 5.1-5.4 we provide the average time and number of iterations to achieve a tolerance $\varepsilon = 10^{-6}$ for each algorithm under study. In the case when an algorithm exceeds 50000 iterations in all cases, we write "⊠" in both columns. From these tables we can observe that when κ increases (and therefore, $\|\mathcal{A}\|$ increases), Condat-Vũ reduces its performance and does not converge within 50000 iterations for big dimensions and large values of κ. Moreover, the number of iterations of FPIHF is considerably lower than its competitors but with expensive computational time by iteration. This can be explained by the fact that FPIHF needs to compute three projections onto the kernel of $(x, w) \mapsto \mathcal{A}x - w$ at each iteration. We can also perceive that, at exception of some cases, the partial inverse-based algorithms increase their computational time to achieve convergence when K is larger. This can be explained by the fact that the dimension of matrix $\mathcal{B}$ is larger as K is larger, and it has to be implemented three times by iteration.

When $\kappa = 1/30$, we observe from Table 5.1, that FPIHF and Condat-Vũ are competitive and both are more efficient than FPIF. When $\kappa = 1/20$, we observe from Table 5.2 that FPIHF outperforms Condat-Vũ and FPIF for large dimensions. When $\kappa = 1/10$, we observe from Table 5.3 that FPIHF is the best algorithm at exception of the smallest

Table 5.1: Comparison of Condat-Vũ, FPIF, and FPIHF for the case $\kappa = 1/30$.

		$K = N/3$		$K = N/2$		$K = 2N/3$	
N	Algorithm	Av. time (s)	Av. iter	Av. time (s)	Av. iter	Av. time (s)	Av. iter
600	Condat-Vũ	0.89	11059	0.80	10047	0.76	9666
	FPIF	3.46	17454	3.91	14353	7.20	17430
	FPIHF	0.99	4851	1.24	4442	1.73	3996
1200	Condat-Vũ	11.32	17321	10.55	16129	10.54	16082
	FPIF	25.52	19930	32.37	13788	51.54	16443
	FPIHF	7.07	5425	13.76	5838	23.83	7570
2400	Condat-Vũ	74.17	34059	70.14	32216	69.48	31963
	FPIF	95.55	17747	138.67	16074	190.06	17216
	FPIHF	43.08	7961	64.68	7464	70.64	6369

Table 5.2: Comparison of Condat-Vũ, FPIF, and FPIHF for the case $\kappa = 1/20$.

		$K = N/3$		$K = N/2$		$K = 2N/3$	
N	Algorithm	Av. time (s)	Av. iter	Av. time (s)	Av. iter	Av. time (s)	Av. iter
600	Condat-Vũ	0.86	10752	0.81	10263	0.87	10992
	FPIF	2.67	13381	3.91	14204	5.88	14258
	FPIHF	0.97	4725	0.82	2900	1.63	3747
1200	Condat-Vũ	13.91	21209	13.35	20359	12.51	19118
	FPIF	23.30	18142	45.16	19222	52.60	16773
	FPIHF	9.07	6943	20.53	8689	10.91	3458
2400	Condat-Vũ	103.92	47673	98.92	45543	91.33	41996
	FPIF	89.77	16659	132.60	15374	145.58	13181
	FPIHF	32.27	5957	45.35	5234	83.48	7539

dimensional case in which it is competitive with Condat-Vũ. The latter does not converge within 50000 for dimension $N = 2400$. When $\kappa = 1/5$, FPIHF is the more efficient algorithm in all the cases under study, as it is illustrated in Table 5.4. Moreover, Condat-Vũ converge before 50000 iterations only in the lower dimensional case when $N = 600$. We conclude that, for higher values of $\|\mathcal{A}\|$ and larger dimensions, is more convenient to implement FPIHF.

Acknowledgment

The first author thanks the support of Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for centers of excellence and ANID under grant FONDECYT 1190871 from ANID-Chile. The third author thanks the support of

Table 5.3: Comparison of Condat-Vũ, FPIF, and FPIHF for the case $\kappa = 1/10$.

		$K = N/3$		$K = N/2$		$K = 2N/3$	
N	Algorithm	Av. time (s)	Av. iter	Av. time (s)	Av. iter	Av. time (s)	Av. iter
600	Condat-Vũ	1.43	18233	1.30	16747	1.25	15577
	FPIF	3.56	18040	3.01	11057	5.17	12389
	FPIHF	1.11	5414	1.30	4696	1.49	3436
1200	Condat-Vũ	30.19	46078	26.98	41243	24.05	36849
	FPIF	25.61	19916	30.70	13095	40.57	12960
	FPIHF	6.96	5343	10.16	4294	17.79	5657
2400	Condat-Vũ	⊠	⊠	⊠	⊠	⊠	⊠
	FPIF	98.90	18363	129.27	14975	172.05	15609
	FPIHF	28.90	5349	46.74	5391	60.61	5484

Table 5.4: Comparison of Condat-Vũ, FPIF, and FPIHF for the case $\kappa = 1/5$.

		$K = N/3$		$K = N/2$		$K = 2N/3$	
N	Algorithm	Av. time (s)	Av. iter	Av. time (s)	Av. iter	Av. time (s)	Av. iter
600	Condat-Vũ	3.76	48078	3.27	40998	2.58	33226
	FPIF	2.68	13527	3.31	11945	4.14	9840
	FPIHF	0.50	2428	0.64	2263	0.79	1780
1200	Condat-Vũ	⊠	⊠	⊠	⊠	⊠	⊠
	FPIF	21.26	16535	27.29	11627	35.55	11399
	FPIHF	7.23	5529	5.72	2424	10.25	3257
2400	Condat-Vũ	⊠	⊠	⊠	⊠	⊠	⊠
	FPIF	88.51	16392	124.71	14444	139.69	12653
	FPIHF	23.95	4414	35.51	4102	41.38	3773

ANID-Subdirección de Capital Humano/Doctorado Nacional/2018-21181024 and by the Dirección de Postgrado y Programas from UTFSM through Programa de Incentivos a la Iniciación Científica (PIIC). The forth author thanks the support of the National Natural Science Foundations of China (12061045, 11661056).

Bibliography

[1] H. Attouch and A. Cabot, *Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 547–598, https://doi.org/10.1007/s00245-019-09584-z.

[2] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2009, https://doi.org/10.1007/978-0-8176-4848-0.

[3] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5. With a foreword by Hédy Attouch.

[4] R. I. Boţ and E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algorithms, 71 (2016), pp. 519–540, https://doi.org/10.1007/s11075-015-0007-5.

[5] R. I. Boţ and E. R. Csetnek, *ADMM for monotone operators: convergence analysis and rates*, Adv. Comput. Math., 45 (2019), pp. 327–359, https://doi.org/10.1007/s10444-018-9619-3.

[6] R. I. Boţ, E. R. Csetnek, and A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM J. Optim., 23 (2013), pp. 2011–2036, https://doi.org/10.1137/12088255X.

[7] R. I. Boţ, E. R. Csetnek, and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Appl. Math. Comput., 256 (2015), pp. 472–487, https://doi.org/10.1016/j.amc.2015.01.017.

[8] R. I. Boţ and C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565, https://doi.org/10.1137/120901106.

[9] L. Briceño, R. Cominetti, C. E. Cortés, and F. Martínez, *An integrated behavioral model of land use and transport system: a hyper-network equilibrium*

approach, Netw. Spat. Econ., 8 (2008), pp. 201–224, https://doi.org/10.1007/s11067-007-9052-5.

[10] L. M. Briceño-Arias, *Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions*, Optimization, 64 (2015), pp. 1239–1261, https://doi.org/10.1080/02331934.2013.855210.

[11] L. M. Briceño-Arias, *Forward-partial inverse-forward splitting for solving monotone inclusions*, J. Optim. Theory Appl., 166 (2015), pp. 391–413, https://doi.org/10.1007/s10957-015-0703-2.

[12] L. M. Briceño-Arias, J. Chen, F. Roldán, and Y. Tang, *Forward-partial inverse-half-forward splitting algorithm for solving monotone inclusions*, 2021, https://doi.org/10.48550/ARXIV.2104.01516, https://arxiv.org/abs/2104.01516.

[13] L. M. Briceño-Arias and P. L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250, https://doi.org/10.1137/10081602X.

[14] L. M. Briceño-Arias and P. L. Combettes, *Monotone operator methods for Nash equilibria in non-potential games*, in Computational and analytical mathematics, vol. 50 of Springer Proc. Math. Stat., Springer, New York, 2013, pp. 143–159, https://doi.org/10.1007/978-1-4614-7621-4_9.

[15] L. M. Briceño-Arias and D. Davis, *Forward-backward-half forward algorithm for solving monotone inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871, https://doi.org/10.1137/17M1120099.

[16] L. Briceño-Arias, J. Deride, S. López-Rivera, and F. J. Silva, *A primal-dual partial inverse splitting for constrained monotone inclusions: Applications to stochastic programming and mean field games*, 2021, https://arxiv.org/abs/2007.01983.

[17] M. Bùi and P. Combettes, *Multivariate monotone inclusions in saddle form*, arXiv eprint, arXiv:2002.06135 (2020).

[18] M. N. Bùi and P. L. Combettes, *Multivariate monotone inclusions in saddle form*, Mathematics of Operations Research, to appear (2022).

[19] V. Cevher and B. Vũ, *A reflected forward-backward splitting method for monotone inclusions involving lipschitzian operators*, Set-Valued Var. Anal., (2020), https://doi.org/10.1007/s11228-020-00542-4.

[20] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188, https://doi.org/10.1007/s002110050258.

[21] J. Colas, N. Pustelnik, C. Oliver, P. Abry, J.-C. Géminard, and V. Vidal, *Nonlinear denoising for characterization of solid friction under low confinement pressure*, Physical Review E , 42 (2019), p. 91, https://doi.org/10.1103/PhysRevE.100.032803, https://hal.archives-ouvertes.fr/hal-02271333.

[22] P. L. Combettes, *Systems of structured monotone inclusions: duality, algorithms, and applications*, SIAM J. Optim., 23 (2013), pp. 2420–2447, https://doi.org/10.1137/130904160.

[23] P. L. Combettes and J. Eckstein, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions*, Math. Program., 168 (2018), pp. 645–672, https://doi.org/10.1007/s10107-016-1044-0.

[24] P. L. Combettes and J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued Var. Anal., 20 (2012), pp. 307–330, https://doi.org/10.1007/s11228-011-0191-y.

[25] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318, https://doi.org/10.1080/02331934.2012.733883.

[26] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479, https://doi.org/10.1007/s10957-012-0245-9.

[27] E. Csetnek, Y. Malitsky, and M. Tam, *Shadow Douglas-Rachford splitting for monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 665–678.

[28] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, https://doi.org/10.1002/cpa.20042.

[29] D. Davis and W. Yin, *A three-operator splitting scheme and its optimization applications*, Set-Valued Var. Anal., 25 (2017), pp. 829–858, https://doi.org/10.1007/s11228-017-0421-z.

[30] Y. Dong, *Weak convergence of an extended splitting method for monotone inclusions*, J. Global Optim., 79 (2021), pp. 257–277, https://doi.org/10.1007/s10898-020-00940-w.

[31] D. Dũng and B. C. Vũ, *A splitting algorithm for system of composite monotone inclusions*, Vietnam J. Math., 43 (2015), pp. 323–341, https://doi.org/10.1007/s10013-015-0121-7.

[32] J. Eckstein, *A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 155–182, https://doi.org/10.1007/s10957-017-1074-7.

[33] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, *Pathwise coordinate optimization*, Ann. Appl. Stat., 1 (2007), pp. 302–332, https://doi.org/10.1214/07-AOAS131.

[34] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Programming, 72 (1996), pp. 1–15, https://doi.org/10.1016/0025-5610(95)00012-7.

[35] D. Gabay, *Chapter IX applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, New York, 1983, pp. 299 – 331, https://doi.org/10.1016/S0168-2024(08)70034-1.

[36] E. M. Gafni and D. P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM J. Control Optim., 22 (1984), pp. 936–964, https://doi.org/10.1137/0322061.

[37] R. Glowinski and A. Marroco, *Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite, d'une classe de problemes de dirichlet non lineares*, Revue Francaise d'Automatique, Informatique et Recherche Operationelle, 9 (1975), pp. 41–76, https://doi.org/10.1051/M2AN/197509R200411.

[38] A. A. Goldstein, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society, 70 (1964), pp. 709 – 710, https://doi.org/bams/1183526263, https://doi.org/.

[39] P. R. Johnstone and J. Eckstein, *Projective splitting with forward steps only requires continuity*, Optim. Lett., 14 (2020), pp. 229–247, https://doi.org/10.1007/s11590-019-01509-7.

[40] P. R. Johnstone and J. Eckstein, *Single-forward-step projective splitting: exploiting cocoercivity*, Comput. Optim. Appl., 78 (2021), pp. 125–166, https://doi.org/10.1007/s10589-020-00238-3.

[41] P. Latafat and P. Patrinos, *Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators*, Comput. Optim. Appl., 68 (2017), pp. 57–93, https://doi.org/10.1007/s10589-017-9909-6.

[42] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.

[43] Y. Malitsky and M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim., 30 (2020), pp. 1451–1472, https://doi.org/10.1137/18M1207260.

[44] M. Ohishi, K. Fukui, K. Okamura, Y. Itoh, and H. Yanagihara, *Coordinate optimization for generalized fused Lasso*, Comm. Statist. Theory Methods, 50 (2021), pp. 5955–5973, https://doi.org/10.1080/03610926.2021.1931888.

[45] H. Raguet, J. Fadili, and G. Peyré, *A generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226, https://doi.org/10.1137/120872802.

[46] J. Rieger and M. K. Tam, *Backward-forward-reflected-backward splitting for three operator monotone inclusions*, Appl. Math. Comput., 381 (2020), pp. 125248, 10, https://doi.org/10.1016/j.amc.2020.125248.

[47] E. K. Ryu and B. C. Vũ, *Finding the forward-Douglas-Rachford-forward method*, J. Optim. Theory Appl., 184 (2020), pp. 858–876, https://doi.org/10.1007/s10957-019-01601-z.

[48] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1997, https://doi.org/10.1090/surv/049.

[49] J. E. Spingarn, *Partial inverse of a monotone operator*, Appl. Math. Optim., 10 (1983), pp. 247–265, https://doi.org/10.1007/BF01448388.

[50] J. E. Spingarn, *Applications of the method of partial inverses to convex programming: decomposition*, Math. Programming, 32 (1985), pp. 199–223, https://doi.org/10.1007/BF01586091.

[51] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, *Sparsity and smoothness via the fused lasso*, J. R. Stat. Soc. Ser. B Stat. Methodol., 67 (2005), pp. 91–108, https://doi.org/10.1111/j.1467-9868.2005.00490.x.

[52] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446, https://doi.org/10.1137/S0363012998338806.

[53] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

Chapter 6

Forward-Backward-Half Forward Algorithm with Line Search for Monotone Inclusions

6.1 Introduction and Main Results

In this chapter we aim at solving numerically the following problem.

Problem 6.1.1. *Let X be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $B : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator, for some $\beta > 0$, let $B_2 : \mathcal{H} \to \mathcal{H}$ be a monotone and L-Lipschitzian operator for some $L > 0$, and let $B_3 : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that B_3 is single valued and continuous in $\operatorname{dom} A \cup X \subset \operatorname{dom} B_3$. Moreover assume that $A + B_3$ is maximally monotone. The problem is to*

$$\textit{find} \quad x \in X \quad \textit{such that} \quad 0 \in Ax + Bx + B_2x + B_3x, \tag{6.1.1}$$

under the assumption that the set of solutions to (6.1.1) is nonempty.

This inclusion encompasses several problems in partial differential equations coming from mechanical models [24, 25, 26], differential inclusions [1, 36], game theory [11], among other disciplines.

The method proposed in this chapter splits the influence of the four operators involved in Problem 6.1.1. The operator A is activated implicitly via its resolvent, B and B_2 are activated explicitly, and B_3 is activated using a backtracking in order to define the step-sizes.

We first study some properties of the monotone operators involved in Problem 6.1.1, which ensure the finite termination of the backtracking procedure in our method.

Lemma 6.1.2. *In the context of Problem 6.1.1, let z and y in $\mathcal{H}$, and define*

$$(\forall \gamma > 0) \qquad x_{z,y}(\gamma) = J_{\gamma A}(z - \gamma y) \quad \textit{and} \quad \varphi_{z,y}(\gamma) = \frac{\|z - x_{z,y}(\gamma)\|}{\gamma}. \tag{6.1.2}$$

Then, the following statements holds:

1. *$\varphi_{z,y}$ is nonincreasing.*
2. *$(\forall z \in \operatorname{dom} A) \quad \lim_{\gamma \downarrow 0} \varphi_{z,y}(\gamma) = \|(A+y)^0 z\| = \min_{w \in Az+y} \|w\|$.*
3. *Set*
$$\mathcal{C} = \big\{z \in \mathcal{H} \;\big|\; \lim_{\gamma \downarrow 0} \varphi_{z,y}(\gamma) < +\infty\big\}. \tag{6.1.3}$$
Then, $\operatorname{dom} A \subset \mathcal{C} \subset \overline{\operatorname{dom}}\, A$.
4. *Suppose that one of the following holds:*
   (a) *$z \in \mathcal{C}$.*
   (b) *$z \in \operatorname{dom} B_3 \setminus \mathcal{C}$, $y = (B_1 + B_2 + B_3)z$ and B_3 is locally bounded at $P_{\overline{\operatorname{dom}}\, A} z$.*
   (c) *$z \in \operatorname{dom} B_3 \setminus \mathcal{C}$, $y = (B_1 + B_2 + B_3)z$, and $\overline{\operatorname{dom}}\, A \subset \operatorname{int} \operatorname{dom} B_3$.*

   Then, for every $\theta \in \,]0,1[$, there exists $\gamma(z) > 0$ such that, for every $\gamma \in \,]0, \gamma(z)]$,
$$\gamma \|B_3 z - B_3 x_{z,y}(\gamma)\| \le \theta \|z - x_{z,y}(\gamma)\|. \tag{6.1.4}$$

Note that Lemma 6.1.2 generalizes [12, Lemma 2.2]. More precisely, in the case when $B_2 = 0$, by setting $y = (B_1 + B_3)z$ in Lemma 6.1.2(1)&(2), we recover [12, Lemma 2.2(1)]. Moreover, realizing that [12, Lemma 2.2(2)] is valid for every $z \in \operatorname{dom} A$, it is a particular case of Lemma 6.1.2(3)&(4a).

The following is our main result from this section.

Theorem 6.1.3. *In the context of Problem 6.1.1, suppose that one of the following holds:*

1. *$X \subset \operatorname{dom} A$.*
2. *$\overline{\operatorname{dom}}\, A \subset \operatorname{dom} B_3$ and B_3 is locally bounded in $\operatorname{dom} B_3$.*
3. *$\overline{\operatorname{dom}}\, A \subset \operatorname{int} \operatorname{dom} B_3$.*

Let $\varepsilon \in \,]0,1[$, set $\rho = \min\{2\beta\varepsilon, \sqrt{1-\varepsilon}/L\}$, let $\sigma \in \,]0,1[$, let $\theta \in \big]0, \sqrt{1-\varepsilon} - L\rho\sigma\big[$, let $z_0 \in \operatorname{dom} B_3$, and consider the sequence $(z_n)_{n \in \mathbb{N}}$ defined by the recurrence

$$(\forall n \in \mathbb{N}) \qquad \left\lfloor \begin{array}{l} x_n = J_{\gamma_n A}(z_n - \gamma_n (B + B_2 + B_3) z_n) \\ z_{n+1} = P_X(x_n + \gamma_n (B_2 + B_3) z_n - \gamma_n (B_2 + B_3) x_n), \end{array} \right. \tag{6.1.5}$$

where, for every $n \in \mathbb{N}$, γ_n is the largest $\gamma \in \{\rho\sigma, \rho\sigma^2, \rho\sigma^3, \cdots\}$ satisfying

$$\gamma\|B_3 z_n - B_3 J_{\gamma A}(z_n - \gamma(B + B_2 + B_3)z_n)\| \leq \theta\|z_n - J_{\gamma A}(z_n - \gamma(B + B_2 + B_3)z_n)\|. \quad (6.1.6)$$

Moreover, assume that at least one of the following additional statements hold:

(i) $\liminf_{n\to\infty} \gamma_n = \delta > 0$.

(ii) B_3 *is uniformly continuous in any weakly compact subset of* $\overline{\mathrm{conv}}(\mathrm{dom}\, A \cup X)$.

Then, $(z_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 6.1.1.*

Theorem 6.1.3 is a generalization of [12, Theorem 2.3]. Indeed, when $B_2 = 0$ and $X \subset \mathrm{dom}\, A$, by taking $L \to 0$, we have $\rho \to 2\beta\varepsilon$ and $\theta \in \left]0, \sqrt{1-\varepsilon}\right[$. Hence, Theorem 6.1.3 recovers [12, Theorem 2.3(2)] noting that the uniform continuity in weakly compact subsets of $\overline{\mathrm{conv}}(\mathrm{dom}\, A \cup X) = \overline{\mathrm{dom}}\, A$ is needed. We hence generalize [12, Theorem 2.3(1)&(2)] to the case when $X \not\subset \mathrm{dom}\, A$.

In addition, the algorithm in (6.1.5) is a generalization of FBHF (Algorithm 1.1.13), FBS (Algorithm 1.1.10), and Tseng's splitting [38]. More precisely, in Theorem 6.1.3, if $B_3 = 0$, we have $\mathrm{dom}\, B_3 = \mathcal{H}$ and, for all $n \in \mathbb{N}$, $\gamma_n = \sigma\rho = \sigma \min\{2\beta\varepsilon, \sqrt{1-\varepsilon}/L\}$. Since, in this case $(\gamma_n)_{n\in\mathbb{N}}$ is constant, the largest step-size is obtained by taking $\varepsilon = \varepsilon(L, \beta) = 2/(1 + \sqrt{1 + 16\beta^2 L^2})$, which satisfies $2\beta\varepsilon = \sqrt{1-\varepsilon}/L = \chi(L, \beta)$, where

$$\chi(L, \beta) = \frac{4\beta}{1 + \sqrt{1 + 16\beta^2 L^2}}, \quad (6.1.7)$$

and $\gamma_n \equiv \gamma = \sigma\chi(L, \beta) \in]0, \chi(L, \beta)[$. Hence, we recover the result in [12, Theorem 2.3(1)] for constant step-sizes. Additionally, if $B_2 = 0$ and $X = \mathcal{H}$, we have $\varepsilon(L, \beta) \to 1$ and $\chi(L, \beta) \to 2\beta$ as $L \to 0$ and $\gamma_n \equiv 2\beta\sigma \in]0, 2\beta[$, recovering the the forward backward algorithm [29]. On the other hand, if $B_1 = 0$, we have $\chi(L, \beta) \to 1/L$ as $\beta \to \infty$ and $\gamma_n \equiv \sigma/L \in]0, 1/L[$, recovering the result in [38] for constant step-sizes.

Next, on this chapter, we derive a method for solving the following optimization problem.

Problem 6.1.4. *Let* $f \in \Gamma_0(\mathcal{H})$*, let* $g \in \Gamma_0(\mathcal{G})$*, let* $h : \mathcal{H} \to \mathbb{R}$ *be a convex Gâteaux differentiable function such that* ∇h *is* β^{-1}*-Lipschitzian for some* $\beta \in]0, +\infty[$*, let* $M : \mathcal{H} \to \mathcal{G}$ *be a bounded linear operator, and let* $e : \mathcal{H} \to \mathbb{R}^p : x \mapsto (e_i(x))_{1\leq i\leq p}$ *be such that, for every* $i \in \{1, \ldots, p\}$*,* e_i *is convex and Gâteaux differentiable in* $\mathrm{int}\,\mathrm{dom}\, e_i$*,* $\mathrm{dom}\, e_i$ *is closed,* $\cap_{i=1}^p \mathrm{intdom}\, e_i \neq \varnothing$*, and* $\mathrm{dom}\, \partial f \subset \cap_{i=1}^n \mathrm{intdom}\, e_i$*. The problem is to*

$$\min_{e(x)\in]-\infty,0]^p} f(x) + g(Mx) + h(x), \quad (6.1.8)$$

and we assume that solutions exist.

By using Lagrangian duality, under qualification conditions, we can write Problem 6.1.4 as Problem 6.1.1. The following proposition is a consequence of Theorem 6.1.3.

Proposition 6.1.5. *In the context of Problem 6.1.4, assume that* $0 \in \operatorname{sri}(\operatorname{dom} g - M(\operatorname{dom} f))$, *let* $X = X_1 \times X_2 \times X_3 \subset \operatorname{dom}\partial f \times \operatorname{dom}\partial g^* \times [0,+\infty[^p$ *be nonempty, closed, and convex, let* $\varepsilon \in]0,1[$, *set* $\rho = \min\{2\beta\varepsilon, \sqrt{1-\varepsilon}/\|M\|\}$, *let* $\sigma \in]0,1[$, *let* $\theta \in \left]0, \sqrt{1-\varepsilon} - \|M\|\rho\sigma\right[$. *For every* $\boldsymbol{z} = (z^1, z^2, z^3) \in \mathcal{H} \times \mathcal{G} \times \mathbb{R}^p$ *define* $\boldsymbol{\Phi}_{\boldsymbol{z}} \colon \gamma \mapsto (\Phi^1_{\boldsymbol{z}}(\gamma), \Phi^2_{\boldsymbol{z}}(\gamma), \Phi^3_{\boldsymbol{z}}(\gamma))$, *where*

$$\begin{aligned}
&\Phi^1_{\boldsymbol{z}} \colon \gamma \mapsto \operatorname{prox}_{\gamma f}\left(z^1 - \gamma\left(\nabla h(z^1) + M^* z^2 + \sum_{i=1}^p z^3_i \nabla e_i(z^1)\right)\right)\\
&\Phi^2_{\boldsymbol{z}} \colon \gamma \mapsto \operatorname{prox}_{\gamma g^*}(z^2 + \gamma M z^1)\\
&\Phi^3_{\boldsymbol{z}} \colon \gamma \mapsto P_{[0,+\infty[^p}\big(z^3 + \gamma e(z^1)\big).
\end{aligned} \tag{6.1.9}$$

Let $\boldsymbol{z}_0 = (z^1_0, z^2_0, z^3_0) \in \mathcal{H} \times \mathcal{G} \times \mathbb{R}^p$ *and consider the recurrence*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{aligned}
x^1_n &= \Phi^1_{\boldsymbol{z}_n}(\gamma_n)\\
x^2_n &= \Phi^2_{\boldsymbol{z}_n}(\gamma_n)\\
x^3_n &= \Phi^3_{\boldsymbol{z}_n}(\gamma_n)\\
z^1_{n+1} &= P_{X_1}\big(x^1_n + \gamma_n(M^* z^2_n + \textstyle\sum_{i=1}^p z^3_{n,i}\nabla e_i(z^1_n)) - \gamma_n(M^* x^2_n + \textstyle\sum_{i=1}^p x^3_{n,i}\nabla e_i(x^1_n))\big)\\
z^2_{n+1} &= P_{X_2}(x^2_n - \gamma_n M z^1_n + \gamma_n M x^1_n)\\
z^3_{n+1} &= P_{X_3}(x^3_n - \gamma_n e(z^1_n) + \gamma_n e(x^1_n))\\
\boldsymbol{z}_{n+1} &= (z^1_{n+1}, z^2_{n+1}, z^3_{n+1}),
\end{aligned}\right. \tag{6.1.10}$$

where, for every $n \in \mathbb{N}$, γ_n *is the largest* $\gamma \in \{\rho\sigma, \rho\sigma^2, \rho\sigma^3, \dots\}$ *satisfying*

$$\gamma^2\left(\left\|\sum_{i=1}^p z^3_{n,i}\nabla e_i(z^1_n) - \Phi^3_{\boldsymbol{z}_n,i}(\gamma)\nabla e(\Phi^1_{\boldsymbol{z}_n}(\gamma))\right\|^2 + \left\|e(z^1_n) - e(\Phi^1_{\boldsymbol{z}_n}(\gamma))\right\|^2\right) \leq \theta^2 |||\boldsymbol{z}_n - \boldsymbol{\Phi}_{\boldsymbol{z}_n}(\gamma)|||^2. \tag{6.1.11}$$

Moreover, assume that at least one of the following additional statements hold:

(i) $\liminf_{n\to\infty} \gamma_n = \delta > 0$.

(ii) For every $i \in \{1,\dots,p\}$, ∇e_i *is bounded and uniformly continuous in every weakly compact subset of* $\overline{\operatorname{dom}}\,\partial f$.

Then, $(z^1_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 6.1.4.*

Finally we implement our algorithm in a constrained total variation least-square problem and we compare its performance with available methods in the literature.

6.2 Article: Four Operator Splitting via a Forward-Backward-Half Forward Algorithm With Line Search[1]

Abstract In this article we provide a splitting method for solving monotone inclusions in real Hilbert space involving four operators: a maximally monotone, a monotone-Lipschitzian, a cocoercive, and a continuous operator. The proposed method takes advantage of the intrinsic properties of each operator, generalizing the forward-back-half forward splitting method and the Tseng's algorithm with line-search. At each iteration, our algorithm defines the step-size by using a line search in which the monotone-Lipschitzian and the cocoercive operators need only one activation. We also derive a method for solving non-linearly constrained composite convex optimization problems in real Hilbert spaces. Finally, we implement our algorithm in a constrained total variation least-square problem and we compare its performance with available methods in the literature.

6.2.1 Introduction

In this paper we aim at solving the following monotone inclusion problem.

Problem 6.2.1. *Let X be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $B_1 : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator, for some $\beta > 0$, let $B_2 : \mathcal{H} \to \mathcal{H}$ be a monotone and L-Lipschitzian operator for some $L > 0$, and let $B_3 : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that B_3 is single valued and continuous in* $\operatorname{dom} A \cup X \subset \operatorname{dom} B_3$. *Moreover assume that $A + B_3$ is maximally monotone. The problem is to*

$$\textit{find} \quad x \in X \quad \textit{such that} \quad 0 \in Ax + B_1x + B_2x + B_3x, \tag{6.2.1}$$

under the assumption that the set of solutions to (6.2.1) *is nonempty.*

This inclusion encompasses several problems in partial differential equations coming from mechanical models [24, 25, 26], differential inclusions [1, 36], game theory [11], among other disciplines. Algorithms proposed in [4, 3, 5, 6, 7, 10, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 27, 28, 30, 31, 32, 34, 39] can solve Problem 6.2.1 under additional assumptions or without exploiting the intrinsic properties of the involved operators, in the case $X = \mathcal{H}$. Indeed, the algorithms in [5, 6, 7, 10, 16, 23] need to compute the resolvents of B_1, B_2, and B_3, which are not explicit in general or they can be numerically expensive. The schemes proposed in [3, 17, 22, 27] take advantage of the properties of B_2, but the cocoercivity of B_1 and the continuity of B_3 are not leveraged. In fact, the algorithms in [3, 17, 22, 27] may

[1] Luis M. Briceño-Arias and Fernando Roldán. Four operator Splitting via a Forward-backward-half forward algorithm with line search, 2022, https://arxiv.org/abs/actualizar.

consider B_3 as a maximally monotone operator and B_1+B_2 as a monotone and Lipschitzian operator and activates it twice by iteration. In contrast, the algorithms in [14, 19, 30, 32, 34] activates $B_1 + B_2$ only once by iteration, but they need to store in the memory the two past iterations and the step size is reduced significantly. Furthermore, methods in [14, 19, 30] consider only one maximally monotone operator, hence, the resolvent of $A + B_3$ must be calculated; on the other hand, methods in [32, 34] need to calculate the resolvent of B_3. In addition, methods proposed in [4, 18, 20, 21, 28, 31, 39] take advantage of the cocoercivity of B_1, but they do not exploit the Lipschitzian property of B_2 or the continuity of B_3 and need to compute the resolvents of B_2 and B_3. The method in [13] exploits the properties of B_1 and B_2 but not considers B_3 which must be treated as any maximally monotone operator via it resolvent.

Other methods solving (6.2.1) including normal cones, when $B_3 = 0$ and either $B_1 =$ or $B_2 = 0$, are discussed in [8, 9, 37].

Methods exploiting the continuity B_3 are proposed in [12, 38]. In particular, the algorithm in [38] solves Problem 6.2.1 when $B_1 = B_2 = 0$. The *forward-backward-forward splitting* (FBF) method, proposed in [12], for solving numerically Problem 6.2.1 either when $B_3 = 0$ or when $B_2 = 0$. In order to provide their respective convergence results, in [12] and in [38], is assumed that $X \subset \operatorname{dom} A$.

In this paper, we propose a fully split method for solving Problem 6.2.1, which take advantage of each of their intrinsic properties of the operators, overcoming the drawbacks of the methods mentioned above. We generalize the proposed methods in [12, 38] allowing $X \not\subset A$ with additional hypothesis on B_3.

Another contribution of this manuscript is to derive an algorithm for solving optimization problems involving convex and Gâteaux differentiable functions, linear compositions and Gâteaux differentiable nonlinear constraints. We study properties of these functions for ensuring the convergence of this method. Finally, we provide numerical experiments which illustrate the efficiency of our algorithm.

The paper is organized as follows. In Section 6.2.2 we set our notation. In Section 6.2.3 we provide some technical lemmas, our splitting method for solving Problem 6.2.1, and our convergence result. In Section 6.2.4 we derive an algorithm for solving a constrained composite optimization problem. Finally, in Section 6.2.5 we provide a numerical experiment illustrating the efficiency of the proposed method in Section 6.2.4.

6.2.2 Preliminaries

Throughout this paper $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces. We denote the scalar product by $\langle \cdot \mid \cdot \rangle$ and the associated norm by $\|\cdot\|$. The symbols $\rightharpoonup$ and $\to$ denotes the weak and strong convergence, respectively. Given a linear bounded operator $M : \mathcal{H} \to \mathcal{G}$, we denote its adjoint by $M^*\colon \mathcal{G} \to \mathcal{H}$. Id denotes the identity operator on $\mathcal{H}$. Let $D \subset \mathcal{H}$ be non-empty, let $T : D \to \mathcal{H}$, and let $\beta \in \left]0, +\infty\right[$. The operator T is β-cocoercive if

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid Tx - Ty\rangle \geq \beta\|Tx - Ty\|^2 \tag{6.2.2}$$

and it is β-Lipschitzian if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leq \beta\|x - y\|. \tag{6.2.3}$$

Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain, range, graph, and the zeros of A are, respectively, $\operatorname{dom} A = \big\{x \in \mathcal{H} \;\big|\; Ax \neq \varnothing\big\}$, $\operatorname{ran} A = \big\{u \in \mathcal{H} \;\big|\; (\exists x \in \mathcal{H})\, u \in Ax\big\}$, $\operatorname{gra} A = \big\{(x, u) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$, and $\operatorname{zer} A = \big\{x \in \mathcal{H} \;\big|\; 0 \in Ax\big\}$. The inverse of A is $A^{-1} \colon \mathcal{H} \to 2^{\mathcal{H}} \colon u \mapsto \big\{x \in \mathcal{H} \;\big|\; u \in Ax\big\}$ and the resolvent of A is $J_A = (\mathrm{Id} + A)^{-1}$. The operator A is monotone if

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v\rangle \geq 0. \tag{6.2.4}$$

Moreover, A it is maximally monotone if it is monotone and there exists no monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$, i.e., for every $(x, u) \in \mathcal{H}\times\mathcal{H}$,

$$(x, u) \in \operatorname{gra} A \quad \Leftrightarrow \quad (\forall (y, v) \in \operatorname{gra} A) \;\; \langle x - y \mid u - v\rangle \geq 0. \tag{6.2.5}$$

A is locally bounded at $x \in \mathcal{H}$, if there exists $\delta \in \left]0, +\infty\right[$ such that $A(\mathcal{B}(x; \delta))$ is bounded, and A is locally bounded in $\varnothing \neq D \subset \mathcal{H}$ if, for every $x \in D$, A is locally bounded at x.

We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$. Let $f \in \Gamma_0(\mathcal{H})$. The Fenchel conjugate of f is defined by $f^* \colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$, which is a function in $\Gamma_0(\mathcal{H})$. The subdifferential of f is the maximally monotone operator

$$\partial f \colon x \mapsto \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H}) \;\; f(x) + \langle y - x \mid u\rangle \leq f(y)\big\}.$$

It turns out that $(\partial f)^{-1} = \partial f^*$ and that $\operatorname{zer}\partial f$ is the set of minimizers of f, which is denoted by $\arg\min_{x\in\mathcal{H}} f(x)$. We denote the proximity operator of f by

$$\operatorname{prox}_f \colon x \mapsto \underset{y\in\mathcal{H}}{\arg\min} \left(f(y) + \frac{1}{2}\|x - y\|^2\right). \tag{6.2.6}$$

We have $\operatorname{prox}_f = J_{\partial f}$. Moreover, it follows from [2, Theorem 14.3] that

$$(\forall \gamma > 0) \quad \operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{f^*/\gamma} \circ \mathrm{Id}/\gamma = \mathrm{Id}. \tag{6.2.7}$$

Given a non-empty closed convex set $C \subset \mathcal{H}$, we denote by P_C the projection onto C and by $\iota_C \in \Gamma_0(\mathcal{H})$ the indicator function of C, which takes the value 0 in C and $+\infty$ otherwise. For further properties of monotone operators, nonexpansive mappings, and convex analysis, the reader is referred to [2].

6.2.3 Main Result

We first study some properties of the monotone operators involved in Problem 6.2.1, which ensure the finite termination of the backtracking procedure in our method.

Lemma 6.2.2. *In the context of Problem 6.2.1, let z and y in $\mathcal{H}$, and define*

$$(\forall \gamma > 0) \qquad x_{z,y}(\gamma) = J_{\gamma A}(z - \gamma y) \quad \textit{and} \quad \varphi_{z,y}(\gamma) = \frac{\|z - x_{z,y}(\gamma)\|}{\gamma}. \tag{6.2.8}$$

Then, the following statements holds:

1. *$\varphi_{z,y}$ is nonincreasing.*
2. $(\forall z \in \operatorname{dom} A) \quad \lim_{\gamma \downarrow 0} \varphi_{z,y}(\gamma) = \|(A+y)^0 z\| = \min_{w \in Az+y} \|w\|.$
3. *Set*
$$\mathcal{C} = \big\{ z \in \mathcal{H} \;\big|\; \lim_{\gamma \downarrow 0} \varphi_{z,y}(\gamma) < +\infty \big\}. \tag{6.2.9}$$
Then, $\operatorname{dom} A \subset \mathcal{C} \subset \overline{\operatorname{dom}}\, A$.
4. *Suppose that one of the following holds:*
   (a) $z \in \mathcal{C}$.
   (b) $z \in \operatorname{dom} B_3 \setminus \mathcal{C}$, $y = (B_1 + B_2 + B_3)z$, *and B_3 is locally bounded at* $P_{\overline{\operatorname{dom}}\, A} z$.
   (c) $z \in \operatorname{dom} B_3 \setminus \mathcal{C}$, $y = (B_1 + B_2 + B_3)z$, *and* $\overline{\operatorname{dom}}\, A \subset \operatorname{int} \operatorname{dom} B_3$.

   Then, for every $\theta \in \left]0,1\right[$, there exists $\gamma(z) > 0$ such that, for every $\gamma \in \left]0, \gamma(z)\right]$,
$$\gamma \|B_3 z - B_3 x_{z,y}(\gamma)\| \leq \theta \|z - x_{z,y}(\gamma)\|. \tag{6.2.10}$$

Proof. Let $z \in \mathcal{H}$. Note that, if $z \in \operatorname{zer}(A+y)$,

$$\begin{aligned} (\forall \gamma > 0) \quad 0 \in Az + y \quad &\Leftrightarrow \quad z - \gamma y \in \gamma A z + z \\ &\Leftrightarrow \quad z = x_{z,y}(\gamma) \\ &\Leftrightarrow \quad \varphi_{z,y}(\gamma) = 0. \end{aligned} \tag{6.2.11}$$

In this case, 1, 2, and 4 are clear. Henceforth, assume $z \in \mathcal{H} \setminus \operatorname{zer}(A+y)$. It follows from (6.2.8) that

$$\frac{z - x_{z,y}(\gamma)}{\gamma} - y \in A x_{z,y}(\gamma). \tag{6.2.12}$$

1: For every γ_1 and γ_2 in $]0,+\infty[$, (6.2.12) and the monotonicity of A yield

$$\begin{aligned}0 &\leq \left\langle \frac{z-x_{z,y}(\gamma_1)}{\gamma_1} - \frac{z-x_{z,y}(\gamma_2)}{\gamma_2} \;\middle|\; x_{z,y}(\gamma_1)-x_{z,y}(\gamma_2)\right\rangle\\
&= \left\langle \frac{z-x_{z,y}(\gamma_1)}{\gamma_1} - \frac{z-x_{z,y}(\gamma_2)}{\gamma_2} \;\middle|\; x_{z,y}(\gamma_1)-z-(x_{z,y}(\gamma_2)-z)\right\rangle\\
&= -\frac{1}{\gamma_1}\|z-x_{z,y}(\gamma_1)\|^2 + \left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right)\langle z-x_{z,y}(\gamma_1) \mid z-x_{z,y}(\gamma_2)\rangle - \frac{1}{\gamma_2}\|z-x_{z,y}(\gamma_2)\|^2.\end{aligned}$$

Hence, we obtain

$$\begin{aligned}\gamma_1\varphi_{z,y}(\gamma_1)^2+\gamma_2\varphi_{z,y}(\gamma_2)^2 &\leq (\gamma_1+\gamma_2)\left\langle \frac{z-x_{z,y}(\gamma_1)}{\gamma_1} \;\middle|\; \frac{z-x_{z,y}(\gamma_2)}{\gamma_2}\right\rangle\\
&\leq \frac{(\gamma_1+\gamma_2)}{2}(\varphi_{z,y}(\gamma_1)^2+\varphi_{z,y}(\gamma_2)^2),\end{aligned}$$

which yields $(\gamma_1-\gamma_2)(\varphi_{z,y}(\gamma_1)^2-\varphi_{z,y}(\gamma_2)^2)\leq 0$ and 1 follows.

2: It follows from the monotonicity of A and (6.2.12) that, for every $w \in Az+y$ and $\gamma \in\,]0,+\infty[$,

$$0 \leq \left\langle \frac{z-x_{z,y}(\gamma)}{\gamma} - w \;\middle|\; x_{z,y}(\gamma)-z\right\rangle,$$

which yields

$$\begin{aligned}\frac{1}{\gamma}\|z-x_{z,y}(\gamma)\|^2 &\leq \langle w \mid z-x_{z,y}(\gamma)\rangle\\
&\leq \|w\|\|z-x_{z,y}(\gamma)\|.\end{aligned}\tag{6.2.13}$$

Thus, $\varphi_{z,y}(\gamma) \leq \|w\|$. Therefore, since [2, Proposition 20.36] implies that, for every $z \in \operatorname{dom} A$, $Az+y$ is nonempty, closed, and convex, [2, Theorem 11.10] yields

$$(\forall \gamma \in\,]0,+\infty[) \qquad \varphi_{z,y}(\gamma) \leq \min_{w\in Az+y}\|w\|.\tag{6.2.14}$$

Hence, since $\varphi_{z,y}\geq 0$, 1 implies that $\lim_{\gamma\downarrow 0}\varphi_{z,y}(\gamma)$ exists. In turn, since $z \in \mathcal{H}\setminus\operatorname{zer}(A+y)$, it follows from (6.2.14) and (6.2.11) that

$$0 < \varphi_{z,y}(1) \leq \lim_{\gamma\downarrow 0}\varphi_{z,y}(\gamma) \leq \min_{w\in Az+y}\|w\|,\tag{6.2.15}$$

which, in view of (6.2.2), implies

$$\lim_{\gamma\downarrow 0} x_{z,y}(\gamma) = z.\tag{6.2.16}$$

Since $(\varphi_{z,y}(\gamma))_{(\gamma>0)}$ is bounded, by [2, Lemma 2.45], there exists a sequence $(\gamma_k)_{k\in\mathbb{N}} \subset]0,+\infty[$ and $\overline{w} \in \mathcal{H}$ such that $\gamma_k \downarrow 0$ and $\frac{z-x_{z,y}(\gamma_k)}{\gamma_k} \rightharpoonup \overline{w}$ as $k \to +\infty$. Therefore, (6.2.12), (6.2.16), and [2, Proposition 20.38(i)] imply $\overline{w} \in Az + y$. Hence, noting that

$$\begin{aligned}(\varphi_{z,y}(\gamma_k))^2 &= \left\|\frac{z - x_{z,y}(\gamma_k)}{\gamma_k} - \overline{w}\right\|^2 + \|\overline{w}\|^2 + 2\left\langle \frac{z - x_{z,y}(\gamma_k)}{\gamma_k} - \overline{w} \;\middle|\; \overline{w}\right\rangle \\ &\geq \|\overline{w}\|^2 + 2\left\langle \frac{z - x_{z,y}(\gamma_k)}{\gamma_k} - \overline{w} \;\middle|\; \overline{w}\right\rangle, \end{aligned} \tag{6.2.17}$$

we deduce

$$\lim_{\gamma\downarrow 0}(\varphi_{z,y}(\gamma_k))^2 = \lim_{k\to+\infty}(\varphi_{z,y}(\gamma_k))^2 \geq \|\overline{w}\|^2 \geq \min_{w\in Az+y}\|w\|^2.$$

Therefore, we obtain from (6.2.15) that

$$\lim_{\gamma\downarrow 0}\varphi_{z,y}(\gamma) = \min_{w\in Az+y}\|w\|, \tag{6.2.18}$$

and 2 follows.

3: It follows from (6.2.9) and 2 that $\operatorname{dom} A \subset \mathcal{C}$. Let $z \in \mathcal{H} \setminus \overline{\operatorname{dom} A}$. The firmly nonexpansiveness of $J_{\gamma A}$ [2, Proposition 23.8(ii)] implies

$$\begin{aligned}\|x_{z,y}(\gamma) - P_{\overline{\operatorname{dom} A}}z\| &= \|J_{\gamma A}(z-\gamma y) - J_{\gamma A}z + J_{\gamma A}z - P_{\overline{\operatorname{dom} A}}z\| \\ &\leq \|J_{\gamma A}(z-\gamma y) - J_{\gamma A}z\| + \|J_{\gamma A}z - P_{\overline{\operatorname{dom} A}}z\| \\ &\leq \gamma\|y\| + \|J_{\gamma A}z - P_{\overline{\operatorname{dom} A}}z\|. \end{aligned} \tag{6.2.19}$$

Hence, by taking $\gamma \downarrow 0$ in (6.2.19) we conclude from [2, Theorem 23.48] that $x_{z,y}(\gamma) \to P_{\overline{\operatorname{dom} A}}z$ as $\gamma \downarrow 0$. Then, by the continuity of the norm and $z \notin \overline{\operatorname{dom} A}$, we deduce

$$\lim_{\gamma\downarrow 0}\|z - x_{z,y}(\gamma)\| = \|z - P_{\overline{\operatorname{dom} A}}z\| > 0.$$

Therefore, $\varphi_{z,y}(\gamma) = \|z - x_{z,y}(\gamma)\|/\gamma \to +\infty$ as $\gamma \downarrow 0$ and, hence, it follows from (6.2.9) that $z \in \mathcal{H} \setminus \mathcal{C}$.

4a: If $z \in \mathcal{C}$, it follows from 1 that

$$0 < \varphi_{z,y}(1) \leq \lim_{\gamma\downarrow 0}\varphi_{z,y}(\gamma) < +\infty. \tag{6.2.20}$$

Therefore $\lim_{\gamma\downarrow 0} x_{z,y} = z$ and the continuity of B_3 implies

$$\lim_{\gamma\downarrow 0} B_3 x_{z,y} = B_3 z. \tag{6.2.21}$$

The result follows from (6.2.20) and (6.2.21).

4b: Set $B = B_1 + B_2 + B_3$ and let $p = P_{\overline{\operatorname{dom} A}} z$. Since B_3 is locally bounded at p, there exists $\delta_p \in]0, +\infty[$ such that $B_3(\mathcal{B}(p; \delta_p))$ is bounded. Now since $y = Bz$ and

$$\frac{z - J_{\gamma A} z}{\gamma} \in A J_{\gamma A} z, \tag{6.2.22}$$

(6.2.12), $y = Bz$, and the it follows from (6.2.12) and the monotonicity of A that

$$\begin{aligned} 0 \leq & \left\langle \frac{z - x_{z,y}(\gamma)}{\gamma} - Bz - \frac{z - J_{\gamma A} z}{\gamma} \;\middle|\; x_{z,y}(\gamma) - J_{\gamma A} z \right\rangle \\ = & -\frac{1}{\gamma} \|x_{z,y}(\gamma) - J_{\gamma A} z\|^2 + \langle Bz \mid J_{\gamma A} z - x_{z,y}(\gamma) \rangle \\ \leq & -\frac{1}{\gamma} \|x_{z,y}(\gamma) - J_{\gamma A} z\|^2 + \|Bz\| \|J_{\gamma A} z - x_{z,y}(\gamma)\|. \end{aligned}$$

Hence, we obtain

$$\|x_{z,y}(\gamma) - J_{\gamma A} z\| \leq \gamma \|Bz\|. \tag{6.2.23}$$

Additionally, by [2, Theorem 23.48], there exists γ_1 such that, for every $\gamma < \gamma_1$, $\|J_{\gamma A} z - p\| \leq \delta_p / 2$. By defining

$$\overline{\gamma} := \begin{cases} \gamma_1, & \text{if } Bz = 0; \\ \min\{\delta_p / (2\|Bz\|), \gamma_1\}, & \text{if } Bz \neq 0, \end{cases} \tag{6.2.24}$$

it follows from (6.2.23) that, for every $\gamma < \overline{\gamma}$,

$$\begin{aligned} \|x_{z,y}(\gamma) - p\| &\leq \|x_{z,y}(\gamma) - J_{\gamma A} z\| + \|J_{\gamma A} z - p\| \\ &\leq \gamma \|Bz\| + \frac{\delta_p}{2} \\ &< \delta_p, \end{aligned}$$

which yields $(x_{z,y}(\gamma))_{0 < \gamma \leq \overline{\gamma}} \subset B(p, \delta_p)$. Therefore, since $z \in \mathcal{H} \setminus \mathcal{C}$ implies $\varphi_{z,\gamma}(\gamma) \to +\infty$ as $\gamma \downarrow 0$ and $(\|B_3 z - B_3 x_{z,y}(\gamma)\|)_{0 < \gamma \leq \overline{\gamma}}$ is bounded, the result follows.

4c: Since $p = P_{\overline{\operatorname{dom} A}} z \notin \operatorname{bdry} \operatorname{dom} B_3$, B_3 is locally bounded in p [2, Theorem 21.18], and the result follows from 4b. ☐

Remark 6.2.3. *1. In the case $B_2 = 0$, by setting $y = (B_1 + B_3) z$ in Lemma 6.2.2(1)&(2), we recover [12, Lemma 2.2(1)].*

2. Realizing that [12, Lemma 2.2(2)] is valid for every $z \in \operatorname{dom} A$, it is a particular case of Lemma 6.2.2(3)&(4a).

Now we state our main result.

Theorem 6.2.4. *In the context of Problem 6.2.1, suppose that one of the following holds:*

1. $X \subset \operatorname{dom} A$.
2. $\overline{\operatorname{dom}}\, A \subset \operatorname{dom} B_3$ *and* B_3 *is locally bounded in* $\operatorname{dom} B_3$.
3. $\overline{\operatorname{dom}}\, A \subset \operatorname{int}\operatorname{dom} B_3$.

Let $\varepsilon \in \left]0,1\right[$, *set* $\rho = \min\{2\beta\varepsilon, \sqrt{1-\varepsilon}/L\}$, *let* $\sigma \in \left]0,1\right[$, *let* $\theta \in \left]0, \sqrt{1-\varepsilon} - L\rho\sigma\right[$, *let* $z_0 \in \operatorname{dom} B_3$, *and consider the sequence* $(z_n)_{n\in\mathbb{N}}$ *defined by the recurrence*

$$(\forall n \in \mathbb{N}) \qquad \left\lfloor \begin{array}{l} x_n = J_{\gamma_n A}(z_n - \gamma_n(B_1 + B_2 + B_3)z_n) \\ z_{n+1} = P_X(x_n + \gamma_n(B_2 + B_3)z_n - \gamma_n(B_2 + B_3)x_n), \end{array} \right. \tag{6.2.25}$$

where, for every $n \in \mathbb{N}$, γ_n *is the largest* $\gamma \in \{\rho\sigma, \rho\sigma^2, \rho\sigma^3, \cdots\}$ *satisfying*

$$\gamma\|B_3 z_n - B_3 J_{\gamma A}(z_n - \gamma(B_1+B_2+B_3)z_n)\| \le \theta\|z_n - J_{\gamma A}(z_n - \gamma(B_1+B_2+B_3)z_n)\|. \tag{6.2.26}$$

Moreover, assume that at least one of the following additional statements hold:

(i) $\liminf\limits_{n\to\infty} \gamma_n = \delta > 0$.

(ii) B_3 *is uniformly continuous in any weakly compact subset of* $\overline{\operatorname{conv}}(\operatorname{dom} A \cup X)$.

Then, $(z_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 6.2.1.*

Proof. Set $B = B_1 + B_2 + B_3$ and fix $n \in \mathbb{N}$. If $z_n \in \mathcal{C}$, where $\mathcal{C}$ is defined in (6.2.9), then γ_n is well defined in view of Lemma 6.2.2(4a). In particular, if 1 holds, γ_n is well defined in view of Lemma 6.2.2(3). Now suppose that $z_n \in \mathcal{H} \setminus \mathcal{C}$. If $n = 0$, it is clear that $z_0 \in \operatorname{dom} B_3 \setminus \mathcal{C}$. Otherwise, since $X \subset \operatorname{dom} B_3$, we have $z_n \in \operatorname{dom} B_3 \setminus \mathcal{C}$. Now, if we assume 2, then B_3 is locally bounded in $P_{\overline{\operatorname{dom}}\, A} z_n \in \overline{\operatorname{dom}}\, A$ and γ_n is well defined from Lemma 6.2.2(4b). Similarly, if we assume 3, γ_n is well defined from Lemma 6.2.2(4c).

Now, let $z^* \in \operatorname{zer}(A + B) \cap X$. Note that, by the maximal monotonicity of $A + B_3$, the full domain of B_1 and B_2, and [2, Corollary 25.5(i)], it follows that $A + B_2 + B_3$ and $A + B$ are maximally monotone. Then, since $B_2 + B_3$ is continuous and single valued in $\operatorname{dom}(B_2 + B_3) = \operatorname{dom} B_3 \supset \operatorname{dom} A \cup X$ and B_1 is β-cocoercive, it follows from [12, Proposition 2.1(1)&(2)] that, for every $n \in \mathbb{N}$, we have

$$\begin{aligned} \|z_{n+1} - z^*\|^2 \le & \|z_n - z^*\|^2 - (1-\varepsilon)\|z_n - x_n\|^2 + \gamma_n^2\|(B_2 + B_3)z_n - (B_2 + B_3)x_n\|^2 \\ & - \frac{\gamma_n}{\varepsilon}(2\beta\varepsilon - \gamma_n)\|B_1 z_n - B_1 z^*\|^2. \end{aligned} \tag{6.2.27}$$

Note that the Lipschitz property of B_2 and (6.2.26) yield

$$\begin{aligned} \gamma_n^2\|(B_2 + B_3)z_n - (B_2 + B_3)x_n\|^2 &\le (L\gamma_n\|z_n - x_n\| + \gamma_n\|B_3 z_n - B_3 x_n\|)^2 \\ &\le (L\gamma_n + \theta)^2\|z_n - x_n\|^2 \\ &\le (L\rho\sigma + \theta)^2\|z_n - x_n\|^2. \end{aligned} \tag{6.2.28}$$

Hence, it follows from (6.2.27) and (6.2.28) that

$$(\forall n\in\mathbb{N})\quad \|z_{n+1}-z^*\|^2 \leq \|z_n-z^*\|^2-((1-\varepsilon)-(L\rho\sigma+\theta)^2)\|z_n-x_n\|^2 \\ -\frac{\gamma_n}{\varepsilon}(2\beta\varepsilon-\gamma_n)\|B_1z_n-B_1z^*\|^2.$$

Therefore, since $(1-\varepsilon-(L\rho\sigma+\theta)^2)\geq 2\beta\varepsilon(1-\sigma)>0$ and $(2\beta\varepsilon-\gamma_n)>0$, [15, Lemma 3.1(i)] implies that $(\|z_n-z^*\|)_{n\in\mathbb{N}}$ is a convergent sequence and

$$z_n-x_n\to 0. \tag{6.2.29}$$

Let $z\in\mathcal{H}$ be a weak limit point of the subsequence $(z_n)_{n\in K}$ for some $K\subset\mathbb{N}$. Then z is also a weak limit point of $(x_n)_{n\in K}$ in view of (6.2.29). Since X is closed and convex, and $(z_n)_{n\in K}$ is a sequence in X we conclude that $z\in X$. Let us prove that $z\in\operatorname{zer}(A+B)$.

(i): Assume that $\liminf_{n\to+\infty}\gamma_n=\delta>0$. Then, there exists $n_0\in\mathbb{N}$ such that $\inf_{n\geq n_0}\gamma_n\geq\delta$. Hence, (6.2.26), (6.2.25), the Lipschitz continuity of B_2, and the cocoercivity of B_1 yield

$$\begin{aligned}(\forall n\geq n_0)\quad \|Bz_n-Bx_n\| &\leq \|B_1z_n-B_1x_n\|+\|B_2z_n-B_2x_n\|+\|B_3z_n-B_3x_n\|\\ &\leq\left(\frac{1}{\beta}+L+\frac{\theta}{\delta}\right)\|z_n-x_n\|,\end{aligned} \tag{6.2.30}$$

which implies $Bz_n-Bx_n\to 0$ in view of (6.2.29). Hence, it follows from (6.2.25) that

$$(\forall n\in\mathbb{N})\quad u_n:=\frac{z_n-x_n}{\gamma_n}-Bz_n+Bx_n\in(A+B)x_n, \tag{6.2.31}$$

and (6.2.29) and $\liminf_{n\to+\infty}\gamma_n=\delta>0$ imply that $u_n\to 0$. Therefore, since $x_n\rightharpoonup z$, $n\in K$, the weak-strong closure of the graph of the maximally monotone operator $A+B$ and (6.2.31) yield $z\in\operatorname{zer}(A+B)$. The convergence follows from [2, Lemma 2.47].

(ii): Without loss of generality, suppose that $\lim_{n\to\infty,n\in K}\gamma_n=0$. Our choice of γ_n guarantee that, for every $n\in K$, we have

$$\widetilde{\gamma}_n\|B_3z_n-B_3J_{\widetilde{\gamma}_nA}(z_n-\widetilde{\gamma}_nBz_n)\|>\theta\|z_n-J_{\widetilde{\gamma}_nA}(z_n-\widetilde{\gamma}_nBz_n)\|. \tag{6.2.32}$$

where $\widetilde{\gamma}_n:=\frac{\gamma_n}{\sigma}>\gamma_n$. Now, by the nonincreasing property of $\gamma\mapsto\frac{1}{\gamma}\|z-J_{\gamma A}(z-\gamma Bz)\|$ provided by Lemma 6.2.2(1) with to $y=Bz$, we have

$$\frac{1}{\widetilde{\gamma}_n}\|z_n-J_{\widetilde{\gamma}_nA}(z_n-\widetilde{\gamma}_nBz_n)\|\leq\frac{1}{\gamma_n}\|z_n-J_{\gamma_nA}(z_n-\gamma_nBz_n)\|, \tag{6.2.33}$$

which is equivalent to

$$\|z_n-J_{\widetilde{\gamma}_nA}(z_n-\widetilde{\gamma}_nBz_n)\|\leq\frac{1}{\sigma}\|z_n-x_n\|. \tag{6.2.34}$$

Then $z_n - x_n \to 0$ implies that $z_n - J_{\widetilde{\gamma}_n A}(z_n - \widetilde{\gamma}_n B z_n) \to 0$ as $n \to \infty$, $n \in K$. Therefore, since $z_n \rightharpoonup z$, $n \in K$, the sequence $(\widetilde{x}_n)_{n\in K}$ defined by

$$(\forall n \in K) \quad \widetilde{x}_n = J_{\widetilde{\gamma}_n A}(z_n - \widetilde{\gamma}_n B z_n) \tag{6.2.35}$$

satisfies $\widetilde{x}_n \rightharpoonup z$ as $n \to \infty$, $n \in K$. Furthermore, (6.2.35) yields

$$\frac{z_n - \widetilde{x}_n}{\widetilde{\gamma}_n} + B\widetilde{x}_n - Bz_n \in (A+B)\widetilde{x}_n. \tag{6.2.36}$$

Since $\{z\} \cup \bigcup_{n\in K}[\widetilde{x}_n, z_n]$ is a weakly compact subset of $\overline{\operatorname{conv}}(\operatorname{dom} A \cup X)$ [35, Lemma 3.2], it follows from the uniform continuity of B_3 that

$$B_3 z_n - B_3 \widetilde{x}_n \to 0 \text{ as } n \to \infty,\ n \in K. \tag{6.2.37}$$

Hence, by (6.2.32), we obtain $\frac{z_n - \widetilde{x}_n}{\widetilde{\gamma}_n} \to 0$ as $n \to \infty$, $n \in K$. Since $B_1 + B_2$ is Lipschitz continuous, (6.2.37) implies that

$$B\widetilde{x}_n - Bz_n \to 0 \text{ as } n \to \infty,\ n \in K. \tag{6.2.38}$$

Altogether, the convergence follows, as in the case (i), from (6.2.36), the weak-strong closure of the graph of the maximally monotone operator $A+B$, and [2, Lemma 2.47]. ☐

Remark 6.2.5. *1. In Theorem 6.2.4, if $B_3 = 0$, we have $\operatorname{dom} B_3 = \mathcal{H}$ and, for all $n \in \mathbb{N}$, $\gamma_n = \sigma\rho = \sigma \min\{2\beta\varepsilon, \sqrt{1-\varepsilon}/L\}$. Since, in this case $(\gamma_n)_{n\in\mathbb{N}}$ is constant, the largest step-size is obtained by taking $\varepsilon = \varepsilon(L,\beta) = 2/(1+\sqrt{1+16\beta^2L^2})$, which satisfies $2\beta\varepsilon = \sqrt{1-\varepsilon}/L = \chi(L,\beta)$, where*

$$\chi(L,\beta) = \frac{4\beta}{1+\sqrt{1+16\beta^2L^2}}, \tag{6.2.39}$$

and $\gamma_n \equiv \gamma = \sigma\chi(L,\beta) \in \left]0, \chi(L,\beta)\right[$. Hence, we recover the result in [12, Theorem 2.3(1)] for constant step-sizes. Additionally, if $B_2 = 0$ and $X = \mathcal{H}$, we have $\varepsilon(L,\beta) \to 1$ and $\chi(L,\beta) \to 2\beta$ as $L \to 0$ and $\gamma_n \equiv 2\beta\sigma \in \left]0, 2\beta\right[$, recovering the the forward backward algorithm [29]. On the other hand, if $B_1 = 0$, we have $\chi(L,\beta) \to 1/L$ as $\beta \to \infty$ and $\gamma_n \equiv \sigma/L \in \left]0, 1/L\right[$, recovering the result in [38] for constant step-sizes.

2. Suppose that $B_2 = 0$ and $X \subset \operatorname{dom} A$. Then by taking $L \to 0$, we have $\rho \to 2\beta\varepsilon$ and $\theta \in \left]0, \sqrt{1-\varepsilon}\right[$. Hence, Theorem 6.2.4 recovers [12, Theorem 2.3(2)] noting that the uniform continuity in weakly compact subsets of $\overline{\operatorname{conv}}(\operatorname{dom} A \cup X) = \overline{\operatorname{dom} A}$ is needed. We hence generalize [12, Theorem 2.3(1)&(2)] to the case when $X \not\subset \operatorname{dom} A$.

6.2.4 Application to Convex Optimization with Nonlinear Constraints

In this section we consider the following optimization problem.

Problem 6.2.6. *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{G})$, let $h : \mathcal{H} \to \mathbb{R}$ be a convex Gâteaux differentiable function such that ∇h is β^{-1}-Lipschitzian for some $\beta \in \left]0, +\infty\right[$, let $M : \mathcal{H} \to \mathcal{G}$ be a bounded linear operator, and let $e : \mathcal{H} \to \left]-\infty, +\infty\right]^p : x \mapsto (e_i(x))_{1 \leq i \leq p}$ be such that, for every $i \in \{1, \ldots, p\}$, e_i is convex and Gâteaux differentiable in* $\operatorname{int}\operatorname{dom} e_i$, $\operatorname{dom} e_i$ *is closed, $\cap_{i=1}^{p} \operatorname{intdom} e_i \neq \varnothing$, and $\operatorname{dom}\partial f \subset \cap_{i=1}^{n} \operatorname{intdom} e_i$. The problem is to*

$$\min_{e(x) \in \left]-\infty, 0\right]^p} f(x) + g(Mx) + h(x), \tag{6.2.40}$$

and we assume that solutions exist.

The Lagrangian associated to (6.2.40) is

$$\mathcal{L} : \mathcal{H} \times \mathbb{R}^p : (x, v) \to \left]-\infty, +\infty\right] \mapsto f(x) + g(Mx) + h(x) + e(x) \cdot v - \iota_{[0,+\infty[^p}(v). \tag{6.2.41}$$

In view of [2, Corollary 19.30], if $(\hat{x}, \hat{v}) \in \mathcal{H} \times \mathbb{R}^p$ is a *saddle point* of (6.2.41), $\hat{x}$ is a solution to (6.2.40) and by [2, Corollary 19.30(v), Theorem 16.3, Theorem 16.47, Example 16.13, & Example 6.42(i)] we deduce that

$$\begin{cases} 0 \in \partial(f + g \circ M + h)(\hat{x}) + \sum_{i=1}^{p} \hat{v}_i \nabla e_i(\hat{x}) \\ 0 \in N_{[0,+\infty[^p}(\hat{v}) - e(\hat{x}). \end{cases} \tag{6.2.42}$$

Under standard qualification conditions, as $0 \in \operatorname{sri}(\operatorname{dom} g - M(\operatorname{dom} f))$, there exists $\hat{u} \in \mathcal{G}$ such that (6.2.42) reduces to

$$\begin{cases} 0 \in \partial f(\hat{x}) + M^*\hat{u} + \nabla h(\hat{x}) + \sum_{i=1}^{p} \hat{v}_i \nabla e_i(\hat{x}) \\ 0 \in \partial g^*(\hat{u}) - M\hat{x} \\ 0 \in N_{[0,+\infty[^p}(\hat{v}) - e(\hat{x}), \end{cases}$$

which is equivalent to

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(\hat{x}) \\ \partial g^*(\hat{u}) \\ N_{[0,+\infty[^p}(\hat{v}) \end{pmatrix} + \begin{pmatrix} \nabla h(\hat{x}) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} M^*\hat{u} \\ -M\hat{x} \\ 0 \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^{p} \hat{v}_i \nabla e_i(\hat{x}) \\ 0 \\ -e(\hat{x}) \end{pmatrix}. \tag{6.2.43}$$

Proposition 6.2.7. *In the context of Problem 6.2.6 let $X = X_1 \times X_2 \times X_3 \subset \operatorname{dom}\partial f \times \operatorname{dom}\partial g^* \times [0, +\infty[^p$ be nonempty, closed, and convex, and define the operator*

$$B_3 : \mathcal{H} \times \mathcal{G} \times \mathbb{R} \to 2^{\mathcal{H} \times \mathcal{G} \times \mathbb{R}}$$
$$(x, u, v) \mapsto \begin{cases} \left\{\left(\sum_{i=1}^{p} v_i \nabla e_i(x), 0, -e(x)\right)\right\}, & \text{if } v \in [0, +\infty[^p \text{ and } x \in \bigcap_{i=1}^{p} \operatorname{intdom} e_i, \\ \varnothing, & \text{otherwise.} \end{cases} \tag{6.2.44}$$

Then, the following hold:

1. *B_3 is maximally monotone.*
2. *Suppose that one of the following holds:*
   (a) *$(\nabla e_i)_{1\leq i\leq p}$ are bounded and uniformly continuous in every weakly compact subset of $\overline{\operatorname{dom}\partial f}$.*
   (b) *$\mathcal{H}$ is finite dimensional and $(\nabla e_i)_{1\leq i\leq p}$ are continuous in every compact subset of $\overline{\operatorname{dom}\partial f}$.*

   Then, B_3 is uniformly continuous in every compact subset of $\overline{\operatorname{dom}\partial f}\times\overline{\operatorname{dom}\partial g^*}\times[0,+\infty[^p$.

Proof. 1: Consider the saddle-function

$$\begin{aligned}\ell\colon \mathcal{H}\times\mathcal{G}\times\mathbb{R}^p&\to[-\infty,+\infty]\\ (x,u,v)&\mapsto\begin{cases} e(x)\cdot v, & \text{if } v\in[0,+\infty[^p \text{ and } x\in\bigcap_{i=1}^p\operatorname{dom}e_i;\\ +\infty, & \text{if } v\in[0,+\infty[^p \text{ and } x\notin\bigcap_{i=1}^p\operatorname{dom}e_i;\\ -\infty, & \text{if } v\notin[0,+\infty[^p.\end{cases}\end{aligned}$$

Note that, if $v\in[0,+\infty[^p$,

$$\ell\colon (x,u,v)\mapsto\begin{cases} e(x)\cdot v, & \text{if } x\in\bigcap_{i=1}^p\operatorname{dom}e_i;\\ +\infty, & \text{otherwise}\end{cases}\tag{6.2.45}$$

and, if $v\notin[0,+\infty[^p$, $\ell(\cdot,\cdot,v)\equiv-\infty$. Hence, for every $v\in\mathbb{R}^p$, $\ell(\cdot,\cdot,v)$ is lower-semicontinuous. Additionally, if $x\in\bigcap_{i=1}^p\operatorname{dom}e_i$, we have

$$-\ell\colon (x,u,v)\mapsto\begin{cases} -e(x)\cdot v, & \text{if } v\in[0,+\infty[^p;\\ +\infty, & \text{otherwise}\end{cases}\tag{6.2.46}$$

and, if $x\notin\bigcap_{i=1}^p\operatorname{dom}e_i$

$$-\ell\colon (x,u,v)\mapsto\begin{cases} -\infty & \text{if } v\in[0,+\infty[^p;\\ +\infty, & \text{otherwise.}\end{cases}\tag{6.2.47}$$

Therefore, for every $(x,u)\in\mathcal{H}\times\mathcal{G}$, $-\ell(x,u,\cdot)$ is lower-semicontinuous. Furthermore,

$$(\forall(x,u)\in\mathcal{H}\times\mathcal{G})(\forall v\in\mathbb{R}^p)\quad B_3(x,u,v)=\partial\ell(\cdot,\cdot,v)(x,u)\times\partial(-\ell(x,u,\cdot))(v).\tag{6.2.48}$$

The result follows from [33, Corollary 1].

2: First, assume 2a. Let $Y = Y_1 \times Y_2 \times Y_3 \subset \overline{\operatorname{dom}}\,\partial f \times \overline{\operatorname{dom}}\,\partial g^* \times [0,+\infty[^p$ be a weakly compact set. Let $\boldsymbol{x} = (x_1, u_1, v_1)$ and $\boldsymbol{y} = (x_2, u_2, v_2)$ in Y, fix $i \in \{1 \ldots, p\}$, define $\rho_i(t)\colon\ [0,1] \to \mathbb{R}\colon t \mapsto e_i(x_1 + t(x_2 - x_1))$ which is differentiable in $]0,1[$. Since Y_1 is weakly compact, by [2, Theorem 3.37], $\operatorname{conv} Y_1$ is also weakly compact. Moreover, we deduce from the boundedness of ∇e_i in $\operatorname{conv} Y_1 \subset \overline{\operatorname{dom}}\,\partial f$ that exists $K_i > 0$ such that $\sup_{x\in\operatorname{conv} Y_1} \|\nabla e_i(x)\| \le K_i$. Therefore, since $\rho_i'\colon t \mapsto \langle \nabla e_i(x_1 + t(x_2 - x_1)) \mid x_2 - x_1\rangle$, we obtain

$$\begin{aligned}
|e_i(x_2) - e_i(x_1)| &= |\rho_i(1) - \rho_i(0)| \\
&= \left|\int_0^1 \rho_i'(t)dt\right| \\
&= \left|\int_0^1 \langle \nabla e_i(x_1 + t(x_2 - x_1)) \mid x_2 - x_1\rangle dt\right| \\
&\le \int_0^1 \|\nabla e_i(x_1 + t(x_2 - x_1))\|\|x_2 - x_1\| dt \\
&\le K_i\|x_2 - x_1\|.
\end{aligned}$$

Thus, we conclude $|e_i(x_1) - e_i(x_2)| \le K_i\|x_2 - x_1\|$ and therefore

$$\|e(x_2) - e(x_1)\| \le K\|x_2 - x_1\| \tag{6.2.49}$$

where $K = \sqrt{\sum_{i=1}^p K_i^2}$. Since Y is weakly compact, it is bounded [2, Lemma 2.36] and there exists $V > 0$ such that $\sup_{v\in Y_3} \|v\| \le V$ for. Let $\varepsilon > 0$. The uniform continuity of $(\nabla e_i)_{1\le i\le p}$ implies the existence of $\delta > 0$ such that

$$(\forall i \in \{1, \ldots, p\})(\forall (z_1, z_2) \in Y_2^2) \quad \|z_1 - z_2\| < \delta \Rightarrow \|\nabla e_i(z_1) - \nabla e_i(z_2)\|^2 \le \frac{\varepsilon^2}{4pV^2}. \tag{6.2.50}$$

Now suppose that

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 \le \min\left\{\frac{\varepsilon^2}{4pK^2}, \delta^2\right\}.$$

Then, (6.2.49), the convexity of $\|\cdot\|^2$, and (6.2.50) imply

$$
\begin{aligned}
\|B_3\boldsymbol{x} - B_3\boldsymbol{y}\|^2 &= \left\|\sum_{i=1}^{p} v_{1,i}\nabla e_i(x_1) - v_{2,i}\nabla e_i(x_2)\right\|^2 + \|e(x_1) - e(x_2)\|^2 \\
&\leq 2\left\|\sum_{i=1}^{p}(v_{1,i} - v_{2,i})\nabla e_i(x_1)\right\|^2 + 2\left\|\sum_{i=1}^{p} v_{2,i}(\nabla e_i(x_1) - \nabla e_i(x_2))\right\|^2 + K^2\|x_1 - x_2\|^2 \\
&\leq 2p\sum_{i=1}^{p}|v_{1,i} - v_{2,i}|^2\,\|\nabla e_i(x_1)\|^2 + 2p\sum_{i=1}^{p}|v_{2,i}|^2\,\|\nabla e_i(x_1) - \nabla e_i(x_2)\|^2 + K^2\|x_1 - x_2\|^2 \\
&\leq 2pK^2\,\|v_1 - v_2\|^2 + 2pV^2\sum_{i=1}^{p}\|\nabla e_i(x_1) - \nabla e_i(x_2)\|^2 + K^2\|x_1 - x_2\|^2 \\
&\leq 2pK^2\,\|\boldsymbol{x} - \boldsymbol{y}\|^2 + \frac{\varepsilon^2}{2} \\
&\leq \varepsilon^2.
\end{aligned}
$$

Therefore B_3 is uniformly continuous in Y.

Now, assume 2b. Since $\mathcal{H}$ is finite dimensional the weak and strong topologies coincide [2, Fact 2.33]. Hence, since ∇e_i is continuous, it is bounded and uniformly continuous in every compact subset of X. The result follows from 2a. ☐

Remark 6.2.8. *Note that if, for every $i \in \{1,\ldots,p\}$, ∇e_i is bounded and uniformly continuous in every weakly compact subset of* $\operatorname{dom} f$*, by Proposition 6.2.7, and* $\overline{\operatorname{dom}}\,\partial f \subset \operatorname{dom} f$*, B_3 is uniformly continuous in every compact subset of* $\overline{\operatorname{dom}}\,\partial f \times \overline{\operatorname{dom}}\,\partial g^* \times [0,+\infty[^p$.

Proposition 6.2.9. *In the context of Problem 6.2.6, assume that* $0 \in \operatorname{sri}(\operatorname{dom} g - M(\operatorname{dom} f))$*, let $X = X_1 \times X_2 \times X_3 \subset \operatorname{dom}\partial f \times \operatorname{dom}\partial g^* \times [0,+\infty[^p$ be nonempty, closed, and convex, let $\varepsilon \in \,]0,1[$, set $\rho = \min\{2\beta\varepsilon, \sqrt{1-\varepsilon}/\|M\|\}$, let $\sigma \in \,]0,1[$, let $\theta \in \left]0, \sqrt{1-\varepsilon} - \|M\|\rho\sigma\right[$. For every $\boldsymbol{z} = (z^1, z^2, z^3) \in \mathcal{H} \times \mathcal{G} \times \mathbb{R}^p$ define $\boldsymbol{\Phi}_{\boldsymbol{z}} \colon \gamma \mapsto (\Phi^1_{\boldsymbol{z}}(\gamma), \Phi^2_{\boldsymbol{z}}(\gamma), \Phi^3_{\boldsymbol{z}}(\gamma))$, where*

$$
\begin{aligned}
&\Phi^1_{\boldsymbol{z}} \colon \gamma \mapsto \operatorname{prox}_{\gamma f}\left(z^1 - \gamma\left(\nabla h(z^1) + M^* z^2 + \sum_{i=1}^{p} z^3_i \nabla e_i(z^1)\right)\right) \\
&\Phi^2_{\boldsymbol{z}} \colon \gamma \mapsto \operatorname{prox}_{\gamma g^*}(z^2 + \gamma M z^1) \\
&\Phi^3_{\boldsymbol{z}} \colon \gamma \mapsto P_{[0,+\infty[^p}\big(z^3 + \gamma e(z^1)\big).
\end{aligned}
\tag{6.2.51}
$$

Let $\boldsymbol{z}_0 = (z_0^1, z_0^2, z_0^3) \in \mathcal{H} \times \mathcal{G} \times \mathbb{R}^p$ *and consider the recurrence*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x_n^1 = \Phi^1_{\boldsymbol{z}_n}(\gamma_n) \\ x_n^2 = \Phi^2_{\boldsymbol{z}_n}(\gamma_n) \\ x_n^3 = \Phi^3_{\boldsymbol{z}_n}(\gamma_n) \\ z_{n+1}^1 = P_{X_1}\big(x_n^1 + \gamma_n(M^* z_n^2 + \sum_{i=1}^p z_{n,i}^3 \nabla e_i(z_n^1)) - \gamma_n(M^* x_n^2 + \sum_{i=1}^p x_{n,i}^3 \nabla e_i(x_n^1))\big) \\ z_{n+1}^2 = P_{X_2}(x_n^2 - \gamma_n M z_n^1 + \gamma_n M x_n^1) \\ z_{n+1}^3 = P_{X_3}(x_n^3 - \gamma_n e(z_n^1) + \gamma_n e(x_n^1)) \\ \boldsymbol{z}_{n+1} = (z_{n+1}^1, z_{n+1}^2, z_{n+1}^3), \end{array} \right. \tag{6.2.52}$$

where, for every $n \in \mathbb{N}$, γ_n *is the largest* $\gamma \in \{\rho\sigma, \rho\sigma^2, \rho\sigma^3, \dots\}$ *satisfying*

$$\gamma^2 \left(\left\| \sum_{i=1}^p z_{n,i}^3 \nabla e_i(z_n^1) - \Phi^3_{\boldsymbol{z}_n,i}(\gamma) \nabla e(\Phi^1_{\boldsymbol{z}_n}(\gamma)) \right\|^2 + \left\| e(z_n^1) - e(\Phi^1_{\boldsymbol{z}_n}(\gamma)) \right\|^2 \right) \leq \theta^2 |||\boldsymbol{z}_n - \boldsymbol{\Phi}_{\boldsymbol{z}_n}(\gamma)|||^2. \tag{6.2.53}$$

Moreover, assume that at least one of the following additional statements hold:

(i) $\liminf_{n\to\infty} \gamma_n = \delta > 0$.

(ii) For every $i \in \{1, \dots, p\}$, ∇e_i *is bounded and uniformly continuous in every weakly compact subset of* $\overline{\operatorname{dom}\partial f}$.

Then, $(z_n^1)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 6.2.6.*

Proof. Let $\boldsymbol{\mathcal{H}} = \mathcal{H} \times \mathcal{G} \times \mathbb{R}^p$, define

$$\begin{cases} A \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon (x, u, v) \mapsto \partial f(x) \times \partial g^*(u) \times N_{[0,+\infty[^p}(v), \\ B_1 \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon (x, u, v) \mapsto (\nabla h(x), 0, 0), \\ B_2 \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon (x, u, v) \mapsto (M^* u, -Mx, 0), \end{cases} \tag{6.2.54}$$

and consider the operator B_3 defined in (6.2.44). Note that A is maximally monotone [2, Proposition 20.23 & Proposition 20.25], B_1 is $\beta-$cocoercive [2, Corollary 18.17], B_2 is $\|M\|$-Lipschitzian [10, Proposition 2.7(ii)] & [2, Fact 2.20], and the operator B_3 is maximally monotone by Proposition 6.2.7. Furthermore, note that $\operatorname{dom} A = \operatorname{dom}(\partial f) \times \operatorname{dom}(\partial g^*) \times [0, +\infty[^p$ and $\operatorname{dom} B_3 = (\cap_{i=1}^n \operatorname{int}\operatorname{dom} e_i) \times \mathcal{G} \times [0, +\infty[^p$. Hence, since $\operatorname{dom}(\partial f) \subset \cap_{i=1}^n \operatorname{int}\operatorname{dom} e_i$, we have $\operatorname{dom} A \cup X \subset \operatorname{dom} B_3$ and $0 \in \operatorname{int}(\operatorname{dom} A - \operatorname{dom} B_3)$ and therefore $A + B_3$ is maximally monotone [2, Corollary 25.5(ii)]. Therefore, the inclusion in (6.2.43) is a particular instance of Problem 6.2.1. Define, for every $n \in \mathbb{N}$, $\boldsymbol{x}_n = (x_n^1, x_n^2, x_n^3)$. Hence, (6.2.51), (6.2.44), and (6.2.54) yield

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \boldsymbol{x}_n = J_{\gamma_n A}(\boldsymbol{z}_n - \gamma_n(B_1 + B_2 + B_3)\boldsymbol{z}_n) \\ \boldsymbol{z}_{n+1} = P_X(\boldsymbol{x}_n + \gamma_n(B_2 + B_3)\boldsymbol{z}_n - \gamma_n(B_2 + B_3)\boldsymbol{x}_n), \end{array} \right. \tag{6.2.55}$$

where γ_n, by (6.2.53) and (6.2.54), satisfies

$$\gamma\|B_3\boldsymbol{z}_n - B_3 J_{\gamma A}(\boldsymbol{z}_n - \gamma(B_1+B_2+B_3)\boldsymbol{z}_n)\| \leq \theta\|\boldsymbol{z}_n - J_{\gamma A}(\boldsymbol{z}_n - \gamma(B_1+B_2+B_3)\boldsymbol{z}_n)\|. \quad (6.2.56)$$

Note that, if we assume (ii), by Proposition 6.2.7, B_3 is uniformly continuous in every weak compact subset of $\overline{\operatorname{dom}\partial f} \times \overline{\operatorname{dom}\partial g^*} \times [0,+\infty[^p = \overline{\operatorname{conv}}(\operatorname{dom} A) = \overline{\operatorname{conv}}(\operatorname{dom} A \cup X)$ [2, Corollary 21.14 & Exercise 3.2]. Altogether, assuming (i) or (ii), since $X \subset \operatorname{dom} A$, by Theorem 6.2.4, there exists $\boldsymbol{z} = (z,u,v) \in \mathcal{H} \times \mathcal{G} \times \mathbb{R}^p$ solution to (6.2.43) such that $\boldsymbol{z}_n \rightharpoonup \boldsymbol{z}$. Since $0 \in \operatorname{sri}(\operatorname{dom} g - M(\operatorname{dom} f))$, by (6.2.42), (6.2.43), and [2, Theorem 16.47] $\boldsymbol{z}$ is a saddle-point of $\mathcal{L}$ in (6.2.41). By [2, Corollary 19.30(v)], z is a solution to Problem 6.2.6 and the result follows. $\square$

6.2.5 Numerical Experiments

In this section we consider the following optimization problem

$$\min_{\substack{y^0 \leq x \leq y^1 \\ x_i(\ln(x_i/a_i)-1)-r_i \leq 0, \\ i\in\{1,\ldots,n\}}} \alpha\|Mx\|_1 + \frac{1}{2}\|Ax - z\|^2, \quad (6.2.57)$$

where $M \in \mathbb{R}^{r\times n}$, $A \in \mathbb{R}^{m\times n}$, $z \in \mathbb{R}^m$, $y^0 = (\eta_i^0)_{1\leq i\leq n} \in \mathbb{R}^n$, $y^1 = (\eta_i^1)_{1\leq i\leq N} \in \mathbb{R}^N$, $r_1,\ldots,r_n \in \,]-1,+\infty[$, and $a_1,\ldots,a_n \in \,]0,+\infty[$. Set

$$\begin{cases} C = \times_{i=1}^n [\eta_i^0, \eta_i^1], \\ f = \iota_C, \\ g = \alpha\|\cdot\|_1, \\ h = \|A\cdot - z\|^2/2, \\ e = (e_i(\cdot))_{i=1}^n, \end{cases} \quad (6.2.58)$$

where

$$(\forall i = 1,\ldots,n) \quad e_i\colon \mathbb{R}^n \to \,]-\infty,+\infty] : x \mapsto \begin{cases} x_i(\ln(x_i/a_i)-1) - r_i, & \text{if } x_i > 0; \\ -r_i, & \text{if } x_i = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then, we have $f \in \Gamma_0(\mathbb{R}^n)$, $g \in \Gamma_0(\mathbb{R}^r)$, and ∇h is $\|A\|^2$-Lipschitzian. Additionally, for every $i \in \{1,\ldots,n\}$, e_i is Gâteaux differentiable in $]0,+\infty[^n$,

$$(\nabla e_i(x))_k = \begin{cases} \ln x_k, & \text{if } k = i, \\ 0, & \text{otherwise,} \end{cases} \quad (6.2.59)$$

$\operatorname{dom} e_i$ is closed, $\cap_{i=1}^n \operatorname{dom} e_i = [0,+\infty[^n$, and $0 \in \operatorname{int}(\operatorname{dom}(\partial f) - \cap_{i=1}^n \operatorname{dom} e_i) = \times_{i=1}^N]-\infty, \eta_i^1[$. Hence, the optimization problem in (5.2.38) is a particular instance of Problem 6.2.6. In this setting, since $g^* = \iota_{[-\alpha,\alpha]^r}$ [2, Example 13.32(v) & Proposition 13.23(i)], we consider $X_1 \times X_2 \times X_3 = C \times [-\alpha,\alpha]^r \times [0,+\infty[^p$ in order to write the recurrence in (6.2.52) as Algorithm 6 below.

Algorithm 6

Fix $\boldsymbol{z}_0 = (z_0^1, z_0^2, z_0^3) \in \mathbb{R}^n \times \mathbb{R}^r \times \mathbb{R}^n$. Let $\sigma \in]0,1[$, let $\varepsilon = \|A\|^4 \frac{\sqrt{1+16\|M\|^2/\|A\|^4}-1}{8\|M\|^2}$, let $\theta = 2\varepsilon\|M\|(1-\sigma)/\|A\|^2$, and let $\epsilon > 0$.
while $r_n > \epsilon$ **do**
  $\gamma = 2\varepsilon\|M\|$
  $\mathtt{V} = 0$
  while $\mathtt{V} = 0$ **do**
    $\gamma \to \gamma \cdot \sigma$
    $\Phi^1(\gamma) = P_{[\eta_1,\eta_2]^n}\Big(z_n^1 - \gamma\big(A^*(Az_n^1 - z) + M^* z_n^2 + \sum_{i=1}^p z_{n,i}^3 \ln(z_{n,i}^1)\big)\Big)$
    $\Phi^2(\gamma) = \gamma(\mathrm{Id} - \operatorname{prox}_{\alpha\|\cdot\|_1/\gamma})(z_n^2/\gamma + Mz_n^1)$
    $\Phi^3(\gamma) = P_{[0,+\infty[^n}\big(z_n^3 + \gamma e(z_n^1)\big)$
    $\boldsymbol{\Phi}(\gamma) = (\Phi^1(\gamma), \Phi^2(\gamma), \Phi^3(\gamma))$
    if $\sum_{i=1}^p |z_{n,i}^3 \ln(z_{n,i}^1) - \Phi_i^3(\gamma)\ln(\Phi_i^1(\gamma))|^2 + \|e(z_n^1) - e(\Phi^1(\gamma))\|^2 \leq \frac{\theta^2}{\gamma^2} |||\boldsymbol{z}_n - \boldsymbol{\Phi}(\gamma)|||^2$ **then**
      $\mathtt{V} = 1$
    end if
  end while
  $\gamma_n = \gamma$
  $(x_n^1, x_n^2, x_n^3) = (\Phi^1(\gamma_n), \Phi^2(\gamma_n), \Phi^3(\gamma_n))$
  $z_{n+1}^1 = P_{[\eta_1,\eta_2]^N}\big(x_n^1 + \gamma_n(M^* z_n^2 + \sum_{i=1}^p z_{n,i}^3 \ln(z_{n,i}^1)) - \gamma_n(M^* x_n^2 + \sum_{i=1}^p x_{n,i}^3 \ln(x_{n,i}^1))\big)$
  $z_{n+1}^2 = P_{[-\alpha,\alpha]^r}(x_n^2 - \gamma_n M z_n^1 + \gamma_n M x_n^1)$
  $z_{n+1}^3 = P_{[0,+\infty[^p}(x_n^3 - \gamma_n e(z_n^1) + \gamma_n e(x_n^1))$
  $\boldsymbol{z}_{n+1} = (z_{n+1}^1, z_{n+1}^2, z_{n+1}^3)$
  $r_n = \mathcal{R}\big(\boldsymbol{z}_{n+1}, \boldsymbol{z}_n\big)$
  $n \to n+1$
end while
return $\boldsymbol{z}_{n+1}$

We compare Algorithm 6 with the algorithm propose in [12] called FBHF and with the MATLAB's `fmincon` (interior point).

To solve problem in (6.2.40) with FBHF algorithm, we consider $X = X_1 \times X_2 \times X_3$ and the followings operators (see (6.2.43) and [12, Theorem 2.3])

Table 6.1: Average time and average number of iterations of 20 random realizations of problem in (5.2.38) for Algorithm 6, FBHF, and `fmincon`.

	$\epsilon = 10^{-6}$	$N_2 = N_1/3$		$N_2 = N_1/2$		$N_2 = 2N_1/3$	
N_1	Algorithm	Av. Time (s)	Av. Iter	Av. Time (s)	Av. Iter	A. Time (s)	A. Iter
600	Alg. 6	6.82	7845	9.61	10384	15.79	16136
	FBHF	10.47	8280	14.22	10885	23.28	16772
	`fmincon`	52.52	238	66.25	276	69.78	251
900	Alg. 6	19.26	8185	28.89	11932	52.69	20653
	FBHF	31.28	8568	46.06	12375	84.17	21757
	`fmincon`	256.71	350	309.33	408	292.21	368
1200	Alg. 6	36.01	8809	62.82	14490	110.76	24778
	FBHF	59.41	9231	98.60	14783	174.08	25633
	`fmincon`	694.86	457	839.06	528	790.66	462

$$A = \begin{pmatrix} \partial f(\hat{x}) \\ \partial g^*(\hat{u}) \\ N_{[0,+\infty[^p}(\hat{v}) \end{pmatrix}, \quad B_1 = \begin{pmatrix} \nabla h(\hat{x}) \\ 0 \\ 0 \end{pmatrix}, \quad B_2 + B_3 = \begin{pmatrix} M^*\hat{u} + \sum_{i=1}^{p} \hat{v}_i \nabla e_i(\hat{x}) \\ -M\hat{x} \\ -e(\hat{x}) \end{pmatrix}. \tag{6.2.60}$$

In our numerical experiments, we generate 20 random realizations of A, M, z, and $r_1, \ldots, r_n$ for dimensions $n = m \in \{600, 900, 1200\}$ and $r \in \{n/3, n/2, 2n/3\}$. In each realization we define $a_i = 9$ for $i \in \{1, \ldots, n\}$, $\alpha = 0.05$, and $y^0 = \hat{y}_0$ and $y^1 = \hat{y}_1 + \text{rand}(n)$, where $\hat{y}_1$ and $\hat{y}_2$ satisfies $e(\hat{y}_0) = e(\hat{y}_1) = 0$. For Algorithm 6 we consider $\sigma = 0.99$. For FBHF we consider $\varepsilon = 0.8$, $\theta = \sqrt{1-\varepsilon}/2$, $\sigma = 0.99$, the maximally monotone operator A, the cocoercive operator B_1, and the continuous operator $B_1 + B_2$ on (6.2.60) (see [12, Theorem 2.3]).

In Table 6.1 we provide the average time and iterations to achieve a tolerance $\epsilon = 10^{-6}$ for the instances mentioned above. We can observe that, for each instance, Algorithm 6 is more efficient than the method FBHF and `fmincon`. Algorithm 6 and FBHF are similar in number of iterations, but each iteration of FBHF is more expensive in time than Algorithm 6. This is because FBHF needs, additionally, to activate the operators M^* and M in each line search. This difference is larger as the dimension of the problem increases. Although `fmincon` needs less iterations than Algorithm 6 and FBHF to reach the stop criterion, each iteration is very expensive in CPU time. Indeed, Algorithm 6 reaches the stop criterion in 20% of the time that `fmincon` takes.

Acknowledgments

The first author thanks the support of Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for centers of excellence and ANID under grant FONDECYT 1190871 from ANID-Chile. The second author thanks the support of ANID-Subdirección de Capital Humano/Doctorado Nacional/2018-21181024 and by the Dirección de Postgrado y Programas from UTFSM through Programa de Incentivos a la Iniciación Científica (PIIC).

Bibliography

[1] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2009, https://doi.org/10.1007/978-0-8176-4848-0.

[2] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5. With a foreword by Hédy Attouch.

[3] R. I. Boţ and E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algorithms, 71 (2016), pp. 519–540, https://doi.org/10.1007/s11075-015-0007-5.

[4] R. I. Boţ and E. R. Csetnek, *ADMM for monotone operators: convergence analysis and rates*, Adv. Comput. Math., 45 (2019), pp. 327–359, https://doi.org/10.1007/s10444-018-9619-3.

[5] R. I. Boţ, E. R. Csetnek, and A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM J. Optim., 23 (2013), pp. 2011–2036, https://doi.org/10.1137/12088255X.

[6] R. I. Boţ, E. R. Csetnek, and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Appl. Math. Comput., 256 (2015), pp. 472–487, https://doi.org/10.1016/j.amc.2015.01.017.

[7] R. I. Boţ and C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565, https://doi.org/10.1137/120901106.

[8] L. M. Briceño-Arias, *Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions*, Optimization, 64 (2015), pp. 1239–1261, https://doi.org/10.1080/02331934.2013.855210.

[9] L. M. Briceño-Arias, *Forward-partial inverse-forward splitting for solving monotone inclusions*, J. Optim. Theory Appl., 166 (2015), pp. 391–413, https://doi.org/10.1007/s10957-015-0703-2.

[10] L. M. Briceño-Arias and P. L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250, https://doi.org/10.1137/10081602X.

[11] L. M. Briceño-Arias and P. L. Combettes, *Monotone operator methods for Nash equilibria in non-potential games*, in Computational and analytical mathematics, vol. 50 of Springer Proc. Math. Stat., Springer, New York, 2013, pp. 143–159, https://doi.org/10.1007/978-1-4614-7621-4_9.

[12] L. M. Briceño-Arias and D. Davis, *Forward-backward-half forward algorithm for solving monotone inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871, https://doi.org/10.1137/17M1120099.

[13] M. N. Bùi and P. L. Combettes, *Multivariate monotone inclusions in saddle form*, Mathematics of Operations Research, to appear (2022).

[14] V. Cevher and B. Vũ, *A reflected forward-backward splitting method for monotone inclusions involving lipschitzian operators*, Set-Valued Var. Anal., (2020), https://doi.org/10.1007/s11228-020-00542-4.

[15] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently parallel algorithms in feasibility and optimization and their applications (Haifa, 2000), vol. 8 of Stud. Comput. Math., North-Holland, Amsterdam, 2001, pp. 115–152, https://doi.org/10.1016/S1570-579X(01)80010-0.

[16] P. L. Combettes and J. Eckstein, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions*, Math. Program., 168 (2018), pp. 645–672, https://doi.org/10.1007/s10107-016-1044-0.

[17] P. L. Combettes and J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued Var. Anal., 20 (2012), pp. 307–330, https://doi.org/10.1007/s11228-011-0191-y.

[18] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318, https://doi.org/10.1080/02331934.2012.733883.

[19] E. Csetnek, Y. Malitsky, and M. Tam, *Shadow Douglas-Rachford splitting for monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 665–678.

[20] D. Davis and W. Yin, *A three-operator splitting scheme and its optimization applications*, Set-Valued Var. Anal., 25 (2017), pp. 829–858, https://doi.org/10.1007/s11228-017-0421-z.

[21] Y. Dong, *Weak convergence of an extended splitting method for monotone inclusions*, J. Global Optim., 79 (2021), pp. 257–277, https://doi.org/10.1007/s10898-020-00940-w.

[22] D. Dũng and B. C. Vũ, *A splitting algorithm for system of composite monotone inclusions*, Vietnam J. Math., 43 (2015), pp. 323–341, https://doi.org/10.1007/s10013-015-0121-7.

[23] J. Eckstein, *A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 155–182, https://doi.org/10.1007/s10957-017-1074-7.

[24] D. Gabay, *Chapter IX applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, New York, 1983, pp. 299 – 331, https://doi.org/10.1016/S0168-2024(08)70034-1.

[25] R. Glowinski and A. Marroco, *Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite, d'une classe de problemes de dirichlet non lineares*, Revue Francaise d'Automatique, Informatique et Recherche Operationelle, 9 (1975), pp. 41–76, https://doi.org/10.1051/M2AN/197509R200411.

[26] A. A. Goldstein, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society, 70 (1964), pp. 709 – 710, https://doi.org/bams/1183526263, https://doi.org/.

[27] P. R. Johnstone and J. Eckstein, *Projective splitting with forward steps only requires continuity*, Optim. Lett., 14 (2020), pp. 229–247, https://doi.org/10.1007/s11590-019-01509-7.

[28] P. R. Johnstone and J. Eckstein, *Single-forward-step projective splitting: exploiting cocoercivity*, Comput. Optim. Appl., 78 (2021), pp. 125–166, https://doi.org/10.1007/s10589-020-00238-3.

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

[30] Y. Malitsky and M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim., 30 (2020), pp. 1451–1472, https://doi.org/10.1137/18M1207260.

[31] H. Raguet, J. Fadili, and G. Peyré, *A generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226, https://doi.org/10.1137/120872802.

[32] J. Rieger and M. K. Tam, *Backward-forward-reflected-backward splitting for three operator monotone inclusions*, Appl. Math. Comput., 381 (2020), pp. 125248, 10, https://doi.org/10.1016/j.amc.2020.125248.

[33] R. T. Rockafellar, *Monotone operators associated with saddle-functions and minimax problems*, in Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968), Amer. Math. Soc., Providence, R.I., 1970, pp. 241–250.

[34] E. K. Ryu and B. C. Vũ, *Finding the forward-Douglas-Rachford-forward method*, J. Optim. Theory Appl., 184 (2020), pp. 858–876, https://doi.org/10.1007/s10957-019-01601-z.

[35] S. Salzo, *The variable metric forward-backward splitting algorithm under mild differentiability assumptions*, SIAM J. Optim., 27 (2017), pp. 2153–2181, https://doi.org/10.1137/16M1073741.

[36] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1997, https://doi.org/10.1090/surv/049.

[37] J. E. Spingarn, *Partial inverse of a monotone operator*, Appl. Math. Optim., 10 (1983), pp. 247–265, https://doi.org/10.1007/BF01448388.

[38] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446, https://doi.org/10.1137/S0363012998338806.

[39] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

Chapter 7

Perspectives

To conclude this thesis we present some future perspectives of work.

- A direct extension of this work is to provide an algorithm which solves Problem 1.1.9 by splitting all the operators. In Chapter 5 we provide a method for solving Problem 1.1.9, when $B_3 = 0$, by using partial inverse techniques and FBHF (Algorithm 1.1.13). In Chapter 6 we present the Forward-Backward Half Forward with line search method (FBHFLS) to solve numerically Problem 1.1.9 when $V = \mathcal{H}$, generalizing FBHF. Hence, by a similar procedure, using partial inverse techniques and FBHFLS algorithm proposed in Chapter 6, it is foreseeable that we can solve Problem 1.1.9.

- We begin this thesis introducing Problem 1.1.1 and we split it in two cases solving both separately. Then, Problem 1.1.1 is still a open question. By solving Problem 1.1.9 and proceeding similarly to Section 5.2.4, we expect to solve Problem 1.1.1.

- In Chapter 3 we derive the convergence of Krasnosel'skiĭ-Mann (KM) iterations defined in the range of monotone self-adjoint linear operators. To obtain the convergence of the relaxed primal-dual algorithm with critical preconditioners we prove that this method defines KM iterations in the range of a particular linear operator. An open question is whether it is possible to apply and extend this result to other methods, with critical step-sizes/preconditioners or to develop new algorithms.

- Several articles provide convergence rates for DRS (Algorithm 1.1.4) (see for instance [2, 7]). In Chapter 2 we generalize DRS by providing the SDR algorithm, additionally we show its relation with the PDS including critical step-sizes. In Chapter 2.2.5 we present numerical simulations and we show instances when an adequate choice of the step-sizes improves, regarding number of iterations and time, the convergence of SDR. Hence, the computation of a convergence rate of SDR will allows us to deduce a relation between the speed of convergence and step-sizes. This is also a perspective for all the methods proposed in this thesis.

- Inertial methods for monotone inclusions update each iteration by using the last two iterates [1]. The inclusion of inertial steps to some methods can accelerate its convergence. For instance, variations including inertial steps of methods as DRS, Tseng's type, ADMM, forward-backward, and forward-backward-forward are available in [1, 4, 5, 6, 3]. In view of that, an interesting future work is to include inertial steps to the methods presented in this thesis.

Bibliography

[1] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Anal., 9 (2001), pp. 3–11.

[2] H. H. Bauschke, J. Y. Bello Cruz, T. T. A. Nghia, H. M. Phan, and X. Wang, *The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle*, J. Approx. Theory, 185 (2014), pp. 63–79, https://doi.org/10.1016/j.jat.2014.06.002.

[3] R. Boţ and E. Csetnek, *An inertial alternating direction method of multipliers*, Minimax Theory Appl., 1 (2016), pp. 29–49.

[4] R. I. Boţ and E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algorithms, 71 (2016), pp. 519–540, https://doi.org/10.1007/s11075-015-0007-5.

[5] R. I. Boţ and E. R. Csetnek, *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems*, J. Optim. Theory Appl., 171 (2016), pp. 600–616, https://doi.org/10.1007/s10957-015-0730-z.

[6] R. I. Boţ, E. R. Csetnek, and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Appl. Math. Comput., 256 (2015), pp. 472–487, https://doi.org/10.1016/j.amc.2015.01.017.

[7] B. He and X. Yuan, *On the convergence rate of Douglas-Rachford operator splitting method*, Math. Program., 153 (2015), pp. 715–722, https://doi.org/10.1007/s10107-014-0805-x.