Universidad Técnica Federico Santa María

Departamento de Matemática
Santiago - Chile

Primal-dual splitting algorithms for constrained monotone inclusions

Tesis presentada por:
Sergio Eduardo López Rivera

Como requisito parcial para optar al grado académico de Magíster en Ciencias Mención Matemática y al título profesional de Ingeniero Civil Matemático

Director de tesis:
Luis Briceño Arias

Co-Director de tesis:
Julio Deride Silva

Julio, 2020

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE MATEMÁTICA
SANTIAGO - CHILE

Primal-dual splitting algorithms for constrained monotone inclusions

Tesis presentada por:
Sergio Eduardo López Rivera

Como requisito parcial para optar al grado académico de Magíster en Ciencias Mención Matemática y al título profesional de Ingeniero Civil Matemático

Director de tesis:
Luis Briceño Arias

Co-Director de tesis:
Julio Deride Silva

Examinadores:
Cristopher Hermosilla

Aris Daniilidis

Juan Peypouquet

Julio, 2020

TÍTULO DE LA TESIS: Primal-dual splitting algorithms for constrained monotone inclusions.

AUTOR: Sergio Eduardo López Rivera.

TESIS, presentada como requisito parcial para optar al grado académico de Magíster en Ciencias Mención Matemática y al título profesional Ingeniero Civil Matemático de la Universidad Técnica Federico Santa María.

COMISIÓN EVALUADORA:

Integrantes	Firma
Luis Briceño Universidad Técnica Federico Santa María	______________
Julio Deride Universidad Técnica Federico Santa María	______________
Cristopher Hermosilla Universidad Técnica Federico Santa María	______________
Juan Peypouquet Universidad de Chile	______________
Aris Daniilidis Universidad de Chile	______________

Santiago, Julio 2020.

Resumen

En esta tesis proponemos un algoritmo eficiente de separación primal-dual para resolver inclusiones monótonas restrictas que incluyen un cono normal a un subespacio vectorial cerrado. Nuestro algoritmo incorpora una proyección adicional sobre un conjunto de restricciones, el cual representa información a priori en la solución. Este trabajo se divide en dos partes. En la primera parte, estudiamos el caso en que el subespacio vectorial es todo el espacio, demostrando convergencia débil de nuestro método, como también convergencia acelerada y convergencia lineal bajo correspondientes hipótesis adicionales sobre los operadores y parámetros del algoritmo. En la segunda parte, consideramos el caso general y demostramos convergencia débil de nuestro método mediante la caracterización de soluciones de la inclusión, usando la técnica de la inversa parcial de un operador. La eficiencia de nuestro método se ve reflejada en el contexto de optimización convexa con restricciones lineales afines y de subespacio vectorial. El uso de la información a priori permite la factibilidad de las iteraciones primales en un subconjunto de restricciones y el uso de la inversa parcial de un operador permite explotar la estructura del subespacio vectorial. Estas dos características de nuestro método mejoran la eficiencia con respecto a varios métodos clásicos en la literatura. Finalmente, también aplicamos nuestro método al problema de asignación de tráfico en redes de transporte con expansión de capacidad de arco bajo incertidumbre, en el cual mostramos la ventaja de usar la estructura del subespacio vectorial del problema.

Abstract

In this thesis we propose an efficient splitting algorithm for solving constrained primal-dual monotone inclusions with a normal cone to a closed vector subspace. Our algorithm incorporates an additional projection onto a set of constraints, which represents a priori information on the solutions. This work is divided in two parts. In the first part, we study the case when the vector subspace is the whole space, in which we provide weak convergence of our method, as well as accelerated convergence and linear convergence under corresponding additional hypotheses on the operators and step sizes of the algorithm. In the second part, we consider the general case and we demonstrate weak convergence of our method by characterizing the solutions to the inclusion using the partial inverse operator. The efficiency of our method is illustrated in the context of convex optimization with affine linear constraints and vector subspace constraints. The use of the a priori information allows the feasibility of primal iterations in a subset of constraints and the use of the partial inverse operator allows to exploit the vector subspace structure. These two features of our method improve the efficiency with respect to several methods in the literature. Finally, we also apply our method to solve the traffic assignment problem with arc-capacity expansion on a network with minimal cost under uncertainty, in which we show the advantage of using the vector subspace structure of the problem.

Contents

Chapter 1

Introduction

This thesis is devoted to the resolution of constrained composite primal-dual monotone inclusions involving a normal cone to a closed vector subspace. The problems of monotone inclusion and convex optimization have applications in many fields of engineering and applied mathematics, such as image processing, inclusions and evolution problems, variational inequalities, learning, partial differential equations (PDE's), game theory, among others (see [2, 3, 32, 33, 40]). In the formulation, we assume that solutions belong to an a priori information set which is modeled by the set of fixed points of an averaged nonexpansive operator T. This feature explains the terminology of "constrained" inclusion. For instance, if $T = P_C$, where C is a closed convex set, the a priori information is given by C. On the other hand, if $T = (\operatorname{Id} + A)^{-1}$, where A is a maximally monotone operator, the a priori information set is $A^{-1}(\{0\})$ and we aim at solving a common solution of two monotone inclusions. In addition, in the context of convex optimization, the vector subspace structure included in our formulation model intrinsic properties of the solution, as regularity in PDE's or signal processing, and nonanticipativity in stochastic problems.

In this thesis we propose generalizations to classical primal-dual splitting algorithms for monotone inclusions [12, 14, 46, 48], by introducing additional projections onto the a priori information set and the vector subspace. We put special attention to convex optimization problems with vector subspace constraints and a priori information on the solution. In this context, we aim at developing efficient methods which take into advantage the a priori information on the solutions and the vector subspace structure of the problem and generalize several algorithms in the literature [12, 14, 20, 29, 48].

In the particular case of convex optimization problems with affine linear constraints, the a priori information may model a selection of the constraints, which leads to a redundancy in the formulation with numerical advantages. Indeed, the proposed method incorporates a projection onto the selection of the constraints, which imposes partial feasibility of the primal iterates and improves the efficiency of the method. In addition, the proposed method exploits the vector subspace structure via the partial inverse operator, first introduced in [46]. The application of this technique also shows numerical improvements with respect to available methods in the literature.

In the case when the vector subspace is the whole space and under strong monotonicity assumptions, we prove acceleration of the proposed method by choosing appropriate stepsizes. We generalize the result obtained in [20, Section 5.1 & Section 5.2] in the optimization context.

The proposed algorithm is applied to several problems as constrained LASSO, constrained ℓ^1 minimization, and capacity expansion problem in traffic networks. We provide numerical comparisons with available methods in the literature in each context [12, 20, 29, 48] and we show the advantage of the additional projection onto the priori information set and the partial inverse technique.

In the engineering application devoted to capacity expansion problem in stochastic traffic networks, the vector subspace models nonanticipativity in the expansion decision. We consider different formulations of vector subspaces and we compare our method in each of these formulations for several random generated instances.

In this chapter we define the main inclusion problem, we briefly review the state of the art of primal-dual methods for solving monotone inclusions problems and non-differentiable convex optimization problems, and we discuss the specific objectives of this thesis.

1.1. Notation and Preliminaries

In this section, we define the concepts that were used throughout this thesis, including some basic facts.

Let $\mathcal{H}$ be a real Hilbert space. We denote the scalar product of $\mathcal{H}$ by $\langle \cdot \,|\, \cdot \rangle$ and its norm associated by $\|\cdot\|$. The class of bounded linear operators from $\mathcal{H}$ to a real Hilbert space $\mathcal{G}$ is denoted by $\mathcal{L}(\mathcal{H},\mathcal{G})$ and if $\mathcal{H}=\mathcal{G}$ this class is denoted by $\mathcal{L}(\mathcal{H})$. Given $L \in \mathcal{L}(\mathcal{H},\mathcal{G})$, its adjoint operator is denoted by $L^* \in \mathcal{L}(\mathcal{G},\mathcal{H})$. The projection on a nonempty closed convex subset $C \subset \mathcal{H}$ is defined by $P_C \colon x \in \mathcal{H} \to \operatorname{argmin}_{y\in C} \|y-x\|$ and its normal cone is the operator defined by

$$N_C \colon \mathcal{H} \rightrightarrows \mathcal{H} \colon x \mapsto N_C x = \begin{cases} \{u \in \mathcal{H} \mid (\forall y \in C)\ \langle u \mid y - x\rangle \leq 0\} & \text{if } x \in C \\ \varnothing & \text{if } x \notin C. \end{cases}$$

Let $M \colon \mathcal{H} \rightrightarrows \mathcal{H}$ a set-valued operator. We denote by $\operatorname{dom} M = \{x \in \mathcal{H} \,|\, Mx \neq \varnothing\}$ to the domain of M and by $\operatorname{zer} M = \{x \in \mathcal{H} \,|\, 0 \in Mx\}$ to the set of the zeros of M. On the other hand, we denote by $\operatorname{ran} M = \{u \in \mathcal{H} \,|\, (\exists x \in \mathcal{H})\, u \in Mx\}$ to the range of M and by $\operatorname{gra} M = \{(x,u) \in \mathcal{H}\times\mathcal{H} \,|\, u \in Mx\}$ to its graph.

The inverse operator of M is defined by $M^{-1} \colon \mathcal{H} \rightrightarrows \mathcal{H} \colon u \mapsto \{x \in \mathcal{H} \,|\, u \in Mx\}$. The resolvent of M is the operator defined by $J_M = (\mathrm{Id} + M)^{-1}$ where Id it is the identity operator in $\mathcal{H}$. We have the following identity

$$(\forall \lambda > 0) \quad J_{\lambda M^{-1}} = \mathrm{Id} - \lambda J_{\lambda^{-1} M} \circ \left(\lambda^{-1}\,\mathrm{Id}\right), \tag{1.1.1}$$

which is deduced from [6, Proposition 23.7(ii)]. Also, M is monotone if

$$(\forall (x,u) \in \operatorname{gra} M)(\forall (y,v) \in \operatorname{gra} M)\ \langle x - y \mid u - v\rangle \geq 0 \tag{1.1.2}$$

and maximally monotone if there exists no monotone operator $N \colon \mathcal{H} \rightrightarrows \mathcal{H}$ such that $\operatorname{gra} M \subsetneq \operatorname{gra} N$. Or equivalently, if

$$(\forall (x,u) \in \mathcal{H}^2)\, \{(x,u) \in \operatorname{gra} M \Leftrightarrow (\forall (y,v) \in \operatorname{gra} M) \quad \langle x-y \mid u-v\rangle \geq 0\}.$$

Moreover, M is ρ-strongly monotone (with $\rho \geq 0$) if

$$(\forall (x,u) \in \operatorname{gra} M)(\forall (y,v) \in \operatorname{gra} M)\ \langle x-y \mid u-v\rangle \geq \rho\,\|x-y\|^2. \tag{1.1.3}$$

An operator $R\colon \mathcal{H} \to \mathcal{H}$ is nonexpansive if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Rx - Ry\| \leq \|x - y\|. \tag{1.1.4}$$

and quasinonexpansive if (1.1.4) is satisfied for every $x \in \mathcal{H}$ and every $y \in \operatorname{Fix} R := \{x \in \mathcal{H} \mid Rx = x\}$. Moreover $R\colon \mathcal{H} \to \mathcal{H}$ is strictly quasinonexpansive if

$$(\forall x \in \mathcal{H} \setminus \operatorname{Fix} R)(\forall y \in \operatorname{Fix} R) \quad \|Rx - y\| < \|x - y\|. \tag{1.1.5}$$

An operator $T\colon \mathcal{H} \to \mathcal{H}$ is α-averaged for some $\alpha \in \,]0,1[$ if there exists a nonexpansive operator $R\colon \mathcal{H} \to \mathcal{H}$ such that $T = (1-\alpha)\operatorname{Id} + \alpha R$, or, equivalently, if [6, Proposition 4.35]

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2 \tag{1.1.6}$$

On the other hand, $T\colon \mathcal{H} \to \mathcal{H}$ is $\beta-$cocoercive for some $\beta > 0$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle \geq \beta\|Tx - Ty\|^2 \tag{1.1.7}$$

and is firmly nonexpansive if it is $\frac{1}{2}$-averaged, or equivalently [6, Remark 4.34(iv)], if it is 1-cocoercive. We have that for every nonempty closed convex subset $C \subset \mathcal{H}$, P_C is firmly nonexpansive [6, Proposition 4.16]. Note that a firmly nonexpansive operator is strictly quasinonexpansive.

Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be a monotone operator, then for every $x \in \mathcal{H}$, $J_A x$ is a singleton. Indeed, let $(y_1, y_2) \in \mathcal{H}^2$ such that $y_1 \in J_A x$ and $y_2 \in J_A x$. Then $x - y_1 \in Ay_1$ and $x - y_2 \in Ay_2$. By monotonicity of A, $0 \leq \langle -y_2 + y_2 \mid y_1 - y_2\rangle = -\|y_1 - y_2\|^2$. Hence $y_1 = y_2$. In addition, if A is monotone, then J_A is firmly nonexpansive [6, Proposition 23.8(i)]. Furthermore, if A is maximally monotone, then $\operatorname{dom} J_A = \mathcal{H}$ [6, Proposition 23.8(ii)].

Let V be a closed vector subspace of $\mathcal{H}$. The partial inverse of $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ with respect to V, denoted by A_V, is defined by [46]:

$$(\forall (x,y) \in \mathcal{H}^2) \quad y \in A_V x \Leftrightarrow P_V y + P_{V^\perp} x \in A(P_V x + P_{V^\perp} y). \tag{1.1.8}$$

The parallel sum of two operators $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ and $B\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is the operator defined by $A \Box B := (A^{-1} + B^{-1})^{-1}\colon \mathcal{H} \rightrightarrows \mathcal{H}$.

Given $f\colon \mathcal{H} \to [-\infty, +\infty]$, the Fenchel conjugate of f is the function defined by $f^*\colon u \in \mathcal{H} \mapsto \sup\{\langle u \mid x\rangle - f(x) \mid x \in \mathcal{H}\} \in [-\infty, +\infty]$, and if $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ is proper (i.e. $\operatorname{dom} f := \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \varnothing$), we define the subdifferential of f as the set-valued operator defined by

$$\partial f\colon \mathcal{H} \rightrightarrows \mathcal{H}\colon x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\, \langle u \mid y - x\rangle \leq f(y) - f(x)\}. \tag{1.1.9}$$

The set of lower semicontinuous, convex and proper functions $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ is denoted by $\Gamma_0(\mathcal{H})$. On the other hand, a proper function $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ is $\rho-$ strongly convex (with $\rho \geq 0$) if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \lambda \in [0,1]) \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\lambda(1-\lambda)\rho}{2}\|x - y\|^2.$$

For every $\lambda > 0$ and $f \in \Gamma_0(\mathcal{H})$, the proximity operator is defined by

$$\operatorname{prox}_{\lambda f}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \left\{ \frac{\|y - x\|^2}{2\lambda} + f(y) \right\}. \tag{1.1.10}$$

The fact of that the proximity operator is well-defined it follows from [6, Corollary 11.16(i)].

Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone [6, Theorem 20.25], $f^* \in \Gamma_0(\mathcal{H})$ [6, Corollary 13.38], $(\partial f)^{-1} = \partial f^*$ [6, Corollary 16.30], and $(\forall \lambda > 0)$ $J_{\lambda \partial f} = \operatorname{prox}_{\lambda f}$ [6, Example 23.3]. Note that in this case the identity (1.1.1) is read as

$$(\forall \lambda > 0) \quad \operatorname{prox}_{\lambda f^*} = \operatorname{Id} - \lambda \operatorname{prox}_{\lambda^{-1} f} \circ \left(\lambda^{-1} \operatorname{Id}\right). \tag{1.1.11}$$

When $C \subset \mathcal{H}$ it is a nonempty closed convex of $\mathcal{H}$, we have that $N_C = \partial \iota_C$ and $J_{N_C} = P_C$, where

$$\iota_C \colon x \mapsto \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C. \end{cases}$$

Moreover, for f and g in $\Gamma_0(\mathcal{H})$, we denote by $f \,\square\, g \colon x \mapsto \inf\{f(y) + g(x-y) \mid y \in \mathcal{H}\}$ the infimal convolution of f and g.

Let see the notion of convex cone. A subset $C \subset \mathcal{H}$ it is a cone if $C = \mathbb{R}_{++} C := \{\lambda x \mid x \in C \wedge \lambda > 0\}$. Then, $\mathcal{H}$ is a cone and is easy to see that the intersection of an arbitrary family of cones is a cone. Therefore, the following notion is well defined. The conical hull of C is intersection of all cones in $\mathcal{H}$ containing C, i.e., the smallest cone in $\mathcal{H}$ that contains C. It is denoted by $\operatorname{cone} C$.

Let C be a nonempty convex subset of $\mathcal{H}$. The strong relative interior of C is

$$\operatorname{sri} C = \{x \in C \mid \operatorname{cone}(C - x) = \overline{\operatorname{span}}\,(C - x)\}, \tag{1.1.12}$$

where, for every $C \subset \mathcal{H}$, $\overline{\operatorname{span}}\, C$ denote the closure of the vector subspace of $\mathcal{H}$ smallest containing C, that is, the closed vector subspace of $\mathcal{H}$ smallest containing C.

Let f and g in $\Gamma_0(\mathcal{H})$. We have that if $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$, then [6, Proposition 15.7(i)& Proposition 25.32]

$$\partial(f \,\square\, g) = (\partial f) \,\square\, (\partial g) \tag{1.1.13}$$

and if $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$, then [6, Proposition 15.7(i)& Theorem 15.3]

$$(f + g)^* = f^* \,\square\, g^*. \tag{1.1.14}$$

1.2. Main Problem

The main problem in this thesis is the following.

Problem 1.2.1 Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $V \subset \mathcal{H}$ and $W \subset \mathcal{G}$ be closed vector subspaces, let $T \colon \mathcal{H} \to \mathcal{H}$ be an $\alpha-$averaged operator with $\alpha \in \left]0,1\right[$, let $L \in \mathcal{L}(\mathcal{H}, \mathcal{G})$ be such that $\operatorname{ran} L \subset W$, let $A \colon \mathcal{H} \rightrightarrows \mathcal{H}$, $B \colon \mathcal{G} \rightrightarrows \mathcal{G}$, and $D \colon \mathcal{G} \rightrightarrows \mathcal{G}$ be maximally monotone operators, and let $C \colon \mathcal{H} \to \mathcal{H}$ be a $\beta-$cocoercive operator. Suppose that A, B^{-1}, and D are ρ, χ, and $\delta-$strongly monotone operators respectively, where $(\rho, \chi) \in [0, +\infty]^2$ and $(\delta, \beta) \in \left]0, +\infty\right]^2$. The problem is to solve the primal and dual inclusions

$$\text{find} \quad x \in \operatorname{Fix} T \quad \text{such that} \quad 0 \in Ax + L^*((B \square D)(Lx)) + Cx + N_V x \tag{$\mathcal{P}$}$$

$$\text{find} \quad u \in W \quad \text{such that} \quad (\exists x \in \operatorname{Fix} T) \begin{cases} -L^* u \in Ax + Cx + N_V x \\ Lx \in B^{-1}u + D^{-1}u, \end{cases} \tag{$\mathcal{D}$}$$

under the assumption that ($\mathcal{P}$)-($\mathcal{D}$) admit solutions.

In the case when $A = \partial f$, $B = \partial g$, $C = \nabla h$, and $D = \partial \ell$, where $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $h\colon \mathcal{H} \to \mathbb{R}$ is a differentiable convex function with $\beta^{-1}-$Lipschitz gradient, and $\ell \in \Gamma_0(\mathcal{G})$ is $\delta-$strongly convex for some $(\delta, \beta) \in \left]0, +\infty\right]^2$, the Problem 1.2.1, under the corresponding qualification conditions

$$\begin{aligned} &0 \in \operatorname{sri}(V - \operatorname{dom}(f + (g \,\Box\, \ell) \circ L)) \\ &0 \in \operatorname{sri}(\operatorname{dom} g + \operatorname{dom} \ell - L(\operatorname{dom} f)) \cap \operatorname{sri}(\operatorname{dom} g^* - \operatorname{dom} \ell^*), \end{aligned} \tag{1.2.1}$$

reduces to the composite constrained convex optimization primal problem

$$\text{find } \overline{x} \in \operatorname{Fix} T \cap \underset{x \in V}{\operatorname{argmin}}\, f(x) + (g \,\Box\, \ell)(Lx) + h(x), \tag{1.2.2}$$

together with the dual problem (by [6, Proposition 13.24(i)&] and (1.1.14))

$$\text{find } \overline{u} \in W \cap \underset{u \in \mathcal{G}}{\operatorname{argmin}}\, g^*(u) + (f^* \,\Box\, h^* \,\Box\, \iota_{V^\perp})(-L^* u) + \ell^*(u). \tag{1.2.3}$$

Indeed, from [6, Proposition 27.1], [6, Theorem 16.47], and (1.1.13) we have that

$$\operatorname{argmin}(f + (g \,\Box\, \ell) \circ L + h + \iota_V) = \operatorname{zer}(\partial f + L^* \circ (\partial g \Box \partial \ell) \circ L + \nabla h + N_V). \tag{1.2.4}$$

In addition, by [6, Corollary 18.17], ∇h is $\beta-$cocoercive.

1.3. State of the art

The present section is devoted to review the existing methods in the literature for solving Problem 1.2.1 in special instances. In addition, we briefly describe the specific objectives of this thesis.

1.3.1. State of the art for the case $V = \mathcal{H}$

In this section we consider the Problem 1.2.1 when $T = \operatorname{Id}$, $W = \mathcal{G}$, and $V = \mathcal{H}$. In this case the problem reduces to the primal-dual inclusion

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + L^*((B \Box D)(Lx)) + Cx, \tag{1.3.1}$$

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad (\exists x \in \mathcal{H}) \begin{cases} -L^* u \in Ax + Cx \\ u \in (B \Box D)(Lx) \end{cases} \tag{1.3.2}$$

The previous primal-dual inclusions are studied in [48], where the following algorithm to solve (1.3.1)-(1.3.2) is provided.

Theorem 1.3.1 *[48, Theorem 3.1(i)] Let $x^0 \in \mathcal{H}$, let $u^0 \in \mathcal{G}$, and let $(\tau, \gamma) \in]0, +\infty[^2$ such that $2\lambda \min\{\beta, \delta\} > 1$, where $\lambda = \min\{\tau^{-1}, \gamma^{-1}\}(1 - \sqrt{\tau\gamma\|L\|^2})$. Consider the routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x^{k+1} = J_{\tau A}(x^k - \tau(L^* u^k + Cx^k)) \\ \overline{x}^k = 2x^{k+1} - x^k \\ u^{k+1} = J_{\gamma B^{-1}}(u^k + \gamma(L\overline{x}^k - D^{-1}u^k)) \end{array} \right. \tag{1.3.3}$$

Then, there exists $(\overline{x}, \overline{u})$ solution to (1.3.1)-(1.3.2) such that $x^k \rightharpoonup \overline{x}$ and $u^k \rightharpoonup \overline{u}$.

When $B\colon u \mapsto \{0\}$ and

$$D\colon u \mapsto \begin{cases} \mathcal{G} & \text{if } u = 0 \\ \varnothing & \text{if } u \neq 0, \end{cases} \tag{1.3.4}$$

it follows that $B \Box D = J_{\gamma B^{-1}} = D^{-1} = 0$. Therefore, the algorithm (1.3.3) reduces to the classical Forward-Backward splitting algorithm [3], which converges to a solution of the following inclusion problem

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Cx. \tag{1.3.5}$$

The problem in (1.3.5) has been studied in [3, 25, 36].

Note that C and D^{-1} are β^{-1} and $\delta^{-1}-$Lipschitz, respectively. Then, we can use the following method proposed in [26], which solves (1.3.1)-(1.3.2) when C and D^{-1} are in general monotone Lipschitzian operators.

Theorem 1.3.2 *[26, Theorem 3.1(ii)] Let $\gamma \in \left]0, (1+\lambda)^{-1}\right[$, where $\lambda = \max\{1/\beta, 1/\delta\} + \|L\|$, let $x^0 \in \mathcal{H}$, and let $u^0 \in \mathcal{G}$. Consider the following routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{aligned} p_1^k &= J_{\gamma A}(x^k - \gamma(Cx^k + L^* u^k)) \\ p_2^k &= J_{\gamma B^{-1}}(u^k + \gamma(Lx^k - D^{-1}u^k)) \\ x^{k+1} &= p_1^k - \gamma(L^* p_2^k + Cp_1^k - L^* u^k - Cx^k) \\ u^{k+1} &= p_2^k + \gamma(Lp_1^k - D^{-1}p_2^k - Lx^k + D^{-1}u^k). \end{aligned} \right. \tag{1.3.6}$$

Then, there exists $(\overline{x}, \overline{u})$ solution to (1.3.1)-(1.3.2) such that $x^k \rightharpoonup \overline{x}$ and $u^k \rightharpoonup \overline{u}$.

When $C = 0$ and D is defined by (1.3.4), the algorithm (1.3.6) reduces to the method proposed first in [14].

Note that although the method (1.3.6) solves a more general case (when C is Lipschitz), in the context of our Problem 1.2.1 it does not use the cocoercivity of C as in the algorithm (1.3.3).

In the context of convex optimization, in this case the problem (1.2.2)-(1.2.3) reduces to the primal-dual problem

$$\underset{x \in \mathcal{H}}{\text{minimize}}\; f(x) + (g \,\Box\, \ell)(Lx) + h(x), \tag{$\mathcal{P}_0$}$$

$$\underset{u \in \mathcal{G}}{\text{minimize}}\; g^*(u) + (f^* \,\Box\, h^*)(-L^* u) + \ell^*(u) \tag{$\mathcal{D}_0$}$$

and ($\mathcal{P}_0$)-($\mathcal{D}_0$) can be solved by (1.3.3) and (1.3.6) using $J_{\partial f} = \operatorname{prox}_f$. A classical method for solving ($\mathcal{P}_0$)-($\mathcal{D}_0$) is the Alternating Direction method of Multipliers (ADMM) [33]. The ADMM consider the following equivalent problem

$$\underset{\substack{(x,y) \in \mathcal{H} \times \mathcal{G} \\ Lx = y}}{\text{minimize}}\; f(x) + (g \,\Box\, \ell)(y) + h(x),$$

from which define the augmented Lagrangian:

$$(\forall \gamma > 0) \quad \mathcal{L}_\gamma(x, y, u) = f(x) + (g \,\Box\, \ell)(y) + h(x) + \langle u \mid Lx - y \rangle + \frac{\gamma}{2}\|Lx - y\|^2. \tag{1.3.7}$$

Given $(x^0, y^0, u^0) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$, ADMM iterates

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{aligned} x^{k+1} &= \operatorname{argmin}_{x \in \mathcal{H}} \mathcal{L}_\gamma(x, y^k, u^k) \\ &= \operatorname{argmin}_{x \in \mathcal{H}} \left\{ f(x) + h(x) + \left\langle u^k \mid Lx \right\rangle + \tfrac{\gamma}{2}\|Lx - y^k\|^2 \right\} \\ y^{k+1} &= \operatorname{argmin}_{y \in \mathcal{G}} \mathcal{L}_\gamma(x^{k+1}, y, u^k) \\ &= \operatorname{argmin}_{y \in \mathcal{G}} \left\{ (g \,\Box\, \ell)(y) - \left\langle u^k \mid y \right\rangle + \tfrac{\gamma}{2}\|Lx^{k+1} - y\|^2 \right\} \\ u^{k+1} &= u^k + \gamma(Lx^{k+1} - y^{k+1}). \end{aligned} \right. \tag{1.3.8}$$

Note that the first step in the algorithm (1.3.8) is not easy to calculate in general since it depends on L and $f+h$. In addition, the algorithm is efficient only in particular cases: when $f+h$ is a quadratic function, $L^*L = \alpha\,\mathrm{Id}$ for some $\alpha \neq 0$ or L^*L is invertible, etc. This disadvantage of ADMM, in the context of convex optimization, is solved by the algorithms (1.3.3) and (1.3.6) since they split the influences of L, f, and $g \,\square\, \ell$ at each step, simplifying the computation of iterations. It should be noted that, in this case, the method (1.3.3) exploit the cocoercivity of ∇h, while the method (1.3.6) uses only the fact that ∇h es Lipschitz, although the latter implies the cocoercivity of ∇h [6, Corollary 18.17].

When $h = 0$ and $\ell = \iota_{\{0\}}$, the problem ($\mathcal{P}_0$)-($\mathcal{D}_0$) reduces to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + g(Lx), \tag{1.3.9}$$

$$\underset{u\in\mathcal{G}}{\text{minimize}}\; g^*(u) + f^*(-L^*u). \tag{1.3.10}$$

An efficient algorithm for solving the previous problem is proposed by Chambolle-Pock [20], which provides an accelerated algorithm that has convergence rate of $O(1/k)$ in the case when some of the non-differentiable functions involved in (1.3.9)-(1.3.10) is strongly convex. More precisely, we have:

Theorem 1.3.3 *Suppose that f and g^* are ρ and $\chi-$ strongly convex, respectively, where $(\rho,\chi) \in [0,+\infty]^2$. Let x^0 and $\overline{x}^0$ in $\mathcal{H}$ such that $x^0 = \overline{x}^0$, and let $u^0 \in \mathcal{G}$. Let $(\tau_k)_{k\in\mathbb{N}}$ and $(\gamma_k)_{k\in\mathbb{N}}$ be sequences in $]0,+\infty[$ such that $\tau_0\gamma_0\|L\|^2 \leq 1$, and let $(\theta_k)_{k\in\mathbb{N}} \subset]0,1]$. Consider the routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = \mathrm{prox}_{\gamma_k g^*}(u^k + \gamma_k L\overline{x}^k) \\ x^{k+1} = \mathrm{prox}_{\tau_k f}(x^k - \tau_k L^* u^{k+1}) \\ \overline{x}^{k+1} = x^{k+1} + \theta_k(x^{k+1} - x^k). \end{array} \right. \tag{1.3.11}$$

Then the following holds:

(i) [20, Theorem 1(c)] Suppose that $(\rho,\chi) = (0,0)$, $\tau_0\gamma_0\|L\|^2 < 1$, and $\theta_k \equiv 1$. Then there exists $(\widehat{x},\widehat{u})$ solution to (1.3.9)-(1.3.10) such that $x^k \rightharpoonup \widehat{x}$ and $u^k \rightharpoonup \widehat{u}$.

(ii) [20, Theorem 2] Suppose $\rho > 0$, $\chi = 0$, and $\tau_0\gamma_0\|L\|^2 = 1$. Define

$$(\forall k \in \mathbb{N}) \quad \theta_k = \frac{1}{\sqrt{1+2\rho\tau_k}}, \quad \tau_{k+1} = \theta_k\tau_k, \quad \gamma_{k+1} = \gamma_k/\theta_k, \tag{1.3.12}$$

Then (x^k) converges with rate $O(1/k)$ to the unique primal solution $\widehat{x} \in \mathcal{H}$ to ($\mathcal{P}_0$). More precisely, we have $(\forall\varepsilon > 0)(\exists N_0 \in \mathbb{N})(\forall k \geq N_0)$

$$\|x^k - \widehat{x}\|^2 \leq \frac{1+\varepsilon}{k^2}\left(\frac{\|\widehat{x} - x^0\|^2}{\rho^2\tau_0^2} + \frac{\|L\|^2}{\rho^2}\|\widehat{u} - u^0\|^2\right),$$

for some dual solution $\widehat{u}$ to ($\mathcal{D}_0$).

(iii) [20, Theorem 3] Suppose $\rho > 0$, $\chi > 0$, and $\tau_0\gamma_0\|L\|^2 = 1$. Define

$$\mu = \frac{2\sqrt{\rho\chi}}{\|L\|}. \tag{1.3.13}$$

If we take $\theta_k \equiv \theta \in](1+\mu)^{-1}, 1]$, $\tau_k \equiv \tau$, and $\gamma_k \equiv \gamma$ with

$$\tau = \frac{\mu}{2\rho} \quad \textit{and} \quad \gamma = \frac{\mu}{2\chi}, \tag{1.3.14}$$

we obtain convergence of rate $O(\omega^{k/2})$ in the iterations, with $\omega = (1+\theta)/(2+\mu) < 1$. More precisely, for the unique primal-dual solution $(\widehat{x}, \widehat{u})$ to $(\mathcal{P}_0)$-$(\mathcal{D}_0)$, we have

$$(\forall k \geq 1) \quad \rho\|x^k - \hat{x}\|^2 + (1-\omega)\chi\|u^k - \hat{u}\|^2 \leq \omega^k(\rho\|x^0 - \hat{x}\|^2 + \chi\|u^0 - \hat{u}\|^2).$$

Suppose that $g = \iota_{\{b\}}$ for some $b \in \mathcal{G}$ in (1.3.9)-(1.3.10). In this case $(\mathcal{P}_0)$ reduces to the convex minimization problem with affine linear constraints:

$$\underset{\substack{x\in\mathcal{H} \\ Lx=b}}{\text{minimize}}\, f(x). \tag{1.3.15}$$

Using the identity (1.1.11), we have that $\text{prox}_{\gamma\iota_{\{b\}}^*} = \text{Id} - \gamma \,\text{prox}_{\frac{1}{\gamma}\iota_{\{b\}}} \circ (\text{Id}/\gamma) = \text{Id} - \gamma b$. Then, for the methods (1.3.3),(1.3.8), and (1.3.11), the dual variable update reads

$$u^{k+1} = u^k + \gamma(L\overline{x}^k - b),$$

where $\overline{x}^k$ is the iteration associated with the primal solution. Therefore, the constraint is imposed via the lagrange multiplier update and we do not necessarily have that primal iterations of previous methods satisfy the affine linear constraint [18] (i.e., $Lx^k \neq b$). The latter makes algorithms inefficient in general, as will be illustrated later in the numerical experiences.

If the projection onto $L^{-1}(\{b\})$ is computable, we can ensure feasibility in the primal iterations. Indeed, note that the problem (1.3.15) is equivalent to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\, \iota_{L^{-1}(\{b\})}(x) + f(\text{Id}\, x). \tag{1.3.16}$$

Then, the algorithm (1.3.11) for solving problem (1.3.16) reads as

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = \text{prox}_{\gamma_k f^*}(u^k + \gamma_k \overline{x}^k) \\ x^{k+1} = P_{L^{-1}(\{b\})}(x^k - \tau_k u^{k+1}) \\ \overline{x}^{k+1} = x^{k+1} + \theta_k(x^{k+1} - x^k), \end{array} \right. \tag{1.3.17}$$

Hence, for every $k \in \mathbb{N}$, $Lx^{k+1} = b$. However, $P_{L^{-1}(\{b\})}$ cannot always be calculated efficiently or explicitly, due to the bad conditioning of the matrices involved.

1.3.2. Objectives for the case $V = \mathcal{H}$

The objectives of this first part are the following:

(i) The main objective is propose an efficient convergent algorithm that solves the Problem 1.2.1 in the case $V = \mathcal{H}$. In the context of the problem in (1.2.2) when $T = P_C$, we expect to obtain a generalization of [20, 29] by including a projection step onto C, which ensures that the primal iterates belong to C.

(ii) In the context of the problem in (1.3.15), set $\mathcal{H} = \mathbb{R}^N$, $\mathcal{G} = \mathbb{R}^M$, $L \in \mathbb{R}^{M\times N}$, and $b \in \mathbb{R}^M$. Let $R \in \mathbb{R}^{m\times N}$ and $c \in \mathbb{R}^m$ (with $m < M$) be a selection of rows of L and the corresponding components of b, respectively, such that $P_{R^{-1}c}$ is computable. Through the implementation of a numerical example in (1.3.15), we aim at improving the efficiency of the algorithm (1.3.11) for solving (1.3.15) by adding the projection onto $R^{-1}c$ in the primal iterates of the algorithm (1.3.11), which ensures the feasibility of primal iterates in $R^{-1}c$.

(iii) Under additional hypotheses on the parameters and operators, study accelerated convergence and linear convergence of the proposed projected algorithm.

1.3.3. State of the art for the case $V \subset \mathcal{H}$

Consider a proper closed vector subspace $V \subset \mathcal{H}$, $T = \mathrm{Id}$, and $W = \mathcal{G}$. In this case the Problem 1.2.1 can be written equivalently as

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + \frac{1}{2}L^*((2B \square 2D)(Lx)) + \frac{1}{2}N_V(\mathrm{Id}\ x) + Cx \tag{1.3.18}$$

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad (\exists x \in \mathcal{H}) \begin{cases} -(1/2)L^*u \in Ax + Cx \\ u \in (2B \square 2D)(Lx), \end{cases} \tag{1.3.19}$$

where we used that, for every $\lambda > 0$, $\lambda(B \square D) = (\lambda B) \square (\lambda D)$. Then, the problem (1.3.18)-(1.3.19) is a particular instance of [48, Problem 4.1] with $m = 2$, $L_1 = L$, and $L_2 = \mathrm{Id}$. Thus, (1.3.18)-(1.3.19) can be solved from the algorithm in [48, Theorem 3.1]. In addition, since C is monotone Lipschitz, we can also apply the algorithm in [26, Theorem 3.1].

The disadvantages of the previous algorithms is that they do not exploit the vector space structure and use N_V as any monotone operator. By introducing product space techniques, previous algorithms include additional dual variables, generating high dimensional and slow methods in general.

When $B = 0$ and D is defined by (1.3.4), the problem (1.3.18)-(1.3.19) reduces to

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Cx + N_V x, \tag{1.3.20}$$

which can be solved by [12,13]. In [12], two splitting algorithms are proposed to solve the inclusion (1.3.20) by characterizing its solutions and taking advantage of the fact that V is a closed vector subspace. The first algorithm is the following.

Theorem 1.3.4 *[12, Theorem 4.2] Let $\gamma \in]0, 2\beta[$, let $\alpha = \max\{2/3, 2\gamma/(\gamma + 2\beta)\}$, let $\{\lambda_n\}_{n\in\mathbb{N}} \subset]0, 1/\alpha[$, and let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$. Suppose that*

$$\sum_{n\in\mathbb{N}} \lambda_n(1 - \alpha\lambda_n) = +\infty \ \text{and} \ \sum_{n\in\mathbb{N}} \lambda_n(\|a_n\| + \|b_n\|) < +\infty.$$

Let $z_0 \in \mathcal{H}$ and consider the routine

$$(\forall k \in \mathbb{N}) \left\lfloor \begin{array}{l} x_k = P_V z_k \\ y_k = (x_k - z_k)/\gamma \\ s_k = x_k - \gamma P_V(Cx_k + a_k) + \gamma y_k \\ p_k = J_{\gamma A}(s_k) + b_k \\ z_{k+1} = z_k + \lambda_k(p_k - x_k). \end{array} \right.$$

Then the sequences $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ are in V and $V^\perp$, respectively, and there are $\overline{x} \in \mathcal{H}$ solution to (1.3.20) and $\overline{y} \in V^\perp \cap (A\overline{x} + P_V C\overline{x})$ such that

(i) $x_k \rightharpoonup \overline{x}$ *and* $y_k \rightharpoonup \overline{y}$.

(ii) $x_{k+1} - x_k \to 0$ *and* $y_{k+1} - y_k \to 0$.

(iii) $\sum_{n\in\mathbb{N}} \lambda_n \|P_V(Cx_n - C\overline{x})\|^2 < +\infty$.

To study the second algorithm in [12], the partial inverse operator of a maximally monotone operator is introduced to characterize the solutions to (1.3.20). This operator is first defined by Spingarn [46]. From this characterization, the following theorem provides an algorithm that converges to a solution to problem in (1.3.20).

Theorem 1.3.5 *[12, Theorem 5.2] Let $\gamma > 0$, let $\varepsilon \in \,]0, \max\{1, \beta/\gamma\}[$, let $\{\delta_n\}_{n\in\mathbb{N}} \subset [\varepsilon, (2\beta/\gamma) - \varepsilon]$, let $\{\lambda_n\}_{n\in\mathbb{N}} \subset [\varepsilon, 1]$, let $x_0 \in V$, and let $y_0 \in V^{\perp}$. For every $n \in \mathbb{N}$, consider the routine*

Step 1 . Find $(p_n, q_n) \in \mathcal{H}^2$ such that $x_n - \delta_n \gamma P_V C x_n + \gamma y_n = p_n + \gamma q_n$ and

$$\frac{P_V q_n}{\delta_n} + P_{V^{\perp}} q_n \in A\left(P_V p_n + \frac{P_{V^{\perp}} p_n}{\delta_n}\right).$$

Step 2 Define $x_{n+1} = x_n + \lambda_n(P_V p_n - x_n)$ and $y_{n+1} = y_n + \lambda_n(P_{V^{\perp}} q_n - y_n)$.

Go to step 1.

Then, the sequences $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ are in V and $V^{\perp}$, respectively, and there are $\overline{x} \in \mathcal{H}$ solution to (1.3.20) and $\overline{y} \in V^{\perp} \cap (A\overline{x} + P_V C\overline{x})$ such that

(i) $x_k \rightharpoonup \overline{x}$ *and* $y_k \rightharpoonup \overline{y}$.

(ii) $x_{k+1} - x_k \to 0$ *and* $y_{k+1} - y_k \to 0$.

(iii) $P_V C x_k \to P_V C\overline{x}$.

It is important to note that the sequence $(\delta_n)_{n\in\mathbb{N}}$ in Theorem 1.3.5 can be defined in order to accelerate the convergence of the algorithm. However, as shown in [46], Step 1 in the previous Theorem 1.3.5 is not always easy to compute. As a consequence of Theorem 1.3.5 by setting $\delta_n \equiv 1$, we have the following particular case where Step 1 can be obtained explicitly when the resolvent of A is computable.

Corollary 1.3.6 *[12, Corollary 5.3] Let $\gamma \in \,]0, 2\beta[$, let $x_0 \in V$, let $y_0 \in V^{\perp}$, let $\varepsilon \in \,]0, 1]$, and let $\{\lambda_n\}_{n\in\mathbb{N}} \subset [\varepsilon, 1]$. Consider the following routine.*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} s_n = x_n - \gamma P_V C x_n + \gamma y_n \\ p_n = J_{\gamma A} s_n \\ y_{n+1} = y_n + \frac{\lambda_n}{\gamma}(P_V p_n - p_n) \\ x_{n+1} = x_n + \lambda_n(P_V p_n - x_n) \end{array} \right. \tag{1.3.21}$$

Then, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are in V and $V^{\perp}$ respectively. Moreover, there are $\overline{x} \in \mathcal{H}$ solution to (1.3.20) and $\overline{y} \in V^{\perp} \cap (A\overline{x} + P_V C\overline{x})$ such that

(i) $x_n \rightharpoonup \overline{x}$ *and* $y_n \rightharpoonup \overline{y}$.

(ii) $x_{n+1} - x_n \to 0$ *and* $y_{n+1} - y_n \to 0$.

(iii) $P_V C x_n \to P_V C\overline{x}$.

1.3.4. Objectives for the case $V \subset \mathcal{H}$

The objectives for the second part of this thesis are:

(i) Characterize the solutions of Problem 1.2.1 in their full generality by using the partial inverse operator, as it has been used in [1, 12, 13] for solving particular instances of Problem 1.2.1.

(ii) Generalize the results obtained in the case $V = \mathcal{H}$, by considering the a priori information modeled by $\operatorname{Fix} T$ and proposing a weak convergent primal-dual algorithm for solving Problem 1.2.1 separating the iterates that belong to V with those that belong to $V^{\perp}$, as it is shown in algorithm (1.3.21).

The purpose of including the operator T in the inclusion is the same as in the objectives of the previous section, i.e., in the case when $T = P_C$, where $C \subset \mathcal{H}$ is a nonempty closed convex, the idea is incorporate this projection in the primal iterates of the proposed algorithm in order to ensure feasibility in the iterations at least in C and thus generate more efficient methods.

(iii) Perform numerical experiences applying the proposed algorithm to the context in (1.2.2) when V is a proper closed subspace, comparing its efficiency with methods already existing in the literature.

(iv) Apply the proposed method to the arc capacity expansion problem in stochastic traffic networks and illustrate the advantages of the vector subspace structure of the problem.

Chapter 2

Monotone inclusions with a priori information

2.1. Summary

In this chapter we provide a convergent algorithm for solving the Problem 1.2.1 in the case when $V = \mathcal{H}$, which reduces to the following primal-dual monotone inclusion problem

$$\text{find} \quad x \in \operatorname{Fix} T \quad \text{such that} \quad 0 \in Ax + L^*((B \square D)(Lx)) + Cx \tag{2.1.1}$$

$$\text{find} \quad u \in W \quad \text{such that} \quad (\exists x \in \operatorname{Fix} T) \quad \begin{cases} -L^*u \in Ax + Cx \\ u \in (B \square D)(Lx). \end{cases} \tag{2.1.2}$$

The operator T represents a priori information on the solutions and has a practical use in the context of convex optimization. Namely, when T is the projection onto a closed convex set C, the proposed method allows that primal iterates belong to C.

The main result of this chapter is the following generalization of Theorem 1.3.3.

Theorem 2.1.1 *Let $\gamma_0 \in \,]0, 2\delta[$ and $\tau_0 \in \,]0, 2\beta[$ be such that*

$$\|L\|^2 \leq \left(\frac{1}{\tau_0} - \frac{1}{2\beta}\right)\left(\frac{1}{\gamma_0} - \frac{1}{2\delta}\right) \tag{2.1.3}$$

and let $(x^0, \overline{x}^0, u^0) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}$ such that $\overline{x}^0 = x^0$. Let $(\theta_k)_{k\in\mathbb{N}}$, $(\gamma_k)_{k\in\mathbb{N}}$, and $(\tau_k)_{k\in\mathbb{N}}$ be sequences in $]0, 1]$, $]0, 2\delta[$ and $]0, 2\beta[$, respectively, and consider

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \eta^{k+1} = J_{\gamma_k B^{-1}}(u^k + \gamma_k(L\overline{x}^k - D^{-1}u^k)) \\ u^{k+1} = P_V\, \eta^{k+1} \\ p^{k+1} = J_{\tau_k A}(x^k - \tau_k(L^*u^{k+1} + Cx^k)) \\ x^{k+1} = T\, p^{k+1} \\ \theta_k = \frac{1}{\sqrt{1+2\rho\tau_k}} \\ \tau_{k+1} = \theta_k \tau_k \\ \gamma_{k+1} = \gamma_k/\theta_k \\ \overline{x}^{k+1} = x^{k+1} + \theta_k(p^{k+1} - x^k). \end{array} \right. \tag{2.1.4}$$

Then, the following hold.

(I) *Suppose that* $(\rho,\chi)=(0,0)$*, set* $\theta_k\equiv 1$*,* $\tau_k\equiv\tau$*,* $\gamma_k\equiv\gamma$*, and assume that* (2.1.3) *holds with strict inequality. Then, there exists a solution* $(\overline{x},\overline{u})$ *to problem* (2.1.1)-(2.1.2) *such that* $x^k \rightharpoonup \overline{x}$ *and* $u^k \rightharpoonup \overline{u}$*.*

(II) *Suppose that* $\rho>0$*,* $\chi=0$*,* $D^{-1}=0$*, and assume that* (2.1.3) *holds with equality. Then for the unique solution* $\overline{x}$ *to problem* (2.1.1)*, we have* $(\forall\varepsilon>0)(\exists N_0\in\mathbb{N})(\forall k\geq N_0)$

$$\left\|x^k-\overline{x}\right\|^2\leq\frac{1+\varepsilon}{k^2}\left(\frac{\left\|x^0-\overline{x}\right\|^2}{\rho^2\tau_0^2}+\frac{2\beta\|L\|^2}{\rho^2(2\beta-\tau_0)}\left\|u^0-\overline{u}\right\|^2\right),$$

where $\overline{u}$ *is a solution to problem* (2.1.2)*.*

(III) *Suppose that* $\rho>0$*,* $\chi>0$*, and define*

$$\mu=\frac{2\sqrt{\rho\chi}}{\|L\|}\quad\text{and}\quad\alpha=\min\left\{\frac{\mu\rho}{\rho+\frac{\mu}{4\beta}},\frac{\mu\chi}{\chi+\frac{\mu}{4\delta}}\right\}.$$

Set $\theta_k\equiv\theta\in\left]{(1+\alpha)^{-1}},1\right]$*,* $\tau_k\equiv 2\beta\mu/(\mu+4\beta\rho)$*, and* $\gamma_k\equiv 2\mu\delta/(\mu+4\delta\chi)$*. Then, for the unique solution* $(\overline{x},\overline{u})$ *to problem* (2.1.1)-(2.1.2)*, we have* $(\forall k\in\mathbb{N})$

$$\begin{aligned}\left(\chi(1-\omega)+\frac{\mu}{4\delta}\right)\left\|u^k-\overline{u}\right\|^2+\left(\rho+\frac{\mu}{4\beta}\right)\left\|x^k-\overline{x}\right\|^2&\\ \leq\omega^k\left(\left(\chi+\frac{\mu}{4\delta}\right)\left\|u^0-\overline{u}\right\|^2+\left(\rho+\frac{\mu}{4\beta}\right)\left\|x^0-\overline{x}\right\|^2\right),&\end{aligned}$$

where $\omega=(1+\theta)/(2+\alpha)\in\left](1+\alpha)^{-1},\theta\right[$*.*

In the previous results, the weak convergence of the algorithm in (2.1.4) is provided by (I) when there is no presence of strong monotony in the operators A and B^{-1}. In the case when A is a $\rho-$strongly monotone operator, (II) claims that the sequences generated by (2.1.4) converges in a rate of $O(1/k)$. Finally, we prove that if A and B^{-1} are ρ and $\chi-$ strongly monotone, respectively, we obtain convergence of rate $O(w^k)$, where $w\in\,]0,1[$ depends of the parameters of (2.1.1)-(2.1.2). We observe that the uniqueness of solutions in (II) and (III) is deduced from the strong monotony of the corresponding monotone operators.

Moreover, by setting $T=\operatorname{Id}$, the proposed method generalizes [48] in the case of monotone operators, and [29] in the case of convex functions. In addition, when $C=D^{-1}=0$, we have that $\beta=\delta=+\infty$. Therefore, (II) and (III) generalizes the results in Theorem 1.3.3 in the cases when $\rho>0$ or $\chi>0$.

In order to test the efficiency of our method, we compare our algorithm (2.1.4) with respect the Chambolle-Pock's algorithm (1.3.11) in the problem of minimize the ℓ^1-norm subject to affine linear constraints:

Problem 2.1.2 Let $R\in\mathbb{R}^{m\times N}$ such that $\ker R^\top=\{0\}$, let $S\in\mathbb{R}^{n\times N}$, let $c\in\mathbb{R}^m$, and let $d\in\mathbb{R}^n$. The problem is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\,\|x\|_1\quad\text{s.t.}\quad Rx=c\quad\text{and}\quad Sx=d,$$

under the assumption that solutions exist.

In this application, we consider $T = P_{R^{-1}c}$, which is computable by the hypothesis on the matrix R. The results show that by increasing the number of constraints on which it is possible to project (i.e. when m increases), the percentage of improvement in the computational time with respect to the Chambolle-Pock's method increases.

2.2. Article: "A Projected Primal-Dual Method for Solving Constrained Monotone Inclusions"

Abstract

In this paper, we provide an algorithm for solving constrained composite primal– dual monotone inclusions, i.e., monotone inclusions in which a priori information on primal–dual solutions is represented via closed and convex sets. The proposed algorithm incorporates a projection step onto the a priori information sets and generalizes methods proposed in the literature for solving monotone inclusions. Moreover, under the presence of strong monotonicity, we derive an accelerated scheme inspired on the primal–dual algorithm applied to the more general context of constrained monotone inclusions. In the particular case of convex optimization, our algorithm generalizes several primal–dual optimization methods by allowing a priori information on solutions. In addition, we provide an accelerated scheme under strong convexity. An application of our approach with a priori information is constrained convex optimization problems, in which available primal–dual methods impose constraints via Lagrange multiplier updates, usually leading to slow algorithms with unfeasible primal iterates. The proposed modification forces primal iterates to satisfy a selection of constraints onto which we can project, obtaining a faster method as numerical examples exhibit. The obtained results extend and improve several results in the literature.

2.2.1. Introduction

This paper is devoted to the numerical resolution of composite primal-dual monotone inclusions in which a priori information on solutions is known. The a priori information on primal-dual solutions is represented via closed and convex sets in primal and dual spaces, following some ideas developed in [15, 47]. We force primal-dual iterates to belong to these information sets by adding additional projections on primal-dual iterates in each iteration of our proposed method.

The advantage of our formulation is illustrated in composite convex optimization with affine linear equality constraints. In this context, the primal-dual methods proposed in [20, 21, 29, 31, 35, 37, 48] impose feasibility through Lagrange multiplier updates. A disadvantage of this approach is that such algorithms are usually slow and their primal iterates do not necessarily satisfy any of the constraints (see, e.g., [18]), leading to unfeasible approximate primal solutions. By projecting onto the affine subspace generated by the constraints, previous problem is solved. However, in several applications this projection is not easy to compute because of bad conditioning on the linear system (see, e.g. [17]). In this context, the a priori information on primal solutions can be set as any selection of the affine linear constraints. Indeed, since any solution is feasible, we know it must satisfy any selection of the constraints. Even if in the previous context the formulation with a priori information may be seen as artificial, it allows us to propose a method with an additional projection onto an arbitrary selection of the constraints, which improves its efficiency (see Section 2.2.5). This method forces primal iterates to satisfy the selected constraints, which can be chosen in order to compute the projection easily.

In this paper, we provide a new projected primal-dual splitting method for solving constrained monotone inclusions, i.e., inclusions in which we count on a priori information on primal-dual solutions. We also provide an accelerated scheme of our method in the presence of strong monotonicity and we derive linear convergence in the fully strongly monotone case. In the case without a priori information, our results give an accelerated scheme of the method proposed in [48] for strongly monotone inclusions. A similar approach in the case without a priori information is used in [9] with a different way to set the step-sizes in order to exploit the strong convexity of the problem. In the context of convex optimization, our method generalize the algorithms proposed in [20, 29] and [37] without inertia, by incorporating a projection onto an a priori primal-dual information set. Our method is applied in the context of convex optimization with equality constraints, when the a priori information set is chosen as a selection of the affine linear constraints. The advantages of this approach with respect to classical primal-dual approaches are justified via numerical examples. Our acceleration scheme in the convex optimization context is obtained as a generalization of [20], complementing the ergodic rates obtained in the case without projection in [21] and, as far as we know, have not been developed in the literature and are interesting in their own right.

The paper is organized as follows. We set our notation and we give a brief background in Section 2.2.2. We propose our algorithm and the main results in Section 2.2.3, together with connections with existing methods in the literature. In Section 2.2.4, we apply previous results to convex optimization problems with equality affine linear constraints. Numerical experiences illustrating the improvement in the efficiency of the algorithm with the additional projection are performed in Section 2.2.5. We finish with some conclusions in Section 2.2.6.

2.2.2. Notation and Preliminaries

Throughout this paper, $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces. We denote the scalar products of $\mathcal{H}$ and $\mathcal{G}$ by $\langle \cdot \mid \cdot \rangle$ and the associated norms by $\| \cdot \|$. The projector operator onto a nonempty, closed, and convex set $C \subset \mathcal{H}$ is denoted by P_C and, for a set-valued operator $M : \mathcal{H} \rightrightarrows \mathcal{H}$, we use $\operatorname{ran}(M)$ for the range of M, $\operatorname{gra}(M)$ for its graph, M^{-1} for its inverse, $J_M = (\operatorname{Id} + M)^{-1}$ for its resolvent, and $\Box$ stands for the parallel sum as in [6]. Moreover, M is ρ-strongly monotone for $\rho \geq 0$ iff, for every (x, u) and (y, v) in $\operatorname{gra}(M)$, $\langle x - y \mid u - v \rangle \geq \rho \|x - y\|^2$, it is $\rho-$cocoercive iff M^{-1} is $\rho-$strongly monotone, M is monotone iff it is ρ-strongly monotone with $\rho = 0$, and it is maximally monotone iff its graph is maximal, in the sens of inclusions in $\mathcal{H} \times \mathcal{H}$, among the graphs of monotone operators. The class of all lower semicontinuous convex functions $f : \mathcal{H} \to (-\infty, +\infty]$ such that $\operatorname{dom}(f) = \{x \in \mathcal{H} : f(x) < +\infty\} \neq \varnothing$ is denoted by $\Gamma_0(\mathcal{H})$ and, for every $f \in \Gamma_0(\mathcal{H})$, the Fenchel conjugate of f is denoted by f^*, its subdifferential by ∂f, and its proximity operator by prox_f, as in [6]. We recall that $(\partial f)^{-1} = \partial f^*$ and $J_{\partial f} = \operatorname{prox}_f$. In addition, when $C \subset \mathcal{H}$ is a nonempty, closed, and convex set, we have that $J_{\partial \iota_C} = \operatorname{prox}_{\iota_C} = P_C$, where ι_C is the indicator function of C, which is 0 in C and $+\infty$ otherwise. Moreover, for $f, g \in \Gamma_0(\mathcal{H})$, we denote by $f \,\Box\, g \colon x \mapsto \inf_{y \in \mathcal{H}} \big(f(y) + g(x - y)\big)$ the infimal convolution of f and g. Moreover, if $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$, where $\operatorname{sri} C$ stands for the strong relative interior of a nonempty set $C \subset \mathcal{H}$, we have $\partial(f \,\Box\, g) = (\partial f) \,\Box\, (\partial g)$ [6, Proposition 15.7(i)& Proposition 25.32]. Given $\alpha \in]0, 1[$, an operator $T \colon \mathcal{H} \to \mathcal{H}$ satisfying $\operatorname{Fix} T \neq \varnothing$ is $\alpha-$averaged quasi-nonexpansive iff, for every $x \in \mathcal{H}$ and $y \in \operatorname{Fix} T$, we have $\|Tx - y\|^2 \leq \|x - y\|^2 - (\frac{1-\alpha}{\alpha})\|x - Tx\|^2$. We refer the reader to [6] for definitions and further results in monotone operator theory and convex optimization.

2.2.3. Problem and Main Results

We consider the following problem.

Problem 2.2.1 Let $T \colon \mathcal{H} \to \mathcal{H}$ be an $\alpha-$averaged quasi-nonexpansive operator with $\alpha \in]0, 1[$, let V be a closed vector subspace of $\mathcal{G}$, let $L \colon \mathcal{H} \to \mathcal{G}$ be a nonzero linear bounded operator satisfying

$\operatorname{ran} L \subset V$, let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $D\colon \mathcal{G} \rightrightarrows \mathcal{G}$ be maximally monotone operators which are ρ and $\delta-$strongly monotone, respectively, and let $B : \mathcal{G} \rightrightarrows \mathcal{G}$ and $C\colon \mathcal{H} \to \mathcal{H}$ be χ and $\beta-$cocoercive, respectively, for $(\rho, \chi) \in [0, +\infty[^2$ and $(\delta, \beta) \in]0, +\infty]^2$. The problem is to solve the primal and dual inclusions

$$\text{find} \quad \hat{x} \in \operatorname{Fix} T \quad \text{such that} \quad 0 \in A\hat{x} + L^*(B \,\square\, D)(L\hat{x}) + C\hat{x} \tag{$\mathcal{P}$}$$

$$\text{find} \quad \hat{u} \in V \quad \text{such that} \quad (\exists\, \hat{x} \in \operatorname{Fix} T) \quad \begin{cases} -L^*\hat{u} \in A\hat{x} + C\hat{x} \\ \hat{u} \in (B \,\square\, D)(L\hat{x}), \end{cases} \tag{$\mathcal{D}$}$$

under the assumption that solutions exist.

When $A = \partial f$, $B = \partial g$, $C = \nabla h$, and $D = \partial \ell$, where $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $h\colon \mathcal{H} \to \mathbb{R}$ is a differentiable convex function with $\beta^{-1}-$Lipschitz gradient, and $\ell \in \Gamma_0(\mathcal{G})$ is $\delta-$strongly convex, Problem 2.2.1 reduces to

$$\text{find} \quad \hat{x} \in \operatorname{Fix} T \cap \operatorname*{argmin}_{x \in \mathcal{H}} F(x) := f(x) + (g \,\square\, \ell)(Lx) + h(x) \tag{$\mathcal{P}_0$}$$

together with the dual problem

$$\text{find} \quad \hat{u} \in V \cap \operatorname*{argmin}_{u \in \mathcal{G}} g^*(u) + (f^* \,\square\, h^*)(-L^*u) + \ell^*(u), \tag{$\mathcal{D}_0$}$$

assuming that some qualification condition holds. Note that, when $T = P_X$, any solution to ($\mathcal{P}_0$) is a solution to $\min_{x \in X} F(x)$, but the converse is not true. The set X in this case represents an a priori information on the primal solution. As you can see in the next section, an application of this formulation is constrained convex optimization, in which X may represent a selection of the affine linear constraints. Even if, in this case, the formulation can be set without considering the set X, its artificial appearance has a practical relevance: the method obtained include a projection onto X which helps to the performance of the method as stated in numerical section.

When $\rho = \chi = 0$, $V = \mathcal{G}$ and $T = \operatorname{Id}$, ($\mathcal{P}_0$)-($\mathcal{D}_0$) can be solved by using [26, Theorem 4.2] or [37, Theorem 5]. In the last method, inertial terms are also included. In the case when $\ell^* = 0$, the algorithm in [29] can be used and if $\ell^* = h = 0$, ($\mathcal{P}_0$)-($\mathcal{D}_0$) can be solved by [14,20] or a version of [20] with line-search proposed in [38]. In [20], the strong convexity is exploited via acceleration schemes. Moreover, when $T = P_X$, $X \subset \mathcal{H}$ is nonempty, closed and convex, $V = \mathcal{G}$ and $\ell^* = h = 0$, ($\mathcal{P}_0$)-($\mathcal{D}_0$) is solved in [18, Theorem 3.1]. When $\rho > 0$ or $\chi > 0$, ergodic convergence rates are derived in [21] when $V = \mathcal{G}$ and $T = \operatorname{Id}$. In its whole generality, as far as we know, ($\mathcal{P}_0$)-($\mathcal{D}_0$) has not been solved and strong convexity has not been exploited.

In Problem 2.2.1, set $T = \operatorname{Id}$, and, for every $i \in \{1, \dots, m\}$, let $L_i\colon \mathcal{H} \to G_i$ be linear and bounded, $B_i\colon G_i \rightrightarrows G_i$ and $D_i\colon G_i \rightrightarrows G_i$ be maximally monotone operators such that D_i is strongly monotone, let $\omega_i > 0$ be such that $\sum_{i=1}^m \omega_i = 1$, and set

$$\begin{aligned}
&V = \mathcal{G} = G_1 \oplus \cdots \oplus G_m \\
&L\colon \mathcal{H} \to \mathcal{G}\colon x \mapsto (L_1 x, \dots, L_m x) \\
&B\colon (u_1, \dots, u_m) \mapsto \omega_1 B_1 u_1 \times \cdots \times \omega_m B_m u_m \\
&D\colon (u_1, \dots, u_m) \mapsto \omega_1 D_1 u_1 \times \cdots \times \omega_m D_m u_m.
\end{aligned}$$

Then, Problem 2.2.1 reduces to [26, Problem 1.1] (see also [48, Problem 1.1]). We prefer to set $m = 1$ for simplicity. In [26], previous problem is solved when C is monotone and Lipschitz by applying the method in [47] to the product primal-dual space. Accelerated versions of previous algorithm under strong monotonicity are proposed in [10] and a different approach for exploiting strong monotonicity is used in [9] in this case. An inertial version in the previous context is developed in [8]. The cocoercivity of C is exploited in [48], where an algorithm is proposed for solving Problem 2.2.1 when $\rho=\chi=0$, $V=\mathcal{G}$ and $T=\operatorname{Id}$. In the following theorem, we provide an algorithm for solving Problem 2.2.1 in its whole generality with weak convergence to a solution when the step sizes are fixed. Moreover, when A or B^{-1} are strongly monotone ($\rho>0$ or $\chi>0$), we provide an accelerated version inspired on (and generalizing) [20, Section 5.1]. Finally, we generalize [20, Section 5.2] for obtaining linear convergence when $\rho>0$ and $\chi>0$.

Theorem 2.2.2 *Let $\gamma_0\in]0,2\delta[$ and $\tau_0\in]0,2\beta[$ be such that*

$$\|L\|^2\leq\left(\frac{1}{\tau_0}-\frac{1}{2\beta}\right)\left(\frac{1}{\gamma_0}-\frac{1}{2\delta}\right) \tag{2.2.1}$$

and let $(x^0,\bar{x}^0,u^0)\in\mathcal{H}\times\mathcal{H}\times\mathcal{G}$ such that $\bar{x}^0=x^0$. Let $(\theta_k)_{k\in\mathbb{N}}$, $(\gamma_k)_{k\in\mathbb{N}}$ and $(\tau_k)_{k\in\mathbb{N}}$ be sequences in $]0,1]$, $]0,2\delta[$ and $]0,2\beta[$, respectively, and consider

$$(\forall k\in\mathbb{N})\quad\left\lfloor\begin{array}{l}\eta^{k+1}=J_{\gamma_kB^{-1}}(u^k+\gamma_k(L\bar{x}^k-D^{-1}u^k))\\ u^{k+1}=P_V\,\eta^{k+1}\\ p^{k+1}=J_{\tau_kA}(x^k-\tau_k(L^*u^{k+1}+Cx^k))\\ x^{k+1}=T\,p^{k+1}\\ \bar{x}^{k+1}=x^{k+1}+\theta_k(p^{k+1}-x^k).\end{array}\right. \tag{2.2.2}$$

Then, the following hold.

(I) *For every $k\geq1$ and for every solution $(\hat{x},\hat{u})$ to Problem 2.2.1, we have*

$$\begin{aligned}\frac{\|x^k-\hat{x}\|^2}{\tau_k}+\frac{\|u^k-\hat{u}\|^2}{\gamma_k}\geq{}&(2\rho\tau_k+1)\frac{\|p^{k+1}-\hat{x}\|^2}{\tau_k}+\left\|p^{k+1}-x^k\right\|^2\left(\frac{1}{\tau_k}-\frac{1}{2\beta}\right)\\&+(2\chi\gamma_k+1)\frac{\|\eta^{k+1}-\hat{u}\|^2}{\gamma_k}+\|\eta^{k+1}-u^k\|^2\left(\frac{1}{\gamma_k}-\frac{1}{2\delta}\right)\\&+2\big\langle L(p^{k+1}-x^k)\mid\eta^{k+1}-\hat{u}\big\rangle-2\theta_{k-1}\big\langle L(p^k-x^{k-1})\mid\eta^k-\hat{u}\big\rangle\\&-2\theta_{k-1}\|L\|\|p^k-x^{k-1}\|\|\eta^{k+1}-u^k\|.\end{aligned} \tag{2.2.3}$$

(II) *Suppose that $(\rho,\chi)=(0,0)$, set $\theta_k\equiv1$, $\tau_k\equiv\tau$, $\gamma_k\equiv\gamma$ and assume that (2.2.1) holds with strict inequality. Then, there exists a solution $(\bar{x},\bar{u})$ to Problem 2.2.1 such that $x^k\rightharpoonup\bar{x}$ and $u^k\rightharpoonup\bar{u}$.*

(III) *Suppose that $\rho>0$, $\chi=0$, and $D^{-1}=0$. If we set*

$$(\forall k\in\mathbb{N})\quad\theta_k=\frac{1}{\sqrt{1+2\rho\tau_k}},\quad\tau_{k+1}=\theta_k\tau_k,\quad\gamma_{k+1}=\gamma_k/\theta_k, \tag{2.2.4}$$

and we assume that (2.2.1) holds with equality, we obtain, for every solution $(\hat{x},\hat{u})$ to Problem 2.2.1, $(\forall\varepsilon>0)(\exists N_0\in\mathbb{N})(\forall k\geq N_0)$

$$\left\|x^k-\hat{x}\right\|^2\leq\frac{1+\varepsilon}{k^2}\left(\frac{\left\|x^0-\hat{x}\right\|^2}{\rho^2\tau_0^2}+\frac{2\beta\|L\|^2}{\rho^2(2\beta-\tau_0)}\left\|u^0-\hat{u}\right\|^2\right).$$

(IV) *Suppose that $\rho>0$ and $\chi>0$ and define*

$$\mu=\frac{2\sqrt{\rho\chi}}{\|L\|}\quad\text{and}\quad\alpha=\min\left\{\frac{\mu\rho}{\rho+\frac{\mu}{4\beta}},\frac{\mu\chi}{\chi+\frac{\mu}{4\delta}}\right\}. \tag{2.2.5}$$

If we set $\theta_k \equiv \theta \in \left]\,(1+\alpha)^{-1}, 1\right]$, $\tau_k \equiv \tau$ *and* $\gamma_k \equiv \gamma$ *with*

$$\tau = \frac{2\beta\mu}{\mu + 4\beta\rho} \quad \textit{and} \quad \gamma = \frac{2\mu\delta}{\mu + 4\delta\chi}, \tag{2.2.6}$$

we obtain linear convergence. That is, for every $k \in \mathbb{N}$,

$$\begin{aligned}\left(\chi(1-\omega) + \frac{\mu}{4\delta}\right)\left\|u^k - \hat{u}\right\|^2 + \left(\rho + \frac{\mu}{4\beta}\right)\left\|x^k - \hat{x}\right\|^2 \\ \leq \omega^k \left(\left(\chi + \frac{\mu}{4\delta}\right)\left\|u^0 - \hat{u}\right\|^2 + \left(\rho + \frac{\mu}{4\beta}\right)\left\|x^0 - \hat{x}\right\|^2\right),\end{aligned}$$

where $\omega = (1+\theta)/(2+\alpha) \in \left]\,(1+\alpha)^{-1}, \theta\,\right[$.

Proof. (I): Fix $k \in \mathbb{N}$ and let $(\hat{x}, \hat{u})$ be a solution to Problem 2.2.1. We have $\hat{x} \in \operatorname{Fix} T$, $\hat{u} \in V$ and, using $B \,\square\, D = (B^{-1} + D^{-1})^{-1}$, we deduce $-(L^*\hat{u} + C\hat{x}) \in A\hat{x}$ and $L\hat{x} - D^{-1}\hat{u} \in B^{-1}\hat{u}$. Moreover, it follows from (2.2.2) that

$$\begin{cases} \frac{x^k - p^{k+1}}{\tau_k} - L^* u^{k+1} - Cx^k \in Ap^{k+1} \\ \frac{u^k - \eta^{k+1}}{\gamma_k} + L\bar{x}^k - D^{-1}u^k \in B^{-1}\eta^{k+1}. \end{cases} \tag{2.2.7}$$

Therefore, since A and B^{-1} are ρ and χ-strongly monotone, respectively, we deduce

$$\begin{aligned}\left\langle \frac{x^k - p^{k+1}}{\tau_k} - L^*(u^{k+1} - \hat{u}) \,\middle|\, p^{k+1} - \hat{x} \right\rangle + \left\langle \frac{u^k - \eta^{k+1}}{\gamma_k} + L(\bar{x}^k - \hat{x}) \,\middle|\, \eta^{k+1} - \hat{u} \right\rangle \\ - \left\langle Cx^k - C\hat{x} \mid p^{k+1} - \hat{x} \right\rangle - \left\langle D^{-1}u^k - D^{-1}\hat{u} \mid \eta^{k+1} - \hat{u} \right\rangle \\ \geq \rho\left\|p^{k+1} - \hat{x}\right\|^2 + \chi\|\eta^{k+1} - \hat{u}\|^2.\end{aligned} \tag{2.2.8}$$

The cocoercivity of C and D^{-1}, and $ab \leq \beta a^2 + b^2/(4\beta)$ yield

$$\begin{aligned}\left\langle Cx^k - C\hat{x} \mid p^{k+1} - \hat{x} \right\rangle &= \left\langle Cx^k - C\hat{x} \mid p^{k+1} - x^k \right\rangle + \left\langle Cx^k - C\hat{x} \mid x^k - \hat{x} \right\rangle \\ &\geq -\|Cx^k - C\hat{x}\|\|p^{k+1} - x^k\| + \beta\|Cx^k - C\hat{x}\|^2 \\ &\geq -\frac{\|p^{k+1} - x^k\|^2}{4\beta},\end{aligned}$$

and, analogously, $\left\langle D^{-1}u^k - D^{-1}\hat{u} \mid \eta^{k+1} - \hat{u} \right\rangle \geq -\frac{\|\eta^{k+1} - u^k\|^2}{4\delta}$. Hence, by using [6, Lemma 2.12(i)] in (2.2.8), we deduce

$$\begin{aligned}\frac{\|x^k - \hat{x}\|^2}{\tau_k} + \frac{\|u^k - \hat{u}\|^2}{\gamma_k} \geq \left(2\rho + \frac{1}{\tau_k}\right)\left\|p^{k+1} - \hat{x}\right\|^2 + \left(2\chi + \frac{1}{\gamma_k}\right)\|\eta^{k+1} - \hat{u}\|^2 \\ + 2\left[\left\langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u} \right\rangle - \left\langle L(\bar{x}^k - \hat{x}) \mid \eta^{k+1} - \hat{u} \right\rangle\right] \\ + \|\eta^{k+1} - u^k\|^2\left(\frac{1}{\gamma_k} - \frac{1}{2\delta}\right) + \|p^{k+1} - x^k\|^2\left(\frac{1}{\tau_k} - \frac{1}{2\beta}\right).\end{aligned} \tag{2.2.9}$$

Moreover, (2.2.2), $\operatorname{ran}(L) \subset V$ and $(u^k - \eta^k)_{k\geq 1} \subset V^\perp$ yield, for every $k \geq 1$,

$$\begin{aligned}
&\left\langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u}\right\rangle - \left\langle L(\bar{x}^k - \hat{x}) \mid \eta^{k+1} - \hat{u}\right\rangle \\
&\quad = \left\langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u}\right\rangle - \left\langle L(x^k - \hat{x}) \mid \eta^{k+1} - \hat{u}\right\rangle \\
&\qquad - \theta_{k-1}\left\langle L(p^k - x^{k-1}) \mid \eta^{k+1} - \hat{u}\right\rangle \\
&\quad = \left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - \theta_{k-1}\left\langle L(p^k - x^{k-1}) \mid \eta^{k+1} - \hat{u}\right\rangle \\
&\quad = \left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - \theta_{k-1}\left\langle L(p^k - x^{k-1}) \mid \eta^{k+1} - u^k\right\rangle \\
&\qquad - \theta_{k-1}\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle \\
&\quad \geq \left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - \theta_{k-1}\|L\|\|p^k - x^{k-1}\|\|\eta^{k+1} - u^k\| \\
&\qquad - \theta_{k-1}\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle,
\end{aligned}$$

which, together with (2.2.9), yield (2.2.3).

(ii): For every $k \in \mathbb{N}$, it follows from Theorem 2.2.2((i)), $\rho = \chi = 0$, $\theta_k \equiv 1$, $\tau_k \equiv \tau$, $\gamma_k \equiv \gamma$

$$\begin{aligned}
\frac{\|p^k - \hat{x}\|^2}{\tau} + \frac{\|\eta^k - \hat{u}\|^2}{\gamma} &\geq \frac{\|p^{k+1} - \hat{x}\|^2}{\tau} + \left(\frac{1-\alpha}{\alpha}\right)\frac{\|x^k - p^k\|^2}{\tau} + \frac{\|u^k - \eta^k\|^2}{\gamma} \\
&\quad + \frac{\|\eta^{k+1} - \hat{u}\|^2}{\gamma} + \|p^{k+1} - x^k\|^2\left(\frac{1}{\tau} - \frac{1}{2\beta}\right) + \|\eta^{k+1} - u^k\|^2\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right) \\
&\quad + 2\left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - 2\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle \\
&\quad - 2\|L\|\|p^k - x^{k-1}\|\|\eta^{k+1} - u^k\| \\
&\geq \frac{\|p^{k+1} - \hat{x}\|^2}{\tau} + \frac{\|u^k - \eta^k\|^2}{\gamma} + \|\eta^{k+1} - u^k\|^2\left(\frac{1}{\gamma} - \frac{1}{2\delta} - \frac{1}{\nu}\right) \\
&\quad + \left(\frac{1-\alpha}{\alpha}\right)\frac{\|x^k - p^k\|^2}{\tau} + \frac{\|\eta^{k+1} - \hat{u}\|^2}{\gamma} + \|p^{k+1} - x^k\|^2\left(\frac{1}{\tau} - \frac{1}{2\beta}\right) \\
&\quad + 2\left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - 2\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle \\
&\quad - \nu\|L\|^2\|p^k - x^{k-1}\|^2,
\end{aligned} \tag{2.2.10}$$

where the first inequality follows from the $\alpha-$averaged quasi-nonexpansiveness of T and the firm nonexpansiveness of P_V, and the last inequality holds for every $\nu > 0$. If we let $\varepsilon = \left[\left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right) - \|L\|^2\right]\left(\frac{\beta\tau}{2\beta-\tau}\right) > 0$, and we choose $\nu = (\frac{1}{\gamma} - \frac{1}{2\delta} - \varepsilon)^{-1} > 0$, we have $\nu\|L\|^2 = (\frac{1}{\tau} - \frac{1}{2\beta}) - \nu\varepsilon(\frac{1}{\tau} - \frac{1}{2\beta})$. Hence, from (2.2.10) we have

$$\begin{aligned}
\Upsilon_k + \frac{\left\|p^k - \hat{x}\right\|^2}{\tau} \geq \Upsilon_{k+1} + \frac{\left\|p^{k+1} - \hat{x}\right\|^2}{\tau} + \left(\frac{1-\alpha}{\alpha}\right)\frac{\|x^k - p^k\|^2}{\tau} + \frac{\|u^k - \eta^k\|^2}{\gamma} \\
+ \varepsilon\|\eta^{k+1} - u^k\|^2 + \nu\varepsilon\left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\|p^k - x^{k-1}\|^2,
\end{aligned} \tag{2.2.11}$$

where, for every $k \in \mathbb{N}$,

$$\Upsilon_k = \frac{\|\eta^k - \hat{u}\|^2}{\gamma} + 2\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle + \left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\|p^k - x^{k-1}\|^2.$$

Note that from (2.2.1) we have, for every $k \in \mathbb{N}$,

$$\begin{aligned}
\Upsilon_k &\geq \frac{\|\eta^k - \hat{u}\|^2}{\gamma} + 2\big\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\big\rangle + \frac{\|L\|^2}{\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right)}\|p^k - x^{k-1}\|^2 \\
&\geq \frac{\|\eta^k - \hat{u}\|^2}{\gamma} + 2\big\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\big\rangle + \gamma\|L\|^2\|p^k - x^{k-1}\|^2 \\
&\geq \frac{1}{\gamma}\|\eta^k - \hat{u} + \gamma L(p^k - x^{k-1})\|^2 \geq 0,
\end{aligned}$$

and, hence, from (2.2.11) we deduce that $(\Upsilon_k + \|p^k - \hat{x}\|^2/\tau)_{k\in\mathbb{N}}$ is a Féjer sequence. We deduce from [6, Lemma 5.31] that $(\eta^k)_{k\in\mathbb{N}}$ and $(p^k)_{k\in\mathbb{N}}$ are bounded,

$$x^k - p^k \to 0, \quad u^k - \eta^k \to 0, \quad \eta^{k+1} - u^k \to 0, \quad \text{and} \quad p^k - x^{k-1} \to 0. \tag{2.2.12}$$

Therefore, there exist weak accumulation points $\bar{x}$ and $\bar{u}$ of the sequences $(p^k)_{k\in\mathbb{N}}$ and $(\eta^k)_{k\in\mathbb{N}}$, respectively, say $p^{k_n} \rightharpoonup \bar{x}$ and $\eta^{k_n} \rightharpoonup \bar{u}$ and, from (2.2.12), we have $u^{k_n} \rightharpoonup \bar{u}$, $u^{k_n+1} \rightharpoonup \bar{u}$, $p^{k_n} \rightharpoonup \bar{x}$, $p^{k_n+1} \rightharpoonup \bar{x}$, $x^{k_n-1} \rightharpoonup \bar{x}$ and $\bar{x}^{k_n} = x^{k_n} + p^{k_n} - x^{k_n-1} \rightharpoonup \bar{x}$. Since T and P_V are nonexpansive, $\mathrm{Id} - T$ and $\mathrm{Id} - P_V$ are maximally monotone [6, Example 20.29] and, therefore, they have weak-strong closed graphs [6, Proposition 20.38]. Hence, it follows from (2.2.12) that $(\mathrm{Id} - T)p^k \to 0$ and $(\mathrm{Id} - P_V)\eta^k \to 0$ and, hence, $(\bar{x}, \bar{u}) \in \operatorname{Fix} T \times V$. Moreover, (2.2.7) can be written equivalently as

$$(v^{k_n}, w^{k_n}) \in (\boldsymbol{M} + \boldsymbol{Q})(p^{k_n+1}, \eta^{k_n+1}),$$

where $\boldsymbol{M}\colon (p,\eta) \mapsto (Ap + L^*\eta) \times (B^{-1}\eta - Lp)$ is maximally monotone [14, Proposition 2.7(iii)], $\boldsymbol{Q}\colon (p,\eta) \mapsto (Cp, D^{-1}\eta)$ is $\min\{\beta,\delta\}-$cocoercive, and

$$\begin{cases} v^k := \frac{x^k - p^{k+1}}{\tau} - L^*(u^{k+1} - \eta^{k+1}) + Cp^{k+1} - Cx^k \\ w^k := \frac{u^k - \eta^{k+1}}{\gamma} + L(x^k - p^{k+1} + p^k - x^{k-1}) + D^{-1}\eta^{k+1} - D^{-1}u^k. \end{cases}$$

It follows from [6, Corollary 25.5] that $\boldsymbol{M} + \boldsymbol{Q}$ is maximally monotone and, since (2.2.12) and the uniform continuity of C, D and L yields $v^{k_n} \to 0$ and $w^{k_n} \to 0$, we deduce from the weak-strong closedness of the graph of $\boldsymbol{M} + \boldsymbol{Q}$ that $(\bar{x}, \bar{u})$ is a solution to Problem 2.2.1. Let $(\hat{x}, \hat{u})$ be a solution to Problem 2.2.1 and consider the Hilbert space $\mathcal{H} \oplus \mathcal{G}$ with the following inner product:

$$\begin{aligned}
&\langle\langle\cdot|\cdot\rangle\rangle\colon (\mathcal{H} \oplus \mathcal{G})^2 \to \mathbb{R} \\
&\qquad ((x,u),(x',u')) \mapsto \frac{1}{\tau}\langle x, x'\rangle + \frac{1}{\gamma}\langle u, u'\rangle,
\end{aligned} \tag{2.2.13}$$

and the succession defined by:

$$\begin{aligned}
(\forall k \in \mathbb{N}) \quad a_k := \Upsilon_k + \frac{\left\|p^k - \hat{x}\right\|^2}{\tau} &= \|(p^k, \eta^k) - (\hat{x}, \hat{u})\|^2_{\mathcal{H}\oplus\mathcal{G}} \\
&\quad + 2\big\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\big\rangle + \left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\|p^k - x^{k-1}\|^2.
\end{aligned}$$

By (2.2.11), it follows that (a_k) is decreasing and, therefore, it converges. Furthermore, by (2.2.12), the last two terms of (a_k) converge (to zero). Then, $\|(p^k, \eta^k) - (\hat{x}, \hat{u})\|^2_{\mathcal{H}\oplus\mathcal{G}}$ also converges. Thus, by the Opial's lemma [42], there exists $(\hat{x}, \hat{u})$ solution to Problem 2.2.1 such that $(p^k, \eta^k) \rightharpoonup (\hat{x}, \hat{u})$. Then, $x^k = x^k - p^k + p^k \rightharpoonup \hat{x}$ and $u^k = u^k - \eta^k + \eta^k \rightharpoonup \hat{u}$.

(iii): Fix $k \in \mathbb{N}$. Since $\rho > 0$, $\delta = +\infty$, $\chi = 0$, we obtain from Theorem 2.2.2((i))

$$\begin{aligned}\frac{\|x^k - \hat{x}\|^2}{\tau_k} + \frac{\|u^k - \hat{u}\|^2}{\gamma_k} &\geq (2\rho\tau_k + 1)\frac{\tau_{k+1}}{\tau_k}\frac{\|p^{k+1} - \hat{x}\|^2}{\tau_{k+1}} + \frac{\gamma_{k+1}}{\gamma_k}\frac{\|\eta^{k+1} - \hat{u}\|^2}{\gamma_{k+1}} \\ &+ 2\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\rangle - 2\theta_{k-1}\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\rangle \\ &+ \left\|p^{k+1} - x^k\right\|^2 \left(\frac{1}{\tau_k} - \frac{1}{2\beta}\right) - \theta_{k-1}^2\gamma_k\|L\|^2\|p^k - x^{k-1}\|^2, \end{aligned} \tag{2.2.14}$$

where we use $2ab \leq a^2/\gamma + \gamma b^2$. Moreover, it follows from (2.2.4) that

$$(\forall k \in \mathbb{N}) \quad (1 + 2\rho\tau_k)\frac{\tau_{k+1}}{\tau_k} = (1 + 2\rho\tau_k)\theta_k = \frac{1}{\theta_k} = \frac{\gamma_{k+1}}{\gamma_k},$$

which, combined with (2.2.14), yields

$$\begin{aligned}\frac{\|x^k - \hat{x}\|^2}{\tau_k} + \frac{\|u^k - \hat{u}\|^2}{\gamma_k} &\geq \frac{1}{\theta_k}\left(\frac{\|p^{k+1} - \hat{x}\|^2}{\tau_{k+1}} + \frac{\|\eta^{k+1} - \hat{u}\|^2}{\gamma_{k+1}}\right) \\ &+ 2\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\rangle - 2\theta_{k-1}\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\rangle \\ &+ \left\|p^{k+1} - x^k\right\|^2 \left(\frac{1}{\tau_k} - \frac{1}{2\beta}\right) - \theta_{k-1}^2\gamma_k\|L\|^2\|p^k - x^{k-1}\|^2. \end{aligned} \tag{2.2.15}$$

Now define

$$(\forall k \in \mathbb{N}) \quad \Delta_k = \frac{\left\|x^k - \hat{x}\right\|^2}{\tau_k} + \frac{\left\|u^k - \hat{u}\right\|^2}{\gamma_k}. \tag{2.2.16}$$

Dividing (2.2.15) by τ_k and using $\theta_k\tau_k = \tau_{k+1}$, we obtain from the nonexpansivity of P_V and T that

$$\begin{aligned}\frac{\Delta_k}{\tau_k} &\geq \frac{\Delta_{k+1}}{\tau_{k+1}} + \frac{2}{\tau_k}\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\rangle - \frac{2}{\tau_{k-1}}\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\rangle \\ &+ \frac{\left\|p^{k+1} - x^k\right\|^2}{\tau_k^2}\left(1 - \frac{\tau_k}{2\beta}\right) - \gamma_k\tau_k\|L\|^2\frac{\|p^k - x^{k-1}\|^2}{\tau_{k-1}^2}. \end{aligned} \tag{2.2.17}$$

In addition, (2.2.1) with equality reduces to

$$\|L\|^2 = \left(\frac{1}{\tau_0} - \frac{1}{2\beta}\right)\frac{1}{\gamma_0} \quad \Leftrightarrow \quad \gamma_0\tau_0\|L\|^2 = \left(1 - \frac{\tau_0}{2\beta}\right). \tag{2.2.18}$$

Since, for every $k \in \mathbb{N} \setminus \{0\}$, $\gamma_k\tau_k = \gamma_0\tau_0$ and $(\tau_k)_{k\in\mathbb{N}}$ is decreasing (see (2.2.4)), we have from (2.2.18) that

$$\gamma_k\tau_k\|L\|^2 = \gamma_0\tau_0\|L\|^2 = \left(1 - \frac{\tau_0}{2\beta}\right) \leq \left(1 - \frac{\tau_{k-1}}{2\beta}\right), \tag{2.2.19}$$

and (2.2.17) yields

$$\begin{aligned}\frac{\Delta_k}{\tau_k} &\geq \frac{\Delta_{k+1}}{\tau_{k+1}} + \frac{\left\|p^{k+1} - x^k\right\|^2}{\tau_k^2}\left(1 - \frac{\tau_k}{2\beta}\right) - \frac{\|p^k - x^{k-1}\|^2}{\tau_{k-1}^2}\left(1 - \frac{\tau_{k-1}}{2\beta}\right) \\ &+ \frac{2}{\tau_k}\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\rangle - \frac{2}{\tau_{k-1}}\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\rangle. \end{aligned} \tag{2.2.20}$$

Now fix $N \geq 1$. By adding from $k = 0$ to $k = N-1$ in (2.2.20), defining $p^0 := x^0$, $x^{-1} := x^0$, and $\tau_{-1} := \tau_0$, we obtain from $u^N = P_V \eta^N$, and $\operatorname{ran} L \subset V$ that

$$
\begin{aligned}
\frac{\Delta_0}{\tau_0} &\geq \frac{\Delta_N}{\tau_N} + \frac{\left\|p^N - x^{N-1}\right\|^2}{\tau_{N-1}^2}\left(1 - \frac{\tau_{N-1}}{2\beta}\right) + \frac{2}{\tau_{N-1}}\left\langle L(p^N - x^{N-1}) \mid u^N - \hat{u}\right\rangle \\
&\geq \frac{\Delta_N}{\tau_N} - \frac{\|L\|^2}{(1 - \frac{\tau_{N-1}}{2\beta})}\left\|u^N - \hat{u}\right\|^2 \\
&= \frac{1}{\tau_N}\left(\Delta_N - \frac{\gamma_N \tau_N \|L\|^2}{(1 - \frac{\tau_{N-1}}{2\beta})}\frac{\left\|u^N - \hat{u}\right\|^2}{\gamma_N}\right) \geq \frac{\|x^N - \hat{x}\|^2}{\tau_N^2},
\end{aligned}
\tag{2.2.21}
$$

where the last inequality follows from (2.2.19) and (2.2.16). Multiplying (2.2.21) by τ_N^2 and using (2.2.18), we conclude that

$$
\left\|x^N - \hat{x}\right\|^2 \leq \tau_N^2\left(\frac{\left\|x^0 - \hat{x}\right\|^2}{\tau_0^2} + \frac{\|L\|^2}{(1 - \frac{\tau_0}{2\beta})}\left\|u^0 - \hat{u}\right\|^2\right). \tag{2.2.22}
$$

The result follows from $\lim_{N\to\infty} N\rho\tau_N = 1$ [20, Corollary 1].

(IV): Fix $k \in \mathbb{N}$. Note that (2.2.6) yields $\left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right) = \|L\|^2$. From (2.2.9), $u^{k+1} = P_V \eta^{k+1}$ and $\operatorname{ran} L \subset V$, we have

$$
\begin{aligned}
\frac{\left\|u^k - \hat{u}\right\|^2}{2\gamma} + \frac{\left\|x^k - \hat{x}\right\|^2}{2\tau} &\geq (2\rho\tau + 1)\frac{\|p^{k+1} - \hat{x}\|^2}{2\tau} + (2\chi\gamma + 1)\frac{\|\eta^{k+1} - \hat{u}\|^2}{2\gamma} \\
&\quad + \frac{\left\|p^{k+1} - x^k\right\|^2}{2}\left(\frac{1}{\tau} - \frac{1}{2\beta}\right) + \frac{\|\eta^{k+1} - u^k\|^2}{2}\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right) \\
&\quad + \left\langle L(p^{k+1} - \overline{x}^k) \mid \eta^{k+1} - \hat{u}\right\rangle.
\end{aligned}
\tag{2.2.23}
$$

Hence, by defining

$$
(\forall k \in \mathbb{N}) \quad \Omega_k := \left(\chi + \frac{\mu}{4\delta}\right)\|u^k - \hat{u}\|^2 + \left(\rho + \frac{\mu}{4\beta}\right)\|x^k - \hat{x}\|^2, \tag{2.2.24}
$$

multiplying (2.2.23) by μ and using (2.2.5), (2.2.6), $u^{k+1} = P_V \eta^{k+1}$, $\operatorname{ran} L \subset V$, and the nonexpansivity of T and P_V, we have

$$
\begin{aligned}
\Omega_k &\geq \Omega_{k+1} + \mu\rho\left\|p^{k+1} - \hat{x}\right\|^2 + \mu\chi\left\|\eta^{k+1} - \hat{u}\right\|^2 + \rho\left\|p^{k+1} - x^k\right\|^2 \\
&\qquad + \chi\left\|\eta^{k+1} - u^k\right\|^2 + \mu\left\langle L(p^{k+1} - \overline{x}^k) \mid \eta^{k+1} - \hat{u}\right\rangle \\
&\geq (1 + \alpha)\Omega_{k+1} + \rho\left\|p^{k+1} - x^k\right\|^2 + \chi\left\|u^{k+1} - u^k\right\|^2 \\
&\qquad + \mu\left\langle L(p^{k+1} - \overline{x}^k) \mid u^{k+1} - \hat{u}\right\rangle.
\end{aligned}
\tag{2.2.25}
$$

Moreover, for every $\omega, \lambda > 0$ we have

$$\begin{aligned}
&\mu\big\langle L(p^{k+1} - \overline{x}^k) \mid u^{k+1} - \hat{u}\big\rangle \\
&\quad = \mu\big\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\big\rangle - \mu\theta\big\langle L(p^k - x^{k-1}) \mid u^{k+1} - \hat{u}\big\rangle \\
&\quad = \mu\big\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\big\rangle - \omega\mu\big\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\big\rangle \\
&\qquad - \omega\mu\big\langle L(p^k - x^{k-1}) \mid u^{k+1} - u^k\big\rangle \\
&\qquad - (\theta - \omega)\mu\big\langle L(p^k - x^{k-1}) \mid u^{k+1} - \hat{u}\big\rangle \\
&\quad \geq \mu\big\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\big\rangle - \omega\mu\big\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\big\rangle \\
&\qquad - \omega\mu\,\|L\|\left(\frac{\lambda\left\|p^k - x^{k-1}\right\|^2}{2} + \frac{\left\|u^{k+1} - u^k\right\|^2}{2\lambda}\right) \\
&\qquad - (\theta - \omega)\mu\,\|L\|\left(\frac{\lambda\left\|p^k - x^{k-1}\right\|^2}{2} + \frac{\left\|u^{k+1} - \hat{u}\right\|^2}{2\lambda}\right) \\
&\quad = \mu\big\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\big\rangle - \omega\mu\big\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\big\rangle \\
&\qquad - \mu\theta\lambda\,\|L\|\,\frac{\left\|p^k - x^{k-1}\right\|^2}{2} - \frac{\omega\mu\,\|L\|\,\left\|u^{k+1} - u^k\right\|^2}{2\lambda} \\
&\qquad - (\theta - \omega)\mu\,\|L\|\,\frac{\left\|u^{k+1} - \hat{u}\right\|^2}{2\lambda}.
\end{aligned} \tag{2.2.26}$$

By choosing $\lambda = \omega\sqrt{\dfrac{\rho}{\chi}}$, from (2.2.25), (2.2.26) and (2.2.5), we obtain

$$\begin{aligned}
\Omega_k \geq{}& \frac{\Omega_{k+1}}{\omega} + \left(1 + \alpha - \frac{1}{\omega}\right)\Omega_{k+1} + \rho\left\|p^{k+1} - x^k\right\|^2 \\
&+ \mu\big\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\big\rangle - \omega\mu\big\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\big\rangle \\
&- \omega\theta\rho\left\|p^k - x^{k-1}\right\|^2 - \left(\frac{\theta - \omega}{\omega}\right)\chi\left\|u^{k+1} - \hat{u}\right\|^2.
\end{aligned} \tag{2.2.27}$$

Since $\theta \in \left](1+\alpha)^{-1}, 1\right]$, by setting $\omega = \dfrac{1+\theta}{2+\alpha} \in \left](1+\alpha)^{-1}, \theta\right[$, we have $1 + \alpha - \dfrac{1}{\omega} = \dfrac{\theta - \omega}{\omega} > 0$. Hence, since (2.2.24) yields $\Omega_{k+1} \geq \chi\left\|u^{k+1} - \hat{u}\right\|^2$, from (2.2.27) and $\theta \leq 1$ we have

$$\begin{aligned}
\Omega_k \geq{}& \frac{\Omega_{k+1}}{\omega} + \rho\left\|p^{k+1} - x^k\right\|^2 - \omega\rho\left\|p^k - x^{k-1}\right\|^2 \\
&+ \mu\big\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\big\rangle - \omega\mu\big\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\big\rangle.
\end{aligned} \tag{2.2.28}$$

Moreover, using $p^0 = x^0 =: x^{-1}$, multiplying (2.2.28) by ω^{-k} and adding from $k = 0$ to $k = N-1$, we conclude from the definition of μ that

$$\begin{aligned}
\Omega_0 &\geq \omega^{-N}\Omega_N + \omega^{-N+1}\rho\left\|p^N - x^{N-1}\right\|^2 + \mu\omega^{-N+1}\big\langle L(p^N - x^{N-1}) \mid u^N - \hat{u}\big\rangle \\
&\geq \omega^{-N}\Omega_N + \omega^{-N+1}\rho\left\|p^N - x^{N-1}\right\|^2 \\
&\quad - \mu\omega^{-N+1}\,\|L\|\left(\sqrt{\frac{\rho}{\chi}}\frac{\left\|p^N - x^{N-1}\right\|^2}{2} + \sqrt{\frac{\chi}{\rho}}\frac{\left\|u^N - \hat{u}\right\|^2}{2}\right) \\
&= \omega^{-N}\Omega_N - \omega^{-N+1}\chi\left\|u^N - \hat{u}\right\|^2,
\end{aligned}$$

or, equivalently,

$$\omega^N \left(\left(\chi + \frac{\mu}{4\delta} \right) \left\| u^0 - \hat{u} \right\|^2 + \left(\rho + \frac{\mu}{4\beta} \right) \left\| x^0 - \hat{x} \right\|^2 \right)$$
$$\geq \left(\chi(1-\omega) + \frac{\mu}{4\delta} \right) \left\| u^N - \hat{u} \right\|^2 + \left(\rho + \frac{\mu}{4\beta} \right) \left\| x^N - \hat{x} \right\|^2,$$

which proves the linear convergence since $\omega < \theta \leq 1$. □

Remark 2.2.3 (I) Note that condition (2.2.1) implies that the whole sequence satisfies

$$(\forall k \in \mathbb{N}) \quad \|L\|^2 \leq \left(\frac{1}{\tau_k} - \frac{1}{2\beta} \right) \left(\frac{1}{\gamma_k} - \frac{1}{2\delta} \right)$$

under the additional assumptions on the sequences $(\gamma_k)_{k\in\mathbb{N}}$ and $(\tau_k)_{k\in\mathbb{N}}$ made in each part.

(II) In the case when $T = \mathrm{Id}$, $V = \mathcal{G}$, $A + C$ is strongly monotone, and C is monotone and β^{-1}–Lipschitz, different step-sizes are considered in [9]. The choice of step-sizes in [9] coincide when $C = 0$, but the sequence $(\theta_k)_{k\in\mathbb{N}}$ is strictly larger that our sequence for the same τ_0 when $C \neq 0$. Therefore, for the same initial $\tau_0 > 0$, our step-sizes $(\tau_k)_{k\in\mathbb{N}}$ are smaller than those in [9], leading to a better estimation when C is cocoercive.

(III) Condition (2.2.1) is weaker than the condition needed in [48]. Indeed, this condition in our case reads $2\rho \min\{\beta, \delta\} > 1$, where $\rho = \min\left\{\gamma^{-1}, \tau^{-1}\right\}(1 - \sqrt{\tau\gamma\|L\|^2})$, which implies $2\min\{\delta, \beta\} > \frac{1}{\rho} > \max\{\gamma, \tau\}$,

$$\left(1 - \frac{\tau}{2\beta}\right) > \sqrt{\tau\gamma\|L\|^2} \quad \text{and} \quad \left(1 - \frac{\gamma}{2\delta}\right) > \sqrt{\tau\gamma\|L\|^2}.$$

Thus, by multiplying last expressions we obtain

$$\left(1 - \frac{\tau}{2\beta}\right)\left(1 - \frac{\gamma}{2\delta}\right) > \tau\gamma\|L\|^2,$$

which implies (2.2.1). Our condition is strictly weaker, as it can be seen in Figure 2.1, in which we plot the case $\|L\| = 1$ and $\delta = \beta = b$, for $b = 1$, $b = 1/2$ and $b = 1/4$. That is, we compare regions

$$R_b = \left\{ (\tau, \gamma) \in [0, 2b] \times [0, 2b] \colon \min\left\{ \frac{1 - \sqrt{\tau\gamma}}{\tau}, \frac{1 - \sqrt{\tau\gamma}}{\gamma} \right\} > \frac{1}{2b} \right\}$$
$$S_b = \left\{ (\tau, \gamma) \in [0, 2b] \times [0, 2b] \colon \left(1 - \frac{\tau}{2b}\right)\left(1 - \frac{\gamma}{2b}\right) > \tau\gamma \right\}.$$

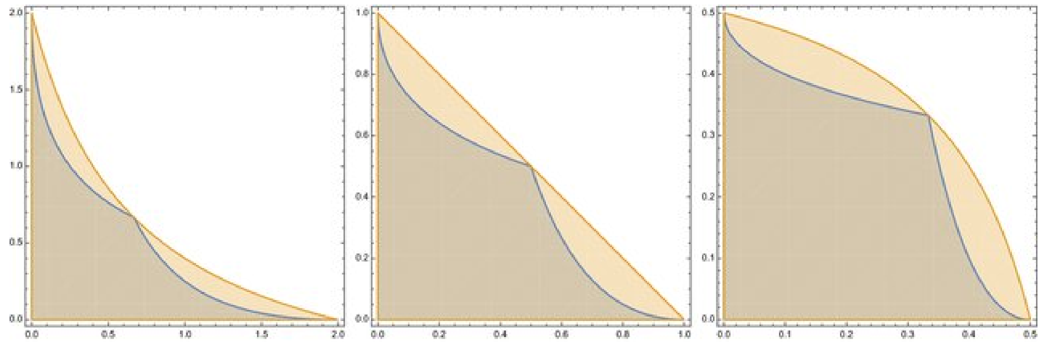

Figure 2.1: We plot regions R_b in blue and S_b in orange. Left: case $b = 1$, Center: case $b = 1/2$, Right: case $b = 1/4$. Note that in the case $\tau = \gamma$ the regions coincide.

(iv) It is not difficult to extend our method by replacing the averaged quasi-nonexpansive operator T by $(\alpha_k)_{k\in\mathbb{N}}-$averaged quasi-nonexpansive operators $(T_k)_{k\in\mathbb{N}}$ varying at each iteration and satisfying $\sup_{k\in\mathbb{N}} \alpha_k < 1$. Indeed, as in [25], we have to assume that $x^k - T_k x^k \to 0$ and $x^k \rightharpoonup x$ implies $x \in \cap_{k\in\mathbb{N}} \operatorname{Fix} T_k$, which is satisfied in several cases. In particular, if we set, for every $k \in \mathbb{N}$, $\gamma_k \in]0, 2\xi[$ and $T_k := J_{\gamma_k M}(\operatorname{Id} - \gamma_k N)$, where $M\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is maximally monotone and $N\colon \mathcal{H} \to \mathcal{H}$ is ξ-cocoercive, we have that T_k is $\gamma_k/2\xi-$cocoercive and $\cap_{k\in\mathbb{N}} \operatorname{Fix} T_k = \operatorname{zer}(M + N)$. Therefore, our method using these operators leads to the common solution to $\operatorname{zer}(M + N)$ and $\operatorname{zer}(A + L^* \circ B \circ L + C)$. Previous example can also be tackled by Theorem 2.2.2 if we use $\gamma_k \equiv \gamma$ and $T_k \equiv T := J_{\gamma M}(\operatorname{Id} - \gamma N)$. We prefer to keep the constant operator case for avoiding additional hypotheses and for the sake of simplicity.

(v) The method proposed in [10] is an accelerated version of the method proposed in [26] under the assumption that $A + C$ is strongly monotone. Of course, this weaker assumption can also be used in our context, but we prefer to keep the statement of Theorem 2.2.2 simpler.

(vi) Theorem 2.2.2((iii)) generalizes the acceleration scheme proposed in [20] to monotone inclusions with a priori information and we obtain an accelerated version of the methods in [48] in the strongly monotone case when $T = \operatorname{Id}$ and $V = \mathcal{G}$. These accelerated versions, as far as we know, have not been developed in the literature.

(vii) In the context of primal-dual problem $(\mathcal{P}_0)$-$(\mathcal{D}_0)$, (2.2.2) reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} \eta^{k+1} & = \operatorname{prox}_{\gamma_k g^*}(u^k + \gamma_k(L\bar{x}^k - \nabla \ell^*(u^k))) \\ u^{k+1} & = P_V\, \eta^{k+1} \\ p^{k+1} & = \operatorname{prox}_{\tau_k f}\left(x^k - \tau_k\left(L^* u^{k+1} + \nabla h(x^k)\right)\right) \\ x^{k+1} & = T\, p^{k+1} \\ \bar{x}^{k+1} & = x^{k+1} + \theta_k(p^{k+1} - x^k), \end{array} \right. \tag{2.2.29}$$

and our conditions on the parameters coincide with [21, 37]. Without strong convexity of f and g^*, we deduce from Theorem 2.2.2((ii)) the weak convergence of the sequences generated by (2.2.29), generalizing results in [18, 29, 37]. When f or g^* is strongly convex, Theorem 2.2.2((iii)) yields an accelerated and projected version of [29]. When $V = \mathcal{G}$ and $T = \operatorname{Id}$, this result complements the ergodic convergence rates obtained in [21] and generalizes [20]. When $\ell^* = 0$, $V = \mathcal{G}$, $T = \operatorname{Id}$, and f and g^* are strongly convex, Theorem 2.2.2((iv)) yields non-ergodic linear convergence of [29], complementing the ergodic linear convergence in [21]. The advantage of the algorithm (2.2.29) with respect to [20,29] is that primal-dual iterates of the former are forced to be in $X \times V$ when $T = P_X$. This feature leads to a faster algorithm in the context of constrained convex optimization, by choosing X to be some of the constraints. This can be observed in the particular instance developed in [18] and in Section 2.2.4, in which we provide some numerical simulations.

2.2.4. Application to Constrained Convex Optimization

In this section, we explore the advantages of the proposed method in constrained convex optimization.

Problem 2.2.4 Let $f \in \Gamma_0(\mathbb{R}^N)$, let R and S be $m \times N$ and $n \times N$ real matrices, respectively, and let $c \in \mathbb{R}^m$ and $d \in \mathbb{R}^n$. The problem is to

$$\min_{x \in \mathbb{R}^N} f(x) \quad \text{s.t.} \quad Rx = c \quad Sx = d, \tag{2.2.30}$$

under the assumption that solutions exist.

Note that (2.2.30) can be written equivalently as $\min_{x\in\mathbb{R}^N} f(x) + \iota_{\{b\}}(Lx)$, where $L\colon x \mapsto (Rx, Sx)$ and $b = (c, d) \in \mathbb{R}^{m+n}$. Assume $0 \in \operatorname{sri}(L(\operatorname{dom} f) - b)$. Note that, since $\operatorname{prox}_{\gamma\iota^*_{\{b\}}} = \operatorname{Id} - \gamma b$ [6, Proposition 24.8(ix)], the method proposed in [20, Algorithm 1] in this case reads: given $x^0 = \bar{x}^0 \in \mathcal{H}$ and $u^0 \in \mathcal{G}$,

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} u^{k+1} & = u^k + \gamma(L\bar{x}^k - b) \\ x^{k+1} & = \operatorname{prox}_{\tau f}(x^k - \tau L^* u^{k+1}) \\ \bar{x}^{k+1} & = 2x^{k+1} - x^k, \end{array} \right. \tag{2.2.31}$$

where $\gamma\tau\|L\|^2 < 1$. The constraint is imposed via the Lagrange multiplier update in the first step of (2.2.31). This implies that the primal sequence $\{x^k\}_{k\in\mathbb{N}}$ does not necessarily satisfy any of the constraints. For ensuring feasibility, we should project onto $L^{-1}b$ by considering the problem $\min_{x\in\mathbb{R}^N} f(x) + \iota_{L^{-1}b}(x)$. However, this is not always possible since, in several applications, the matrices involved are singular or very bad conditioned (see discussion in [17, 18]). If it is difficult to compute $P_{L^{-1}b}$ but we can project onto $R^{-1}c$, we can rewrite (2.2.30) as the problem of finding $\hat{x} \in \operatorname{Fix} P_{R^{-1}c} \cap \operatorname{argmin}_{x\in\mathbb{R}^N} f(x) + \iota_{\{b\}}(Lx)$, which is $(\mathcal{P}_0)$ when $T = P_{R^{-1}c}$, $h = 0$, $\ell = \iota_{\{0\}}$ and $g = \iota_{\{b\}}$. Next corollary follows from Theorem 2.2.2, (2.2.29) and $P_{R^{-1}c}\colon x \mapsto x - R^*(RR^*)^{-1}(Rx - c)$, when RR^* is invertible.

Corollary 2.2.5 *Let $\gamma > 0$ and $\tau > 0$ be such that $\gamma\tau\|L\|^2 < 1$ and let $(x^0, \bar{x}^0, u^0) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^{m+n}$ be such that $x^0 = \bar{x}^0$. Consider the routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} u^{k+1} & = u^k + \gamma(L\bar{x}^k - b) \\ p^{k+1} & = \operatorname{prox}_{\tau f}(x^k - \tau L^* u^{k+1}) \\ x^{k+1} & = p^{k+1} - R^*(RR^*)^{-1}(Rp^{k+1} - c) \\ \bar{x}^{k+1} & = x^{k+1} + p^{k+1} - x^k. \end{array} \right. \tag{2.2.32}$$

Then, there exist a solution $\hat{x}$ to Problem 2.2.4 and an associated multiplier $\hat{u}$ such that $x^k \to \hat{x}$ and $u^k \to \hat{u}$.

2.2.5. Numerical Experiences

In this section, we consider some particular instances of Problem 2.2.4. We consider the case when $f = \|\cdot\|_1 \in \Gamma_0(\mathbb{R}^N)$, $N = 1000$, $\tau = \frac{0.99}{\gamma\|L\|^2}$ and the relative error in (2.2.32) is $r_k =$

$\sqrt{\frac{\|u^{k+1}-u^k\|^2+\|x^{k+1}-x^k\|^2}{\|u^k\|^2+\|x^k\|^2}}$, for every $k \in \mathbb{N}$. We set $\gamma = 10^{-2}$ and $(x^0, \bar{x}^0, u^0) = (0,0,0) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^{m+n}$ and, in each test we show the average execution time and the number of average iterations of both methods, obtained by considering 20 random realizations of matrices R, S and vectors $c \in \mathbb{R}^m$ and $d \in \mathbb{R}^n$. The random generated matrices and vectors are obtained via the `rand` function of matlab. PCP and CP denote the algorithms (2.2.32) and (2.2.31), respectively.

Test 1. In Problem 2.2.4, Table 2.1 show the efficiency of CP and PCP for the case $m = 1$ and $n = 100$.

Table 2.1: Average time and number of iterations when $m = 1$ for obtaining $r_k < e$.

$m = 1$, $n = 100$	$e = 10^{-4}$		$e = 5 \cdot 10^{-5}$		$e = 10^{-5}$	
	iter	time (s)	iter	time (s)	iter	time (s)
PCP	9265	22.28	14570	37.02	46191	116.26
CP	9732	23.04	15718	39.21	50544	125.49
%improv.	4.8	3.3	7.3	5.6	8.6	7.4

We see that both algorithms are similar in terms of the execution time and the number of iterations, with a small advantage for the PCP algorithm. In addition, by decreasing the tolerance e, the percentage of improvement, computed as $100 \cdot (x_{\text{CP}} - x_{\text{PCP}})/x_{\text{CP}}$, slightly increases.

Test 2. In Problem 2.2.4, Table 2.2 show the efficiency of CP and PCP for the case $m = 10$ and $n = 100$.

Table 2.2: Average time and number of iterations when $m = 10$ for obtaining $r_k < e$.

$m = 10$, $n = 100$	$e = 10^{-4}$		$e = 5 \cdot 10^{-5}$		$e = 10^{-5}$	
	iter	time (s)	iter	time (s)	iter	time (s)
PCP	6865	18.65	10229	27.86	22855	65.05
CP	9280	23.72	16033	39.13	49526	129.78
%improv.	26.0	21.4	36.2	28.8	53.9	49.9

In this case, there are clear differences between both algorithms and, as before, PCP is more efficient as tolerance decreases. In fact, when tolerance is 10^{-5}, there is an improvement of approximately 50% with respect to the CP in the execution time and the number of iterations is less than a half.

Test 3. Finally, in Problem 2.2.4, Table 2.3 show the efficiency of CP and PCP for the case $m = 30$ and $n = 100$.

Table 2.3: Average time and number of iterations when $m = 30$ for obtaining $r_k < e$.

$m = 30$, $n = 100$	$e = 10^{-4}$		$e = 5 \cdot 10^{-5}$		$e = 10^{-5}$	
	iter	time (s)	iter	time (s)	iter	time (s)
PCP	5146	7.68	7143	10.67	13421	19.70
CP	9941	12.93	16438	21.37	50841	64.23
%improv.	48.2	40.6	56.5	50.1	73.6	69.3

We note that the improvement in execution times are considerably higher than in the previous cases. For example, in the case of $e = 10^{-4}$, the improvement increases by approximately 20% with respect to the case $m = 10$ and by approximately 40% in the case of $m = 1$. As in the previous cases, if we decrease the tolerance to 10^{-5}, PCP has better efficiency reaching almost 70% improvement with respect to CP. Table 2.4 summarizes the percentage of improvements for each test.

Table 2.4: Comparison of improvement of average iterations and average times.

% improv.	$m = 1$		$m = 10$		$m = 30$	
	iter	time (s)	iter	time (s)	iter	time (s)
$e = 10^{-4}$	4.8	3.3	26.0	21.4	48.2	40.6
$e = 5 \cdot 10^{-5}$	7.3	5.6	36.2	28.8	56.5	50.1
$e = 10^{-5}$	8.6	7.4	53.9	49.9	73.6	69.3

We observe a better relative performance of PCP with respect to CP for larger values of m. Note that, the larger is m, the larger is the proportion of constraints on which we project.

2.2.6. Conclusions

In this paper, we provide a projected primal–dual method for solving composite monotone inclusions with a priori information on solutions. We provide acceleration schemes in the presence of strong monotonicity, and we derive linear convergence in the fully strongly monotone case. The importance of the a priori information set is illustrated via a numerical example in convex optimization with equality constraints, in which the proposed method outperforms [20].

Chapter 3

Monotone inclusions with a priori information and vector subspaces

3.1. Summary

In this second part, we consider the Problem 1.2.1 in the general case. The first important result obtained is the following characterization of the solutions to Problem 1.2.1, using the partial inverse technique introduced in [46].

Proposition 3.1.1 *Let $\tau > 0$ and consider the inclusion*

$$\text{find } (z,v) \in \operatorname{Fix}(P_V T P_V + P_{V^\perp}) \times W \text{ s.t. } \begin{cases} -P_V L^* v \in (\tau A)_V z + \tau P_V C P_V z \\ L P_V z \in (\tau B)^{-1} v + (\tau D)^{-1} v. \end{cases} \tag{$\mathcal{I}_\tau$}$$

Then the following assertions hold.

(I) *If $(x,u) \in \mathcal{H} \times \mathcal{G}$ is a solution to ($\mathcal{P}$)-($\mathcal{D}$), then*

$$(\exists y \in V^\perp) \quad \big(x + \tau(y - P_{V^\perp}(L^*u + Cx)), \tau u\big) \quad \text{is a solution to } (\mathcal{I}_\tau).$$

(II) *The set of solutions to ($\mathcal{I}_\tau$) is nonempty.*

(III) *If $(z,v) \in \mathcal{H} \times \mathcal{G}$ is a solution to ($\mathcal{I}_\tau$), then $(P_V z, v/\tau)$ is a solution to ($\mathcal{P}$)-($\mathcal{D}$).*

The relevance of the previous result is the fact that the solutions of the coupled inclusion ($\mathcal{P}$)-($\mathcal{D}$) are obtained explicitly from the solutions of an auxiliary inclusion, which does not contain the normal cone of V and admits solutions since we assume the existence of solutions to ($\mathcal{P}$)-($\mathcal{D}$).

The second result provides a weak convergent algorithm for solving Problem 1.2.1.

Theorem 3.1.2 *In the context of Problem 1.2.1, let $x^0 \in V$, let $\overline{x}^0 = x^0$, let $y^0 \in V^\perp$, let $u^0 \in \mathcal{G}$, let $\tau \in \left]0, 2\beta\right[$, and let $\gamma \in \left]0, 2\delta\right[$ be such that*

$$\|L\|^2 < \left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right). \tag{3.1.1}$$

Consider the following routine:

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l}
\eta^{k+1} = J_{\gamma B^{-1}}(u^k + \gamma(L\overline{x}^k - D^{-1}u^k)) \\
u^{k+1} = P_W \eta^{k+1} \\
\widetilde{z}^k = x^k + \tau y^k - \tau P_V(L^* u^{k+1} + Cx^k) \\
w^{k+1} = J_{\tau A} \widetilde{z}^k \\
r^{k+1} = P_V w^{k+1} \\
x^{k+1} = P_V T r^{k+1} \\
y^{k+1} = y^k + (r^{k+1} - w^{k+1})/\tau \\
\overline{x}^{k+1} = x^{k+1} + r^{k+1} - x^k.
\end{array} \right. \tag{3.1.2}
$$

Then there exists $(\widehat{x}, \widehat{u}) \in (V \cap \operatorname{Fix} T) \times W$ *solution to* $(\mathcal{P})$-$(\mathcal{D})$ *such that* $x^k \rightharpoonup \widehat{x}$ *and* $u^k \rightharpoonup \widehat{u}$.

The algorithm (3.1.2) is inspired in the algorithm (2.1.4), which coincide with (3.1.2) when $V = \mathcal{H}$ and there is no presence of strong monotony ($\rho = \chi = 0$). Moreover, in the case without a priori information ($T = \operatorname{Id}$), we generalize the algorithm in [12] by taking $B = D^{-1} = 0$, which generalize the classical method of partial inverses [46].

When T is weak-weak continuous, we also deduce that $Tr^k \rightharpoonup \widehat{x}$. For example, if $\mathcal{H}$ is finite dimensional and $T = P_C$ where C is a nonempty convex closed set. In this case, we obtain a sequence that belongs to the subset of constraints C and that converges to a solution to Problem 1.2.1. This fact produces a faster convergence as will be shown later in the numerical applications.

We apply the proposed algorithm in two convex optimization problems. The first is the constrained LASSO [34], in which we see that our algorithm without a priori information improve the efficiency of the algorithms in [12, 48]. This is explained by the fact that our algorithm exploit the vector subspace structure of the problem. The results show that if the vector subspace is larger, then the percentage of improvement in execution time increase. The second problem is the constrained ℓ^1 minimization, where we use our algorithm with a priori information. The results show that by increasing the number of affine linear constraints considered as a priori information, the percentage of improvement with respect to the Chambolle-Pock's method increases, following a behavior similar to the algorithm (2.1.4).

Finally, we use the proposed algorithm for solving the arc-capacity expansion problem in transport networks [22]. This problem consists in find the optimal investment decision in arc capacity and network flow operation under an uncertain environment, in order to minimize the operational cost along with the cost of expanding the arc capacity by imposing a non-anticipativity constraint in the expansion capacity vector, a demand constraint on the route flow vector, and a constraint associated with the arc capacity limit.

We compare our vector subspace formulation with respect to the formulation that does not use vector subspaces in two particular networks. The percentage of improvement in computational time when we use proper vector subspaces is up to $26,75\%$, which occurs when we consider the larger network and when there are more scenarios for the problem.

3.2. Article: "Primal-Dual Partial Inverse Splitting for Constrained Monotone Inclusions"

Abstract

In this work we study a constrained monotone inclusion involving the normal cone to a closed vector subspace and a priori information on primal solutions. We model this information by imposing that solutions belongs to the fixed point set of an averaged nonexpansive mapping. We characterize the solutions using an auxiliary inclusion that involves the partial inverse operator. Then, we propose the primal-dual partial inverse splitting and we prove its weak convergence to a solution of the inclusion, generalizing several methods in the literature. The efficiency of the proposed method is illustrated in two non-smooth convex optimization problems whose constraints have vector subspace structure. Finally, the proposed algorithm is applied to find a solution to a stochastic arc capacity expansion problem in transport networks.

3.2.1. Introduction

In this paper we propose a convergent algorithm for solving composite monotone inclusions involving a normal cone to a closed vector subspace and a priori information on the solutions. Monotone inclusions model several applications such as evolution inclusions [3, 28, 43], variational inequalities [6, 32], partial differential equations (PDEs) [2, 33, 40], and various optimization problems. In particular, when the monotone operators are subdifferentials of convex functions, the inclusion we study reduces to an optimization problem with a vector subspace constraint and a priori information. This class of problems appears in PDEs [41, Section 3], signal and image processing [4, 23, 30], and stochastic traffic theory [22, 45], among other fields. In the aforementioned applications, the vector subspace constraint models intrinsic properties of the solution, as regularity in PDEs and image processing, or non-anticipativity in stochastic problems. In turn, the a priori information can be used to reinforce feasibility in the iterates, resulting in more efficient algorithms, as explored in [19].

Our problem, in the particular case when the vector subspace is the whole space, can be solved by the algorithm in [19]. This method uses the a priori information to improve the efficiency, generalizing the algorithm in [48] for monotone inclusions and in [29] for convex optimization.

In addition, when no a priori information is considered, the methods proposed in [12, 46] solve particular instances of our problem using the partial inverse introduced in [46]. This mathematical tool exploits the vector subspace structure of the inclusion and has been used, for example, in [1, 13]. Our problem in the more general context without a priori information can be solved by algorithms in [14, 26, 48], using product space techniques without special consideration on the vector subspace structure. The product space formulation generates methods that include updates of high dimensional dual variables at each iteration, which reduce their performance.

The objective of this paper is to provide an algorithm for solving the inclusion under study in its full generality, by taking advantage of the vector subspace structure and the a priori information of the inclusion. Our method is obtained from the combination of the algorithm in [19] with partial inverse techniques developed in [12, 13, 46]. We illustrate the advantages of the partial inverse approach and the use of the a priori information by means of numerical experiences on constrained convex optimization. In this context, the a priori information is modeled by a set formed by some of the constraints of the problem and the additional projections in our method improve the speed of the convergence with respect to existing methods.

The paper is organized as follows. In Section 3.2.2, we set our notation and preliminaries. In Section 3.2.3, we formulate our problem, we characterize the solutions by using the partial inverse operator, we prove the weak convergence of our algorithm, and we discuss connections with existing methods in the literature. In Section 3.2.4, we implement the proposed algorithm in the context of constrained convex optimization. In particular, in Section 3.2.4 we compare the performance of our method with classical algorithms used in the field for the constrained LASSO problem and in Section 3.2.4 for the constrained ℓ^1 minimization problem. Moreover, we apply our method to the engineering application of stochastic arc capacity expansion in a transport network in Section 3.2.4 [22]. Finally, we provide our conclusions and perspectives in Section 3.2.5.

3.2.2. Notation and Background

Let $\mathcal{H}$ be a real Hilbert space. We denote the scalar product of $\mathcal{H}$ by $\langle \cdot | \cdot \rangle$ and its norm associated by $\|\cdot\|$. The class of bounded linear operators from $\mathcal{H}$ to a real Hilbert space $\mathcal{G}$ is denoted by $\mathcal{L}(\mathcal{H},\mathcal{G})$ and if $\mathcal{H}=\mathcal{G}$ this class is denoted by $\mathcal{L}(\mathcal{H})$. Given $L \in \mathcal{L}(\mathcal{H},\mathcal{G})$, its adjoint operator is denoted by $L^* \in \mathcal{L}(\mathcal{G},\mathcal{H})$. The projection operator onto a nonempty closed convex set $C \subset \mathcal{H}$ is denoted by P_C and the normal cone to C is denoted by N_C. Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator. We denote by $\operatorname{gra} A$ its graph, by A^{-1} its inverse operator, and by $J_A := (\operatorname{Id} + A)^{-1}$ its resolvent, where Id is the identity operator on $\mathcal{H}$. Moreover, A is ρ-strongly monotone for $\rho \geq 0$ iff, for every (x,u) and (y,v) in $\operatorname{gra} A$, we have $\langle x-y \mid u-v\rangle \geq \rho\,\|x-y\|^2$, it is monotone iff it is $\rho-$strongly monotone with $\rho=0$, and it is maximally monotone iff there is no exists a monotone operator B such that $\operatorname{gra} A \subsetneq \operatorname{gra} B$. Let V be a closed vector subspace of $\mathcal{H}$. The partial inverse of A with respect to V , denoted by A_V, is the operator whose graph is $\operatorname{gra} A_V := \{(x,u) \in \mathcal{H}\times\mathcal{H} : (P_V x + P_{V^\perp} u, P_V u + P_{V^\perp} x) \in \operatorname{gra} A\}$ [46]. Let $T\colon \mathcal{H} \to \mathcal{H}$. The set of fixed points of T is denoted by $\operatorname{Fix} T$. The operator T is α-averaged for some $\alpha \in \,]0,1[$ iff, for every $(x,y) \in \mathcal{H}^2$, we have $\|Tx-Ty\|^2 \leq \|x-y\|^2 - \left(\frac{1-\alpha}{\alpha}\right)\|(\operatorname{Id}-T)x - (\operatorname{Id}-T)y\|^2$ and it is $\beta-$cocoercive for some $\beta>0$ iff, for every $(x,y)\in\mathcal{H}^2$, we have $\langle x-y \mid Tx-Ty\rangle \geq \beta\|Tx-Ty\|^2$. The class of lower semicontinuous convex proper functions $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ is denoted by $\Gamma_0(\mathcal{H})$. The subdifferential of $f \in \Gamma_0(\mathcal{H})$ is denoted by ∂f and for every $\lambda > 0$, the proximity operator on x, denoted by $\operatorname{prox}_{\lambda f} x$, is the unique minimizer of $\lambda f + \|\cdot - x\|^2/2$. For further information on convex analysis and monotone operator theory, the reader is referred to [6].

3.2.3. Problem and Results

We consider the following composite primal-dual inclusion problem with a priori information.

Problem 3.2.1 Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces and let $V \subset \mathcal{H}$ and $W \subset \mathcal{G}$ be closed vector subspaces. Let $T\colon \mathcal{H} \to \mathcal{H}$ be an $\alpha-$averaged nonexpansive operator for some $\alpha \in \,]0,1[$, let $L \in \mathcal{L}(\mathcal{H},\mathcal{G})$ be such that $\operatorname{ran} L \subset W$, let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$, $B\colon \mathcal{G} \rightrightarrows \mathcal{G}$, and $D\colon \mathcal{G} \rightrightarrows \mathcal{G}$ be maximally monotone operators, let $C\colon \mathcal{H} \to \mathcal{H}$ be a $\beta-$cocoercive operator, and suppose that D is $\delta-$strongly monotone for some $(\delta,\beta) \in \,]0,+\infty]^2$. The problem is to

$$\text{find} \quad (x,u) \in \operatorname{Fix} T \times W \quad \text{such that} \quad \begin{cases} -L^*u \in Ax + Cx + N_V x \\ Lx \in B^{-1}u + D^{-1}u, \end{cases} \tag{$\mathcal{I}$}$$

under the assumption that ($\mathcal{I}$) admits solutions.

Consider the case when $W=\mathcal{G}$, $A=\partial F$, $B=\partial G$, $C=\nabla H$, and $D=\partial\ell$, where $F\in\Gamma_0(\mathcal{H})$, $G\in\Gamma_0(\mathcal{G})$, $H\colon\mathcal{H}\to\mathbb{R}$ is a differentiable convex function with $\beta^{-1}-$Lipschitz gradient, and $\ell\in\Gamma_0(\mathcal{G})$ is $\delta-$strongly convex. By defining $G\,\square\,\ell$ as the infimal convolution of G and ℓ, it follows from $\partial(G \,\square\, \ell) = ((\partial G)^{-1} + (\partial \ell)^{-1})^{-1}$ [6, Proposition 15.7(i) & Proposition 25.32] that Problem 3.2.1 reduces to the constrained optimization problem

$$\text{find } x \in \operatorname{Fix} T \cap \underset{x \in V}{\operatorname{argmin}}\, F(x) + (G \,\square\, \ell)(Lx) + H(x), \tag{3.2.1}$$

under standard qualification conditions. Note that from [6, Corollary 18.17], ∇H is $\beta-$cocoercive. In the case when $T = P_C$ for some nonempty closed convex set C, (3.2.1) models optimization problems with a priori information on the solution [19]. Moreover, (3.2.1) models the problem of finding a common solution to two convex optimization problems when T is such that $\operatorname{Fix} T = \arg\min \Phi$, for some convex function Φ (e.g., if $T = \operatorname{prox}_\Phi$, $\operatorname{Fix} T = \arg\min \Phi$). Analogously, we can incorporate a priori information on solutions and we can find common solutions to two problems in the context of monotone inclusions.

In the particular case when $V = \mathcal{H}$, $W = \mathcal{G}$, and $T = \operatorname{Id}$, [48] solves Problem 3.2.1, and the corresponding optimization problem can be solved by the method proposed in [29]. Previous methods are generalizations of several classical splitting algorithms in particular instances, such as the proximal-point algorithm [39, 44], the forward-backward splitting [36], and the Chambolle-Pock's algorithm [20]. In [19], previous methods are generalized for solving Problem 3.2.1 in the case when $W \subset \mathcal{G}$ and T is a general averaged nonexpansive operator.

In the case when $V \subset \mathcal{H}$, the algorithms proposed in [12, 46] solve Problem 3.2.1 in particular instances, exploiting the vector subspace structure by using the partial inverse of a monotone operator. A convergent splitting method generalizing the partial inverse algorithm in [46] is proposed in [12] and solves Problem 3.2.1 when $B = 0$. The algorithms proposed in [14, 26, 48] solves Problem 3.2.1 when $W = \mathcal{G}$ and $T = \operatorname{Id}$, using product space techniques without special consideration on the vector subspace structure of N_V. The resulting method involves higher dimensional dual variables to be updated at each iteration, affecting the performance of the algorithm.

In this section we provide our algorithm to solve Problem 3.2.1 in its full generality. Our method exploits the vector subspace structure and the a priori information of the inclusion. We first characterize the solutions to Problem 3.2.1 as solutions to an auxiliary monotone inclusion and then, we obtain our algorithm by applying the method in [19] to that inclusion, guaranteeing its convergence.

Characterization of solutions

In this section we prove that Problem 3.2.1 is equivalent to an auxiliary constrained monotone inclusion that does not include the normal cone to the closed vector subspace V. In order to construct this inclusion, we first identify the a priori information $\operatorname{Fix} T$ with the fixed point set of a suitable averaged nonexpansive operator.

Proposition 3.2.2 *Let $T\colon \mathcal{H} \to \mathcal{H}$ be an $\alpha-$averaged nonexpansive operator, for some $\alpha \in \,]0, 1[$, and define*

$$M := P_V \circ T \circ P_V + P_{V^\perp}. \tag{3.2.2}$$

Then, the following holds.

(I) *M is $\alpha-$averaged.*

(II) $\operatorname{Fix} M = P_V^{-1}(V \cap \operatorname{Fix} T) = (V \cap \operatorname{Fix} T) + V^\perp$.

Proof. (I): Let $(x,z) \in \mathcal{H}^2$. Since T is $\alpha-$averaged, P_V is linear and nonexpansive, and $\operatorname{Id} - M = P_V - P_V \circ T \circ P_V = P_V \circ (\operatorname{Id} - T) \circ P_V$, we have

$$\begin{aligned}\|Mx - Mz\|^2 &= \|P_V T P_V x + P_{V^\perp} x - P_V T P_V z - P_{V^\perp} z\|^2 \\ &= \|P_V T P_V x - P_V T P_V z\|^2 + \|P_{V^\perp} x - P_{V^\perp} z\|^2 \\ &\leq \|T P_V x - T P_V z\|^2 + \|P_{V^\perp} x - P_{V^\perp} z\|^2 \\ &\leq \|P_V x - P_V z\|^2 - \left(\frac{1-\alpha}{\alpha}\right) \|(\operatorname{Id} - T) P_V x - (\operatorname{Id} - T) P_V z\|^2 \\ &\quad + \|P_{V^\perp} x - P_{V^\perp} z\|^2 \\ &\leq \|x - z\|^2 - \left(\frac{1-\alpha}{\alpha}\right) \|P_V(\operatorname{Id} - T) P_V x - P_V(\operatorname{Id} - T) P_V x\|^2 \\ &= \|x - z\|^2 - \left(\frac{1-\alpha}{\alpha}\right) \|(\operatorname{Id} - M)x - (\operatorname{Id} - M)z\|^2.\end{aligned}$$

(II): Since Problem 3.2.1 has solutions, $\operatorname{Fix} P_V \cap \operatorname{Fix} T = V \cap \operatorname{Fix} T \neq \varnothing$. Then, by [6, Proposition 4.49 (i)], it follows that $\operatorname{Fix}(P_V T) = V \cap \operatorname{Fix} T$. Therefore, for every $z \in \mathcal{H}$ we have

$$\begin{aligned} z \in \operatorname{Fix} M &\iff z = P_V T P_V z + z - P_V z \\ &\iff P_V z \in \operatorname{Fix}(P_V T) = (V \cap \operatorname{Fix} T)\end{aligned}$$

and the result follows. □

The following proposition characterizes the solutions of Problem 3.2.1.

Proposition 3.2.3 *In the context of Problem 3.2.1, let M be the operator defined in (3.2.2), let $\tau > 0$, and consider the inclusion*

$$\text{find } (z,v) \in \operatorname{Fix} M \times W \text{ s.t. } \begin{cases} -P_V L^* v \in (\tau A)_V z + \tau P_V C P_V z \\ L P_V z \in (\tau B)^{-1} v + (\tau D)^{-1} v. \end{cases} \qquad (\mathcal{I}_\tau)$$

The following assertions hold.

(I) *If $(x,u) \in \mathcal{H} \times \mathcal{G}$ is a solution to $(\mathcal{I})$, then*

$$(\exists y \in V^\perp) \quad \big(x + \tau(y - P_{V^\perp}(L^* u + Cx)), \tau u\big) \quad \text{is a solution to } (\mathcal{I}_\tau).$$

(II) *The set of solutions to $(\mathcal{I}_\tau)$ is nonempty.*

(III) *If $(z,v) \in \mathcal{H} \times \mathcal{G}$ is a solution to $(\mathcal{I}_\tau)$, then $(P_V z, v/\tau)$ is a solution to $(\mathcal{I})$.*

Proof. (I): Let $(x,u) \in \mathcal{H} \times \mathcal{G}$. Note that

$$\begin{aligned} B^{-1} + D^{-1} &= (B^{-1} + D^{-1}) \circ \big(\tau^{-1} \operatorname{Id}\big) \circ (\tau \operatorname{Id}) \\ &= \big(B^{-1} \circ \big(\tau^{-1} \operatorname{Id}\big) + D^{-1} \circ \big(\tau^{-1} \operatorname{Id}\big)\big) \circ (\tau \operatorname{Id}) \\ &= \big((\tau B)^{-1} + (\tau D)^{-1}\big) \circ (\tau \operatorname{Id}).\end{aligned}$$

Therefore, (x,u) is a solution to $(\mathcal{I})$ if and only if $(x,u) \in (V \cap \operatorname{Fix} T) \times W$ and there exists $y \in N_V x = V^\perp$ such that

$$\begin{cases} y - L^*u - Cx \in Ax \\ Lx \in B^{-1}u + D^{-1}u \end{cases} \Leftrightarrow \begin{cases} \tau(y - L^*u - Cx) \in (\tau A)x \\ Lx \in (\tau B)^{-1}(\tau u) + (\tau D)^{-1}(\tau u) \end{cases} \tag{3.2.3}$$

$$\Leftrightarrow \begin{cases} -\tau P_V(L^*u + Cx) \in (\tau A)_V z \\ Lx \in (\tau B)^{-1}(\tau u) + (\tau D)^{-1}(\tau u) \end{cases}$$

$$\Leftrightarrow \begin{cases} -P_V L^* v \in (\tau A)_V z + \tau P_V C P_V z \\ L P_V z \in (\tau B)^{-1} v + (\tau D)^{-1} v, \end{cases} \tag{3.2.4}$$

where $z := x + \tau(y - P_{V^\perp}(L^*u + Cx))$ and $v := \tau u \in W$. In addition, since $P_V z = x \in \operatorname{Fix} T \cap V$, Proposition 3.2.2((ii)) yields $z \in \operatorname{Fix} M$.

(ii): It is clear from the assumptions in Problem 3.2.1 and (i).

(iii): Suppose that $(z,v) \in \operatorname{Fix} M \times W$ is a solution to $(\mathcal{I}_\tau)$. Then, Proposition 3.2.2((ii)) yields $(P_V z, v/\tau) \in (V \cap \operatorname{Fix} T) \times W$. Moreover, by setting $x := P_V z$, $u := v/\tau$, and $y := P_{V^\perp}(\frac{1}{\tau} z + L^*u + CP_V z) \in V^\perp$, (3.2.4) yields (3.2.3) and the result follows. □

Algorithm and convergence

In this section we propose the primal-dual partial inverse algorithm for solving Problem 3.2.1, which is derived from [19, Theorem 3.1] applied to $(\mathcal{I}_\tau)$.

Theorem 3.2.4 *In the context of Problem 3.2.1, let $x^0 \in V$, let $\overline{x}^0 = x^0$, let $y^0 \in V^\perp$, let $u^0 \in \mathcal{G}$, let $\tau \in \left]0, 2\beta\right[$, and let $\gamma \in \left]0, 2\delta\right[$ be such that*

$$\|L\|^2 < \left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right). \tag{3.2.5}$$

Consider the following routine.

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \eta^{k+1} = J_{\gamma B^{-1}}\big(u^k + \gamma(L\overline{x}^k - D^{-1}u^k)\big) \\ u^{k+1} = P_W \eta^{k+1} \\ \widetilde{z}^k = x^k + \tau y^k - \tau P_V(L^*u^{k+1} + Cx^k) \\ w^{k+1} = J_{\tau A}\widetilde{z}^k \\ r^{k+1} = P_V w^{k+1} \\ x^{k+1} = P_V T r^{k+1} \\ y^{k+1} = y^k + (r^{k+1} - w^{k+1})/\tau \\ \overline{x}^{k+1} = x^{k+1} + r^{k+1} - x^k. \end{array} \right. \tag{3.2.6}$$

Then there exists a solution $(\widehat{x}, \widehat{u})$ to Problem 3.2.1 such that $(x^k, u^k) \rightharpoonup (\widehat{x}, \widehat{u})$.

Proof. For every $k \in \mathbb{N}$, define $z^k := x^k + \tau y^k$, $p^k := r^k + \tau y^k$, $\overline{z}^{k+1} := z^{k+1} + p^{k+1} - z^k$, and $\overline{z}^0 := z^0$. Note that $\{r^k\}_{k\in\mathbb{N}} \subset V$, $\{x^k\}_{k\in\mathbb{N}} \subset V$, and $\{y^k\}_{k\in\mathbb{N}} \subset V^\perp$. Hence, since for every $k \in \mathbb{N}$, $r^k = P_V p^k$ and $x^k = P_V z^k$, it follows from (3.2.6) that

$$(\forall k \in \mathbb{N}) \left\{ \begin{array}{l} z^{k+1} = P_V T r^{k+1} + \tau y^{k+1} = P_V T P_V p^{k+1} + P_{V^\perp} p^{k+1} = M p^{k+1} \\ P_V \overline{z}^{k+1} = x^{k+1} + r^{k+1} - x^k = \overline{x}^{k+1} \end{array} \right.$$

In addition, $P_V\overline{z}^0 = P_V z^0 = x^0 = \overline{x}^0$ from (3.2.6) and [13, Proposition 3.1(i)] we deduce

$$\begin{aligned} p^{k+1} &= r^{k+1} + \tau y^{k+1} \\ &= 2r^{k+1} - w^{k+1} + \tau y^k \\ &= (2P_V - \mathrm{Id}\,)J_{\tau A}\widetilde{z}^k + P_{V^\perp}\widetilde{z}^k \\ &= J_{(\tau A)_V}\widetilde{z}^k. \end{aligned} \tag{3.2.7}$$

Note that

$$\begin{aligned} \tau J_{\gamma B^{-1}} &= \tau(\mathrm{Id}\, + \gamma B^{-1})^{-1} \\ &= \big((\mathrm{Id}\, + \gamma B^{-1}) \circ (\tau^{-1}\,\mathrm{Id}\,)\big)^{-1} \\ &= \big(\tau^{-1}\,\mathrm{Id}\, + \gamma(\tau B)^{-1}\big)^{-1} \\ &= \big(\tau^{-1}(\mathrm{Id}\, + \tau\gamma(\tau B)^{-1})\big)^{-1} \\ &= J_{\tau\gamma(\tau B)^{-1}} \circ (\tau\,\mathrm{Id}\,). \end{aligned}$$

Define $\widetilde{A} = (\tau A)_V$, $\widetilde{B} = \tau B$, $\widetilde{C} = \tau P_V C P_V$, $\widetilde{D} = \tau D$, $\widetilde{L} = LP_V$, $\sigma = \tau\gamma \in \,]0, 2\tau\delta[$, and for every $k \in \mathbb{N}$, $v^k = \tau u^k$ and $\zeta^k = \tau\eta^k$. Thus, from $P_V^* = P_V$ and the linearity of P_V and P_W, we deduce that (3.2.6) reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \zeta^{k+1} = J_{\sigma\widetilde{B}^{-1}}(v^k + \sigma(\widetilde{L}\overline{z}^k - \widetilde{D}^{-1}v^k)) \\ v^{k+1} = P_W\zeta^{k+1} \\ p^{k+1} = J_{\widetilde{A}}(z^k - (\widetilde{L}^*v^{k+1} + \widetilde{C}z^k)) \\ z^{k+1} = Mp^{k+1} \\ \overline{z}^{k+1} = z^{k+1} + p^{k+1} - z^k. \end{array} \right. \tag{3.2.8}$$

In addition, we have that $\widetilde{A}$ and $\widetilde{B}$ are maximally monotone operators [6, Proposition 20.44(v) & Proposition 20.22], $\widetilde{D}$ is $\tau\delta$-strongly monotone, $\widetilde{L}$ is a bounded linear operator such that $\operatorname{ran}\widetilde{L} \subset \operatorname{ran} L \subset W$, and by Proposition 3.2.2((I)), M is $\alpha-$averaged. Moreover, since P_V is nonexpansive and $P_V^* = P_V$, for every $(x, y) \in \mathcal{H} \times \mathcal{H}$ we have

$$\begin{aligned} \left\langle \widetilde{C}x - \widetilde{C}y \mid x - y \right\rangle &= \tau\langle CP_Vx - CP_Vy \mid P_Vx - P_Vy\rangle \\ &\geq \tau\beta\|CP_Vx - CP_Vy\|^2 \\ &\geq \tau\beta\|P_VCP_Vx - P_VCP_Vy\|^2 \\ &= \frac{\beta}{\tau}\|\widetilde{C}x - \widetilde{C}y\|^2, \end{aligned}$$

which implies that $\widetilde{C}$ is $\dfrac{\beta}{\tau}-$cocoercive. Furthermore, from (3.2.5) we obtain

$$\|\widetilde{L}\|^2 \leq \|L\|^2 < \left(1 - \frac{1}{2\frac{\beta}{\tau}}\right)\left(\frac{1}{\sigma} - \frac{1}{2\tau\delta}\right) \tag{3.2.9}$$

and $1 \in \,]0, 2\beta/\tau[$. Altogether, since Proposition 3.2.3((II)) yields the existence of solutions to

$$\text{find} \quad (z, v) \in \operatorname{Fix} M \times W \quad \text{such that} \quad \left\{ \begin{array}{l} -\widetilde{L}^*v \in \widetilde{A}z + \widetilde{C}z \\ \widetilde{L}z \in \widetilde{B}^{-1}v + \widetilde{D}^{-1}v, \end{array} \right. \tag{3.2.10}$$

it follows from [19, Theorem 3.1(ii)] that there exists $(\widehat{z}, \widehat{v}) \in \operatorname{Fix} M \times W$ solution to (3.2.10) such that $(z^k, v^k) \rightharpoonup (\widehat{z}, \widehat{v})$. Therefore, $u^k = v^k/\tau \rightharpoonup \widehat{v}/\tau =: \widehat{u}$ and, from weak continuity of P_V, we have $x^k = P_V z^k \rightharpoonup P_V \widehat{z} =: \widehat{x}$. Finally, Proposition 3.2.3((III)) implies that $(\widehat{x}, \widehat{u}) \in (V \cap \operatorname{Fix} T) \times W$ is solution to Problem 3.2.1 and the result follows. □

Remark 3.2.5 (I) When T is weakly continuous, we have $Tr^k \rightharpoonup \widehat{x}$, where $(r^k)_{k\in\mathbb{N}}$ is defined in (3.2.6). Indeed, by the proof of [19, Theorem 3.1(ii)], the sequence $(p^k)_{k\in\mathbb{N}}$ in the algorithm (3.2.8) satisfies that $p^k - z^k \to 0$. Then $p^k = p^k - z^k + z^k \rightharpoonup \widehat{z}$ and, since P_V is weakly continuous, it follows that $r^k = P_V p^k \rightharpoonup P_V \widehat{z} = \widehat{x}$. Thus, since $\widehat{x} \in \operatorname{Fix} T$, $Tr^k \rightharpoonup T\widehat{x} = \widehat{x}$.

This fact helps to obtain a faster convergence in the context of convex optimization with affine linear constraints, as we will see in our numerical experiences. Indeed, in this case $\{Tr^k\}_{k\in\mathbb{N}} \subset C$, where C represents some selection of the affine linear constraints, which impose feasibility of the converging iterates explicitly.

(II) If $\mathcal{H}$ is finite dimensional and $T = P_C$ for a nonempty closed convex $C \subset \mathcal{H}$, we deduce $Tr^k \to \widehat{x}$ from Remark 3.2.5((I)) and the fact that P_C is continuous.

(III) When $V = \mathcal{H}$, we have that $V^\perp = \{0\}$ and $P_V = \operatorname{Id}$. Thus, the algorithm (3.2.6) reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \eta^{k+1} = J_{\gamma B^{-1}}(u^k + \gamma(L\overline{x}^k - D^{-1}u^k)) \\ u^{k+1} = P_W \eta^{k+1} \\ w^{k+1} = J_{\tau A}(x^k - \tau(L^* u^{k+1} + Cx^k)) \\ x^{k+1} = Tw^{k+1} \\ \overline{x}^{k+1} = x^{k+1} + w^{k+1} - x^k, \end{array} \right.$$

which is the algorithm proposed in [19, Theorem 3.1] when the stepsizes τ and γ are fixed.

(IV) When $T = \operatorname{Id}$, $W = \mathcal{G}$, and $B = D^{-1} = 0$, we have that for every $\lambda > 0$, $J_{\lambda B} = \operatorname{Id}$. Then, the algorithm (3.2.6) reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \widetilde{z}^k = x^k + \tau y^k - \tau P_V C x^k \\ w^{k+1} = J_{\tau A} \widetilde{z}^k \\ x^{k+1} = P_V w^{k+1} \\ y^{k+1} = y^k + (x^{k+1} - w^{k+1})/\tau, \end{array} \right.$$

which is the algorithm proposed in [12, Corollary 5.3] without relaxation ($\lambda_n \equiv 1$).

(V) In the context of the convex optimization problem (3.2.1), the proposed algorithm (3.2.6) reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \eta^{k+1} = \operatorname{prox}_{\gamma G^*} \left(u^k + \gamma(L\overline{x}^k - \nabla \ell^* u^k)\right) \\ u^{k+1} = P_W \eta^{k+1} \\ \widetilde{z}^k = x^k + \tau y^k - \tau P_V(L^* u^{k+1} + \nabla H(x^k)) \\ w^{k+1} = \operatorname{prox}_{\tau F} \widetilde{z}^k \\ r^{k+1} = P_V w^{k+1} \\ x^{k+1} = P_V T r^{k+1} \\ y^{k+1} = y^k + (r^{k+1} - w^{k+1})/\tau \\ \overline{x}^{k+1} = x^{k+1} + r^{k+1} - x^k. \end{array} \right. \tag{3.2.11}$$

In particular, when $T = \operatorname{Id}$, $V = \mathcal{H}$, $W = \mathcal{G}$, and $\ell = \iota_{\{0\}}$, we deduce that the algorithm

(3.2.11) is equivalent to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = \operatorname{prox}_{\gamma G^*}\left(u^k + \gamma L\overline{x}^k\right) \\ \widetilde{z}^k = x^k - \tau(L^* u^{k+1} + \nabla H(x^k)) \\ x^{k+1} = \operatorname{prox}_{\tau F} \widetilde{z}^k \\ \overline{x}^{k+1} = 2x^{k+1} - x^k, \end{array} \right. \tag{3.2.12}$$

which is an error-free version of the algorithm proposed in [29, Algorithm 3.1]. If additionally $H = 0$, the method (3.2.12) reduces to [20, Algorithm 1].

3.2.4. Numerical Experiences and Applications

In this section, we illustrate the efficiency of the proposed method in three instances. In the first instance we consider a non-differentiable constrained convex optimization problem without including a priori information on the solution ($T = \mathrm{Id}$), called constrained LASSO [34]. The constraint is given by the kernel of a linear operator and we apply our primal-dual method exploiting the vector subspace structure of the problem. For the second test, we depict the advantage of including such a priori information in a constrained ℓ^1-minimization problem, in which this feature is represented by a selection of the constraints. The last experience is an application of the proposed method to the capacity expansion problem in transport networks. We solve the two-stage stochastic arc capacity expansion problem over a directed graph using our primal-dual partial inverse method. The problem is to find the optimal investment decision in arc capacity and network flow operation under an uncertain environment. All proposed algorithms are implemented in MATLAB and run in a Mac mini (2018) 3 GHz 6-Core Intel Core i5 8GB RAM.

Constrained LASSO

We consider the following problem

$$\underset{\substack{x\in\mathbb{R}^n \\ Rx=0}}{\text{minimize}} \;\; \alpha\|x\|_1 + \frac{1}{2}\|Ax-b\|_2^2, \tag{3.2.13}$$

where $\alpha > 0$, $A \in \mathbb{R}^{p\times n}$, $R \in \mathbb{R}^{m\times n}$ satisfies $\ker R^\top = \{0\}$, and $b \in \mathbb{R}^p$. Note that, by setting $f = \alpha\|\cdot\|_1$, the problem in (3.2.13) can be written in at least three equivalent manners:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f(x) + \frac{1}{2}g_2(Ax) + \frac{1}{2}\iota_{\{0\}}(Rx), \;\text{ where }\; g_2 = \|\cdot - b\|_2^2; \tag{3.2.14}$$

$$\underset{x\in\ker R}{\text{minimize}}\, f(x) + h(x), \;\text{ where }\; h = \frac{1}{2}\|A(\cdot) - b\|_2^2, \tag{3.2.15}$$

and

$$\underset{x\in\ker R}{\text{minimize}}\, f(x) + g_1(Ax), \;\text{ where }\; g_1 = \frac{1}{2}\|\cdot - b\|_2^2. \tag{3.2.16}$$

Observe that $f \in \Gamma_0(\mathbb{R}^n)$, $g_1 \in \Gamma_0(\mathbb{R}^p)$, and $g_2 = 2g_1 \in \Gamma_0(\mathbb{R}^p)$. We have that $h\colon \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function with $\|A^\top A\|-$ Lipschitz continuous gradient. Therefore, the problem in (3.2.14) satisfies the hypotheses in [48, Corollary 4.2(i)]. Thus, since [6, Proposition 24.8(i)] yields $(\forall\gamma > 0)\;\; \operatorname{prox}_{\gamma g_1}\colon x \mapsto (x+\gamma b)/(\gamma+1)$, the primal-dual method proposed in [48, Corollary 4.2(i)] reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x^{k+1} = \operatorname{prox}_{\tau\alpha\|\cdot\|_1}(x^k - \frac{\tau}{2}A^\top v_1^k - \frac{\tau}{2}R^\top v_2^k) \\ y^k = 2x^{k+1} - x^k \\ v_1^{k+1} = 2(v_1^k + \sigma_1 Ay^k - \sigma_1 b)/(\sigma_1 + 2) \\ v_2^{k+1} = v_2^k + \sigma_2 Ry^k, \end{array} \right. \tag{3.2.17}$$

where $(x^0, v_1^0, v_2^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ and the strictly positive constants τ, σ_1, σ_2 satisfy the condition $\sqrt{\frac{\tau}{2}\sigma_1\|A\|^2 + \frac{\tau}{2}\sigma_2\|L\|^2} < 1$. In addition, by setting $V = \ker R$, the problem in (3.2.15) satisfies the hypotheses in [12, Proposition 6.7] and the Forward-Backward method with vector subspaces reduces to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} w^{k+1} = \operatorname{prox}_{\tau\alpha\|\cdot\|_1}(x^k + \lambda y^k - \lambda P_{\ker R}\nabla h(x^k)) \\ x^{k+1} = P_{\ker R} w^{k+1} \\ y^{k+1} = y^k + (x^{k+1} - w^{k+1})/\lambda \end{array} \right. , \tag{3.2.18}$$

where $x^0 \in \ker R$, $y^0 \in (\ker R)^\perp$, and $\lambda \in \left]0, 2/\|A^\top A\|\right[$. Moreover, by setting $\mathcal{H} = \mathbb{R}^n$, $\mathcal{G} = \mathbb{R}^p$, and $V = \ker R$, the problem in (3.2.16) satisfies the hypotheses in (3.2.1) which is a particular instance of Problem 3.2.1. Therefore, from (3.2.11) we obtain the following result in the case $T = \operatorname{Id}$.

Proposition 3.2.6 *Let $\{x^0, \overline{x}^0\} \subset \ker R$ such that $x^0 = \overline{x}^0$, let $y^0 \in (\ker R)^\perp$, let $u^0 \in \mathbb{R}^p$, and let $(\tau, \gamma) \in \left]0, +\infty\right[^2$ such that $\tau\gamma\|A\|^2 < 1$. Consider the following routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = (u^k + \gamma(A\overline{x}^k - b))/(\gamma + 1) \\ \widetilde{z}^k = x^k + \tau y^k - \tau P_{\ker R} A^\top u^{k+1} \\ w^{k+1} = \operatorname{prox}_{\tau\alpha\|\cdot\|_1} \widetilde{z}^k \\ x^{k+1} = P_{\ker R} w^{k+1} \\ y^{k+1} = y^k + (x^{k+1} - w^{k+1})/\tau \\ \overline{x}^{k+1} = 2x^{k+1} - x^k \end{array} \right. \tag{3.2.19}$$

Then there exists $x \in \ker R$ solution to (3.2.16) such that $x^k \to x$.

Note that $RR^\top$ is invertible since $\ker R^\top = \{0\}$. Then, by [6, Example 29.17(iii)], we have $P_{\ker R} = \operatorname{Id} - R^\top(RR^\top)^{-1}R$. On the other hand, by [6, proposition 24.11 & Example 24.20], we have

$$(\forall \tau > 0) \quad \operatorname{prox}_{\tau\|\cdot\|_1} \colon x \mapsto (\operatorname{prox}_{\tau|\cdot|} x_i)_{i=1}^n, \tag{3.2.20}$$

where

$$\operatorname{prox}_{\tau|\cdot|} \colon x \mapsto \begin{cases} x + \tau & \text{if } x < -\tau \\ 0 & \text{if } x \in [-\tau, \tau] \\ x - \tau & \text{if } x > \tau. \end{cases} \tag{3.2.21}$$

For each method, we obtain the average execution time and the average number of iterations from 20 random instances for the matrices A, R, and b, using $\alpha = 1$. We measure the efficiency for different values of m, n, and p. We label the algorithm in (3.2.17) as *PD generalized*, algorithm in (3.2.18) as *FB with subspaces*, and algorithm in (3.2.19) as *PD with subspaces*. For every algorithm, we obtain the values of τ, γ, λ, σ_1, and σ_2 by discretizing the parameter set in which the algorithm converges and selecting the parameters such that the method stops in a minimum number of iterations. This procedure is repeated for every dimension of matrices and vectors. In particular, we fix $\sigma_1 = \sigma_2$ for the method in (3.2.17). The results are shown in Table 3.1.

(n, p, m)	PD with subspaces	FB with subspaces	PD generalized
$(500, 250, 25)$	0.639 (3284)	3.909 (21059)	1.317 (4469)
$(500, 250, 50)$	0.822 (3565)	5.014 (22145)	1.254 (5081)
$(500, 250, 100)$	1.289 (3523)	7.956 (21374)	1.817 (5527)
$(500, 750, 25)$	0.488 (2184)	3.615 (16577)	1.012 (2991)
$(500, 750, 50)$	0.579 (2229)	4.197 (16445)	0.862 (3063)
$(500, 750, 100)$	0.854 (2117)	6.345 (15853)	1.129 (3066)
$(1000, 500, 50)$	3.032 (8910)	16.781 (49199)	4.665 (11125)
$(1000, 500, 100)$	5.278 (9716)	27.937 (51287)	5.591 (12615)
$(1000, 500, 200)$	10.830 (9036)	57.976 (48314)	7.283 (13014)
$(1000, 1500, 50)$	6.252 (4869)	44.335 (34553)	7.610 (6378)
$(1000, 1500, 100)$	7.911 (4992)	54.217 (34507)	8.691 (6484)
$(1000, 1500, 200)$	11.570 (4642)	79.844 (32110)	9.882 (6169)

Table 3.1: Average execution time (number of iterations) with relative error tolerance $e = 10^{-6}$.

We observe a substantial gain in efficiency when we use *PD with subspaces* with respect to the other two methods. The number of iterations is reduced in $25-30\%$ with respect to *PD generalized*. The construction of *PD with subspaces* exploiting the vector subspace and primal-dual structure of the problem explains these benefits. However, the computational time used by *PD with subspaces* is larger than that of *PD generalized* when the vector subspace is smaller ($m = 200$). The presence of two projections onto $\ker R$ at each iteration of *PD with subspaces* explains this behavior.

Constrained ℓ^1 minimization

In this subsection, we consider the following problem

$$\underset{\substack{x\in\mathbb{R}^n\\ Rx=0\\ Mx=c,\, Nx=d}}{\text{minimize}} \quad \alpha\|x\|_1, \tag{3.2.22}$$

where $\alpha > 0$, $R \in \mathbb{R}^{m_1\times n}$ satisfies $\ker R^\top = \{0\}$, $M \in \mathbb{R}^{m_2\times n}$ satisfies $\ker M^\top = \{0\}$, $N \in \mathbb{R}^{(p-m_1-m_2)\times n}$ with $p \geq m_1 + m_2$ (when $p = m_1 + m_2$, we remove the constraint $Nx = d$), $c \in \mathbb{R}^{m_2}$, and $d \in \mathbb{R}^{p-m_1-m_2}$. By defining $f = \alpha\|\cdot\|_1$, $g_1 = \iota_{\{b\}}$ with $b^\top = (c^\top, d^\top)$, $g_2 = \iota_{\{e\}}$ with $e^\top = (b^\top, 0^\top)$, $A^\top = [M^\top|N^\top]$, and $K^\top = [A^\top|R^\top]$ the problem in (3.2.22) can be written in the following three equivalent manners:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x) + g_2(Kx) \tag{3.2.23}$$

$$\underset{x\in\ker R}{\text{minimize}} \;\; f(x) + g_1(Ax) \tag{3.2.24}$$

$$\text{find } x \in \operatorname{Fix} P_{M^{-1}c} \cap \underset{x\in\ker R}{\operatorname{argmin}}\, f(x) + g_1(Ax). \tag{3.2.25}$$

We assume that $b \in \operatorname{ran} A$ for the existence of solutions to problem (3.2.22). Note that $f \in \Gamma_0(\mathbb{R}^n)$, $g_1 \in \Gamma_0(\mathbb{R}^{p-m_1})$, and $g_2 \in \Gamma_0(\mathbb{R}^p)$. To solve the problem in (3.2.23) we use the method proposed by Chambolle-Pock [20, Algorithm 1] (algorithm (3.2.12) with $H = 0$), which reduces in this case to

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = u^k + \gamma(K\overline{x}^k - e) \\ x^{k+1} = \operatorname{prox}_{\tau\alpha\|\cdot\|_1}(x^k - \tau K^\top u^{k+1}) \\ \overline{x}^{k+1} = 2x^{k+1} - x^k \end{array} \right. \tag{3.2.26}$$

where $(x^0, \overline{x}^0, u^0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p$ satisfies $x^0 = \overline{x}^0$, and $(\tau, \gamma) \in]0, +\infty[^2$ satisfies $\tau\gamma\|K\|^2 < 1$.

On the other hand, in order to solve the problems in (3.2.24) and (3.2.25), we use the algorithm in (3.2.11) with $W = \mathcal{G}$, $\ell = \iota_{\{0\}}$, and $H = 0$. We deduce the following convergence results from Theorem 3.2.4 for the cases with and without a priori information, represented by $M^{-1}c$.

Proposition 3.2.7 *Let $\{x^0, \overline{x}^0\} \subset \ker R$ such that $x^0 = \overline{x}^0$, let $y^0 \in (\ker R)^\perp$, let $u^0 \in \mathbb{R}^{p-m_1}$, and let $(\tau, \gamma) \in]0, +\infty[^2$ such that $\tau\gamma\|A\|^2 < 1$. Consider the following routine*

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = u^k + \gamma(A\overline{x}^k - b) \\ \widetilde{z}^k = x^k + \tau(y^k - P_{\ker R}A^\top u^{k+1}) \\ w^{k+1} = \operatorname{prox}_{\tau\alpha\|\cdot\|_1} \widetilde{z}^k \\ x^{k+1} = P_{\ker R} w^{k+1} \\ y^{k+1} = y^k + (x^{k+1} - w^{k+1})/\tau \\ \overline{x}^{k+1} = 2x^{k+1} - x^k. \end{array} \right. \tag{3.2.27}
$$

Then there exists $x \in \mathbb{R}^n$ solution to (3.2.24) such that $x^k \to x$.

Proposition 3.2.8 *With the same hypotheses as the Proposition 3.2.7, consider the following routine*

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = u^k + \gamma(A\overline{x}^k - b) \\ \widetilde{z}^k = x^k + \tau(y^k - P_{\ker R}A^\top u^{k+1}) \\ w^{k+1} = \operatorname{prox}_{\tau\alpha\|\cdot\|_1} \widetilde{z}^k \\ r^{k+1} = P_{\ker R} w^{k+1} \\ x^{k+1} = P_{\ker R} P_{M^{-1}c} r^{k+1} \\ y^{k+1} = y^k + (r^{k+1} - w^{k+1})/\tau \\ \overline{x}^{k+1} = x^{k+1} + r^{k+1} - x^k. \end{array} \right. \tag{3.2.28}
$$

Then there exists $x \in \mathbb{R}^n$ solution to (3.2.25) such that $x^k \to x$.

It follows from $\ker M^\top = \{0\}$ that $MM^\top$ is invertible, and we have $P_{M^{-1}c} \colon x \mapsto x - M^\top(MM^\top)^{-1}(Mx - c)$ [6, Example 29.17(iii)].

For the numerical experience, we set $n = 1000$, $p = 100$, and $\alpha = 1$. In Table 3.2 we illustrate, for each method, the average execution time and the average number of iterations for 20 random instances of K and $b \in \operatorname{ran} A$. For every instance, we set R as the last m_1 rows of K with $m_1 \in \{1, 10, 20, 40\}$ and we apply the algorithm in (3.2.28) selecting M as the first m_2 rows of K with $m_2 \in \{1, 10, 20, 30, 50\}$. For every algorithm, we obtain the values of τ and γ as in Section 3.2.4.

Algorithm	$m_1 = 1$	$m_1 = 10$	$m_1 = 20$	$m_1 = 40$
Chambolle-Pock	8.374 (28905)	7.915 (26716)	7.611 (25960)	7.754 (26378)
(3.2.27)	4.807 (29559)	6.123 (27002)	6.372 (26143)	7.205 (24455)
(3.2.28) with $m_2 = 1$	4.533 (27491)	6.664 (25886)	7.052 (25034)	8.368 (23492)
(3.2.28) with $m_2 = 10$	4.209 (21877)	6.279 (22292)	6.627 (21847)	7.490 (19715)
(3.2.28) with $m_2 = 20$	3.826 (19703)	5.871 (20546)	6.353 (20469)	6.921 (18073)
(3.2.28) with $m_2 = 30$	3.768 (18608)	5.827 (19725)	6.365 (19947)	6.891 (17466)
(3.2.28) with $m_2 = 50$	4.023 (17914)	6.050 (19253)	6.591 (19529)	7.093 (16988)

Table 3.2: Average execution time (number of iterations) with relative error tolerance $e = 10^{-5}$.

We observe that the vector subspace structure of algorithm in (3.2.27) provides an improvement up to 43% in computational time with respect to Chambolle-Pock's splitting. The improvement is larger when the vector subspace is larger ($m_1 = 1$). This improvement reduces to 7% for smaller vector subspaces ($m_1 = 40$). Moreover, by including the a priori information via the projection onto $M^{-1}c$, we observe an important reduction on the number of iterations for achieving convergence with respect to the algorithm in (3.2.27), following the same behavior perceived in [19]. We explain this behavior from Remark 3.2.5(I), since $\widetilde{x}^k := P_{M^{-1}c}r^k \to x$ and $(\widetilde{x}^k)_{k\in\mathbb{N}} \subset M^{-1}c$. Indeed, the method forces iterations to satisfy at least a selection of the constraints and its efficiency increases as we enlarge the selection m_2. In fact, we obtain 7% of improvement on number of iterations when $m_2 = 1$ and 39% when $m_2 = 50$ with respect to the algorithm in (3.2.27). The improvement is again more important when the vector subspace is larger. Moreover, for larger values of m_2, the computation of $P_{M^{-1}c}$ becomes more time expensive, increasing the total computational time in (3.2.28). In the case when $m_2 = 1$, even if there is a reduction in iterations, there is a slight increase on the average execution time because of an additional projection onto $\ker R$ at each iteration.

Capacity expansion problem in transport networks

In this section we aim at solving the traffic assignment problem with arc-capacity expansion on a network with minimal cost under uncertainty. Let $\mathcal{A}$ be the set of arcs and let $\mathcal{O}$ and $\mathcal{D}$ be the sets of origin and destination nodes of the network, respectively. The set of routes from $o \in \mathcal{O}$ to $d \in \mathcal{D}$ is denoted by R_{od} and $R := \cup_{(o,d)\in\mathcal{O}\times\mathcal{D}} R_{od}$ is the set of all routes. The arc-route incidence matrix $N \in \mathbb{R}^{|\mathcal{A}|\times|R|}$ is defined by $N_{ar} := 1$, if arc a belongs to the route r, and $N_{ar} := 0$, otherwise.

The uncertainty is modeled by a finite set Ξ of possible scenarios. For every scenario $\xi \in \Xi$, $p_\xi \in [0,1]$ is its probability of occurrence, $h_{od,\xi} \in \mathbb{R}_+$ is the forecasted demand from $o \in \mathcal{O}$ to $d \in \mathcal{D}$, $c_{a,\xi} \in \mathbb{R}_+$ is the corresponding capacity of the arc $a \in \mathcal{A}$, $t_{a,\xi} \colon \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing and $\chi_{a,\xi}-$Lipschitz continuous travel time function on arc $a \in \mathcal{A}$, for some $\chi_{a,\xi} > 0$, and the variable $f_{r,\xi} \in \mathbb{R}_+$ stands for the flow in route $r \in R$.

In the problem of this section, we consider the expansion of flow capacity at each arc in order to improve the efficiency of the network operation. We model this decision making process in a two-stage stochastic problem. The first stage reflects the investment in capacity and the second corresponds to the operation of the network in an uncertain environment.

In order to solve this problem, we take a non-anticipativity approach [22], letting our first stage decision variable depend on the scenario and imposing a non-anticipativity constraint. We denote by $x_{a,\xi} \in \mathbb{R}_+$ the variable of capacity expansion on arc $a \in \mathcal{A}$ in scenario $\xi \in \Xi$ and the non-anticipativity condition is defined by the constraint

$$\mathcal{N} := \left\{x \in \mathbb{R}^{|\mathcal{A}|\times|\Xi|} \,:\, (\forall(\xi,\xi') \in \Xi^2) \quad x_\xi = x_{\xi'}\right\},$$

where $x_\xi \in \mathbb{R}^{|\mathcal{A}|}$ is the vector of capacity expansion for scenario $\xi \in \Xi$ and we denote $f_\xi \in \mathbb{R}^{|R|}$ analogously. We restrict the capacity expansion variables by imposing, for every $a \in \mathcal{A}$ and $\xi \in \Xi$, $x_{a,\xi} \in [0, M_a]$, where $M_a > 0$ represents the upper bound of capacity expansion on arc $a \in \mathcal{A}$. Additionally, we model the investment cost of expansion via a quadratic function defined by a symmetric positive definite matrix $Q \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$.

Problem 3.2.9 The problem is to

$$\underset{(x,f)\in(\mathcal{N}\cap D^{|\Xi|})\times\mathbb{R}_+^{|R||\Xi|}}{\text{minimize}} \sum_{\xi\in\Xi} p_\xi \left[\sum_{a\in\mathcal{A}} \int_0^{(Nf_\xi)_a} t_{a,\xi}(z)dz + \frac{1}{2}x_\xi^\top Q x_\xi\right]$$

$$\text{s.t.} \quad (\forall \xi \in \Xi)(\forall a \in \mathcal{A}) \quad (Nf_\xi)_a - x_{a,\xi} \leq c_{a,\xi}, \tag{3.2.29}$$

$$(\forall \xi \in \Xi)(\forall (o,d) \in \mathcal{O} \times \mathcal{D}) \quad \sum_{r \in R_{od}} f_{r,\xi} = h_{od,\xi}, \tag{3.2.30}$$

where $D := \bigtimes_{a \in \mathcal{A}} [0, M_a]$, and we assume the existence of solutions.

The first term of the objective function in Problem 3.2.9 represents the expected operational cost of the network. The optimality conditions of the optimization problem with this objective cost related to the pure traffic assignment problem, defines a Wardrop equilibrium [7]. The second term in the objective function is the expansion investment cost. Constraints in (3.2.29) represent that, for every arc $a \in \mathcal{A}$, the flow cannot exceed the expanded capacity $c_{a,\xi} + x_{a,\xi}$ at each scenario $\xi \in \Xi$, while (3.2.30) are the demand constraints.

We solve Problem 3.2.9 following the structure of the problem in (3.2.1) with $T = \operatorname{Id}$. We consider the following two equivalent formulations.

Primal-Dual Formulation

Note that Problem 3.2.9 can be equivalently written as

$$\underset{(x,f) \in \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|}}{\text{minimize}} F(x,f) + G(L(x,f)) + H(x,f), \tag{$\mathcal{P}$}$$

where

$$\begin{cases} (\forall \xi \in \Xi) \quad V_\xi^+ := \left\{ f \in \mathbb{R}_+^{|R|} \,:\, (\forall (o,d) \in \mathcal{O} \times \mathcal{D}) \ \sum_{r \in R_{od}} f_r = h_{od,\xi} \right\} \\ \Lambda := (D^{|\Xi|} \cap \mathcal{N}) \times \left(\bigtimes_{\xi \in \Xi} V_\xi^+ \right) \\ F := \iota_\Lambda \\ (\forall \xi \in \Xi) \quad \Theta_\xi := \left\{ (x,u) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|} \,:\, (\forall a \in \mathcal{A}) \ u_a - x_a \leq c_{a,\xi} \right\} \\ G := \iota_{\bigtimes_{\xi \in \Xi} \Theta_\xi} \\ L \colon (x,f) \mapsto (x_\xi, Nf_\xi)_{\xi \in \Xi} \\ H \colon (x,f) \mapsto \displaystyle\sum_{\xi \in \Xi} p_\xi \left[\sum_{a \in \mathcal{A}} \int_0^{(Nf_\xi)_a} t_{a,\xi}(z) dz + \frac{1}{2} x_\xi^\top Q x_\xi \right]. \end{cases} \tag{3.2.31}$$

Observe that F and G are lower semicontinuous convex proper functions, and L is linear and bounded with $\|L\| \leq \max\{1, \|N\|\}$. Moreover, we have that H is a convex function with gradient Lipschitz. Indeed, since $(t_{a,\xi})_{a \in \mathcal{A}, \xi \in \Xi}$ are increasing, $f \mapsto Nf$ is linear, and Q is definite positive, we conclude that H is a separable convex function. On the other hand, by the chain's rule, the fundamental theorem of calculus, and by defining

$$\psi \colon f \mapsto (p_\xi N^\top (t_{a,\xi}((Nf_\xi)_a))_{a \in \mathcal{A}})_{\xi \in \Xi},$$

it follows that

$$\begin{aligned} \nabla H \colon (x,f) \mapsto & \left((p_\xi Q x_\xi)_{\xi \in \Xi}, \left(p_\xi N^\top \nabla \varphi_\xi (Nf_\xi) \right)_{\xi \in \Xi} \right) \\ & = \left((p_\xi Q x_\xi)_{\xi \in \Xi}, \psi(f) \right). \end{aligned} \tag{3.2.32}$$

Then, for every (x, f) and (x', f') in $\mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|}$, we have

$$\begin{aligned}
\|\nabla H(x,f) - \nabla H(x',f')\|^2 &= \sum_{\xi\in\Xi} \|p_\xi Q(x_\xi - x'_\xi)\|^2 + \|p_\xi N^\top (t_{a,\xi}((Nf_\xi)_a) - t_{a,\xi}((Nf'_\xi)_a))_{a\in\mathcal{A}}\|^2 \\
&\leq \sum_{\xi\in\Xi} p_\xi^2 \left(\|Q\|^2 \|x_\xi - x'_\xi\|^2 + \|N^\top\|^2 \sum_{a\in\mathcal{A}} \chi_{a,\xi}^2 \|N_{a\bullet}\|^2 \|f_\xi - f'_\xi\|^2 \right) \\
&\leq \sum_{\xi\in\Xi} p_\xi^2 (\|Q\|^2 \|x_\xi - x'_\xi\|^2 + \|N^\top\|^2 \|N\|^2 \max_{a\in\mathcal{A}} \chi_{a,\xi}^2 \|f_\xi - f'_\xi\|^2) \\
&\leq \sum_{\xi\in\Xi} p_\xi^2 \max\{\|Q\|^2, \|N\|^4 \max_{a\in\mathcal{A}} \chi_{a,\xi}^2\} (\|x_\xi - x'_\xi\|^2 + \|f_\xi - f'_\xi\|^2) \\
&\leq \max_{\xi\in\Xi} \left(p_\xi^2 \max\{\|Q\|^2, \|N\|^4 \max_{a\in\mathcal{A}} \chi_{a,\xi}^2\} \right) \|(x,f) - (x',f')\|^2. \quad (3.2.33)
\end{aligned}$$

Thus, ∇H is Lipschitz continuous with constant

$$\beta^{-1} = \max_{\xi\in\Xi} \left(p_\xi \max\left\{ \|Q\|, \|N\|^2 \max_{a\in\mathcal{A}} \chi_{a,\xi} \right\} \right). \quad (3.2.34)$$

Altogether, $(\mathcal{P})$ is a particular instance of problem in (3.2.1) with $V = \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|}$ and $\ell = \iota_{\{0\}}$. Therefore, from the algorithm in (3.2.12) we obtain the following result.

Proposition 3.2.10 *Let (x^0, f^0) and $(\overline{x}^0, \overline{f}^0)$ in $\mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|}$ such that $(x^0, f^0) = (\overline{x}^0, \overline{f}^0)$, let $(p^0, u^0) \in \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|\mathcal{A}||\Xi|}$, let $\tau \in]0, 2\beta[$, and let $\gamma \in]0, +\infty[$ such that $\tau\gamma \max\{1, \|N\|^2\} < 1 - \tau/2\beta$. Consider the following routine:*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l}
\text{For every } \xi \in \Xi \\
\left\lfloor \begin{array}{l}
(\widetilde{p}_\xi^k, \widetilde{u}_\xi^k) = (p_\xi^k + \gamma \overline{x}_\xi^k, u_\xi^k + \gamma N \overline{f}_\xi^k) \\
(p_\xi^{k+1}, u_\xi^{k+1}) = (\widetilde{p}_\xi^k, \widetilde{u}_\xi^k) - \gamma P_{\Theta_\xi}(\gamma^{-1}(\widetilde{p}_\xi^k, \widetilde{u}_\xi^k))
\end{array} \right. \\
\widetilde{x}^k = x^k - \tau(p^{k+1} + (p_\xi Q x_\xi^k)_{\xi\in\Xi}) \\
\widetilde{f}^k = f^k - \tau(N^\top u^{k+1} + \psi(f^k)) \\
x^{k+1} = P_{D^{|\Xi|} \cap \mathcal{N}} \widetilde{x}^k \\
(\forall \xi \in \Xi) \quad f_\xi^{k+1} = P_{V_\xi^+} \widetilde{f}_\xi^k \\
\overline{x}^{k+1} = 2x^{k+1} - x^k \\
\overline{f}^{k+1} = 2f^{k+1} - f^k.
\end{array} \right. \quad (3.2.35)$$

Then there exists $(\widehat{x}, \widehat{f})$ solution to Problem 3.2.9 such that $(x^k, f^k) \to (\widehat{x}, \widehat{f})$.

Remark 3.2.11 Note that the projections appearing in (3.2.35), $(P_{\Theta_\xi})_{\xi\in\Xi}$, $P_{D^{|\Xi|}\cap\mathcal{N}}$, $(P_{V_\xi^+})_{\xi\in\Xi}$ can be computed efficiently.

(i) Let $\xi \in \Xi$ and let $(x, u) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|}$. We deduce from [6, Proposition 29.3 & Example 29.20] that $P_{\Theta_\xi}(x, u) = (P_{a,\xi}(x, u))_{a\in\mathcal{A}}$, where

$$(\forall a \in \mathcal{A}) \ P_{a,\xi}(x, u) = \begin{cases} \left(\frac{x_a + u_a - c_{a,\xi}}{2}, \frac{x_a + u_a + c_{a,\xi}}{2} \right), & \text{if } u_a - x_a - c_{a,\xi} > 0; \\ (x_a, u_a), & \text{otherwise.} \end{cases}$$

(ii) Let $\xi \in \Xi$ and note that

$$V_\xi^+ = \bigtimes_{(o,d)\in\mathcal{O}\times\mathcal{D}} V_{od,\xi},$$

where, for every $(o,d) \in \mathcal{O}\times\mathcal{D}$, $V_{od,\xi} := \{f \in \mathbb{R}_+^{|R_{od}|} \mid \sum_{r\in R_{od}} f_r = h_{od,\xi}\}$. It follows from [6, Proposition 29.3] that

$$P_{V_\xi^+} \colon f = (f_{od})_{\substack{o\in\mathcal{O}\\ d\in\mathcal{D}}} \mapsto (P_{V_{od,\xi}} f_{od})_{\substack{o\in\mathcal{O}\\ d\in\mathcal{D}}}. \tag{3.2.36}$$

For every $(o,d) \in \mathcal{O}\times\mathcal{D}$, the projection $P_{V_{od,\xi}}$ can be computed efficiently by using the quasi-Newton algorithm developed in [27].

(iii) Note that

$$D^{|\Xi|} \cap \mathcal{N} = \bigtimes_{a\in\mathcal{A}} C_a,$$

where, for every $a \in \mathcal{A}$, $C_a = \{y \in [0, M_a]^{|\Xi|} \mid (\forall(\xi,\xi') \in \Xi^2)\ y_\xi = y_{\xi'}\}$. It follows from [6, Proposition 29.3] that $P_{D^{|\Xi|}\cap\mathcal{N}} \colon x = (x_a)_{a\in\mathcal{A}} \mapsto (P_{C_a} x_a)_{a\in\mathcal{A}}$ and

$$(\forall a \in \mathcal{A}) \quad P_{C_a} \colon y \mapsto \operatorname{mid}(0, \overline{y}, M_a)\mathbf{1}, \tag{3.2.37}$$

where $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbb{R}^{|\Xi|}$, $\overline{y} = \frac{1}{|\Xi|}\sum_{\xi\in\Xi} y_\xi$, and $\operatorname{mid}(a,b,c)$ is the middle value among a, b, and c. In order to prove (3.2.37), let $a \in \mathcal{A}$, $y \in \mathbb{R}^{|\Xi|}$, and $x \in C_a$. Then $x = \eta\mathbf{1}$ from some $\eta \in [0, M_a]$. Hence, defining $\lambda = \operatorname{mid}(0, \overline{y}, M_a)$, we have that

$$\begin{aligned}
(x - \lambda\mathbf{1})^\top(y - \lambda\mathbf{1}) &= \sum_{\xi\in\Xi}(x_\xi - \lambda)(y_\xi - \lambda)\\
&= (\eta - \lambda)|\Xi|(\overline{y} - \lambda)\\
&= \begin{cases} \eta|\Xi|\overline{y} & \text{if} \quad \overline{y} < 0\\ 0 & \text{if} \quad \overline{y} \in [0, M_a]\\ (\eta - M_a)|\Xi|(\overline{y} - M_a) & \text{if} \quad \overline{y} > M_a \end{cases}\\
&\leq 0.
\end{aligned}$$

On the other hand, it is clear that $\lambda \in [0, M_a]$.

Primal-Dual Partial Inverse Formulation

For the second equivalent formulation of Problem 3.2.9 consider the closed vector subspace

$$S = \left\{ f \in \mathbb{R}^{|R||\Xi|} \;:\; (\forall\xi \in \Xi)(\forall(o,d) \in \mathcal{O}\times\mathcal{D}) \sum_{r\in R_{od}} f_{r,\xi} = 0 \right\}$$

and let $\widehat{f}^0$ defined by

$$(\forall\xi \in \Xi)(\forall(o,d) \in \mathcal{O}\times\mathcal{D})(\forall r \in R_{od}) \quad \widehat{f}^0_{r,\xi} = h_{od,\xi}/|R_{od}|,$$

which satisfies (3.2.30). Then, under the notation in (3.2.31), the Problem 3.2.9 is equivalent to

$$\underset{(x,f)\in\mathcal{N}\times S}{\text{minimize}}\ \widehat{F}(x, f + \widehat{f}^0) + G(L(x, f + \widehat{f}^0)) + H(x, f + \widehat{f}^0), \tag{$\mathcal{P}_V$}$$

where $\widehat{F} = \iota_{D^{|\Xi|}\times\mathbb{R}_+^{|R||\Xi|}}$. Note that, the difference with respect to $(\mathcal{P})$ is that in $(\mathcal{P}_V)$ we propose a vector subspace splitting on function F defined in (3.2.31).

In addition, observe that $\widehat{F}$ and $G(\cdot + L(0,\widehat{f}^0))$ are lower semicontinuous, convex, and proper, and $H(\cdot + (0,\widehat{f}^0))$ is convex differentiable with $\beta^{-1}-$Lipschitz gradient. Thus, $(\mathcal{P}_V)$ satisfies the hypotheses of problem (3.2.1) with $V = \mathcal{N}\times S$ and $\ell = \iota_{\{0\}}$. Hence, by using [6, Proposition 29.1(i)] in (3.2.11) with $T = \mathrm{Id}$, we obtain the following result.

Proposition 3.2.12 *Let (x^0, f^0) and $(\overline{x}^0, \overline{f}^0)$ in $\mathcal{N}\times S$ such that $(x^0, f^0) = (\overline{x}^0, \overline{f}^0)$, let $(p^0, u^0) \in \mathbb{R}^{|\mathcal{A}||\Xi|}\times\mathbb{R}^{|\mathcal{A}||\Xi|}$, let (y^0, g^0) in $\mathcal{N}^\perp\times S^\perp$, let $\tau\in\,]0,2\beta[$, and let $\gamma\in\,]0,+\infty[$ be such that $\tau\gamma\max\{1,\|N\|^2\} < 1-\tau/2\beta$. Consider the following routine:*

$$(\forall k\in\mathbb{N})\quad \left\lfloor \begin{array}{l} \widetilde{p}^k = p^k + \gamma\overline{x}^k \\ \widetilde{u}^k = u^k + \gamma N\overline{f}^k \\ (\forall\xi\in\Xi)\quad (p_\xi^{k+1}, u_\xi^{k+1}) = (\widetilde{p}_\xi^k, \widetilde{u}_\xi^k + \gamma N\widehat{f}_\xi^0) - \gamma P_{\Theta_\xi}(\gamma^{-1}(\widetilde{p}_\xi^k, \widetilde{u}_\xi^k + \gamma N\widehat{f}_\xi^0)) \\ \widetilde{x}^k = x^k + \tau y^k - \tau P_{\mathcal{N}}(p^{k+1} + (p_\xi Q x_\xi^k)_{\xi\in\Xi}) \\ \widetilde{f}^k = f^k + \tau g^k - \tau P_S(N^\top u^{k+1} + \psi(f^k + \widehat{f}^0)) \\ z^{k+1} = P_{D^{|\Xi|}}\widetilde{x}^k \\ \ell^{k+1} = P_{\mathbb{R}_+^{|R||\Xi|}}(\widetilde{f}^k + \widehat{f}^0) - \widehat{f}^0 \\ x^{k+1} = P_{\mathcal{N}} z^{k+1} \\ f^{k+1} = P_S \ell^{k+1} \\ y^{k+1} = y^k + (x^{k+1} - z^{k+1})/\tau \\ g^{k+1} = g^k + (f^{k+1} - \ell^{k+1})/\tau \\ \overline{x}^{k+1} = 2x^{k+1} - x^k \\ \overline{f}^{k+1} = 2f^{k+1} - f^k. \end{array}\right. \tag{3.2.38}$$

Then there exists $(\widehat{x}, \widehat{f})$ solution to Problem 3.2.9 such that $(x^k, f^k)\to(\widehat{x},\widehat{f})$.

Remark 3.2.13 The projections in (3.2.38) are explicit. Indeed, for every $x\in\mathbb{R}^{|\mathcal{A}||\Xi|}$ we have, for every $\xi\in\Xi$, $(P_{\mathcal{N}}x)_\xi = \frac{1}{|\Xi|}\sum_{\xi'\in\Xi} x_{\xi'}$ and $(P_{D^{|\Xi|}}x)_\xi = (\mathrm{mid}(0, x_{a,\xi}, M_a))_{a\in\mathcal{A}}$. Moreover, for every $f\in\mathbb{R}^{|R||\Xi|}$ we have, for every $(o,d)\in\mathcal{O}\times\mathcal{D}$ and $r\in R_{od}$, $(P_S f)_r = (f_{r,\xi} - \frac{1}{|R_{od}|}\sum_{r'\in R_{od}} f_{r',\xi})_{\xi\in\Xi}$ and $(P_{\mathbb{R}_+^{|R||\Xi|}} f)_r = (\max\{0, f_{r,\xi}\})_{\xi\in\Xi}$.

Numerical experiences

In this subsection we compare the efficiency of the algorithms in (3.2.35) and (3.2.38) to solve the arc capacity expansion problem. We consider two networks used in [22]. Network 1, represented in Figure 3.1, has 7 arcs and 6 paths and Network 2, in Figure 3.2, has 19 arcs and 25 paths.

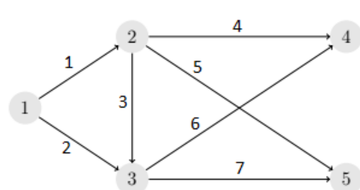

Figure 3.1: Network 1.

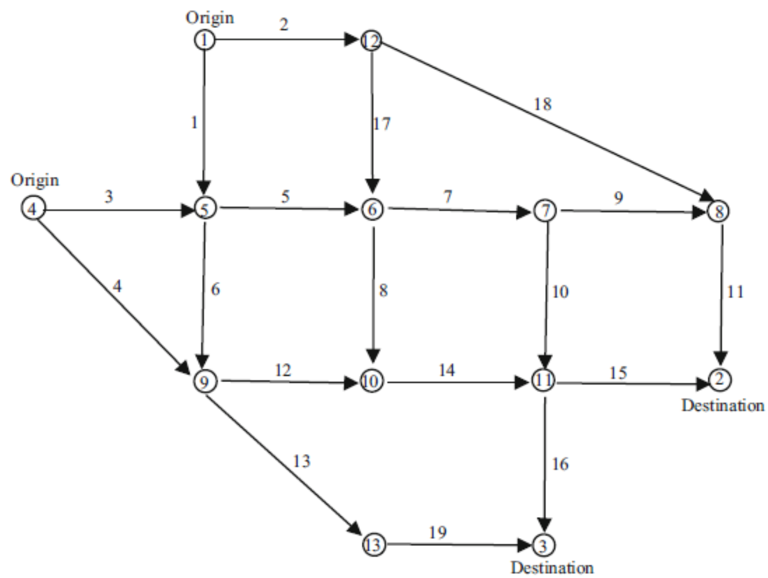

Figure 3.2: Network 2.

In our numerical experiences we set $p_\xi \equiv 1/|\Xi|$, $(c_\xi)_{\xi\in\Xi}$ as a sample of the random variable $100 \cdot b + d \cdot \text{beta}(2,2)$, where

$$b = \begin{cases} (1,1,2,2,1,1,1) & \text{in Network 1} \\ (10,4.4,1.4,10,3,4.4,10,2,2,4,7,7,7,7,4,3.5,2.2,4.4,7) & \text{in Network 2} \end{cases}$$

and

$$d = \begin{cases} (15,15,30,30,15,15,15) \quad \text{in Network 1} \\ (15,6.6,2.1,15,4.5,6.6,15,3,3,6,10.5, \\ \qquad\qquad 10.5,10.5,10.5,6,5.25,3.3,6.6,10.5) \quad \text{in Network 2,} \end{cases}$$

and $(h_\xi)_{\xi\in\Xi}$ as a sample of the random variable

$$h = \begin{cases} (h_{1,4},h_{1,5}) \sim (150,180) + (120,96)\cdot \text{beta}(5,1) \quad \text{in Network 1} \\ (h_{1,2},h_{1,3},h_{4,2},h_{4,3}) \sim (300,700,500,350) \\ \qquad\qquad +(120,120,120,120)\cdot \text{beta}(50,10) \quad \text{in Network 2.} \end{cases}$$

We set the capacity limits $(M_a)_{a\in\mathcal{A}} = \theta d$, where $\theta = 40$ in Network 1 and $\theta = 200$ in Network 2. We also set $Q = \text{Id} \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$ and, for every $a \in \mathcal{A}$ and $\xi \in \Xi$, the travel time function is $t_{a,\xi}\colon u \mapsto \eta_a(1 + 0.15\, u/c_{a,\xi})$, where

$$\eta = \begin{cases} (6,4,3,5,6,4,1) \quad \text{in Network 1} \\ (7,9,9,12,3,9,5,13,5,9,9,10,9,6,9,8,7,14,11) \quad \text{in Network 2.} \end{cases}$$

We implement the algorithms in (3.2.35) and (3.2.38) for different values of $|\Xi|$. We obtain the following results by considering 20 random realizations of $(c_\xi)_{\xi\in\Xi}$ and $(h_\xi)_{\xi\in\Xi}$.

Network 1	$\lvert\Xi\rvert = 1$	$\lvert\Xi\rvert = 3$	$\lvert\Xi\rvert = 5$	$\lvert\Xi\rvert = 10$
algorithm (3.2.35)	0.082 (1143)	0.731 (3217)	1.363 (4199)	4.388 (5698)
algorithm (3.2.38)	0.075 (1160)	0.607 (3284)	1.098 (4294)	3.485 (5804)
% improvement of time	8.54%	16.96%	19.44%	20.58%

Network 2	$\lvert\Xi\rvert = 1$	$\lvert\Xi\rvert = 3$	$\lvert\Xi\rvert = 5$	$\lvert\Xi\rvert = 10$
algorithm (3.2.35)	0.864 (4801)	10.195 (27285)	16.166 (27660)	45.327 (39790)
algorithm (3.2.38)	0.637 (4816)	7.627 (28147)	12.069 (28885)	33.204 (40848)
% improvement of time	26.27%	25.19%	25.34%	26.75%

Table 3.3: Average execution time (number of iterations) with relative error tolerance $e = 10^{-10}$.

Note that the algorithm with vector subspaces in (3.2.38) is more time efficient compared to the classical primal-dual algorithm in (3.2.35). Indeed, the percentage of improvement reaches up to 26.75%, for the larger dimensional case of Network 2 and Ξ. It is worth to notice that the number of iterations is lower in average for the approach without vector subspaces, but it is explained by the subroutines that compute the projections onto $(V_\xi^+)_{\xi\in\Xi}$, which lead to a larger computational time by iteration. In order to show the difference of both algorithms, in Figure 3.1 and Figure 3.2 we illustrate the relative error depending on the execution time for the best and the worst instance respect to convergence time.

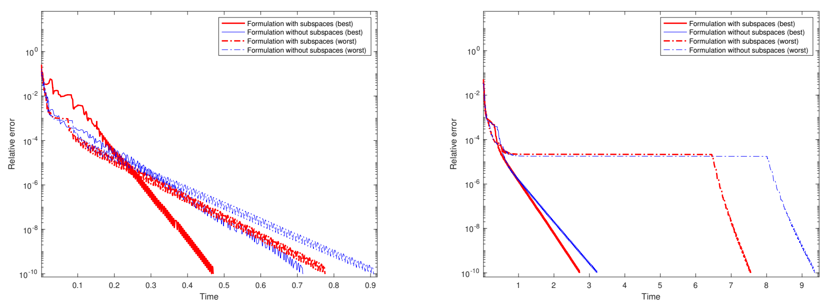

Figure 3.3: Relative error in Network 1 (semilog) with $\lvert\Xi\rvert = 3$ (left) and $\lvert\Xi\rvert = 10$ (right).

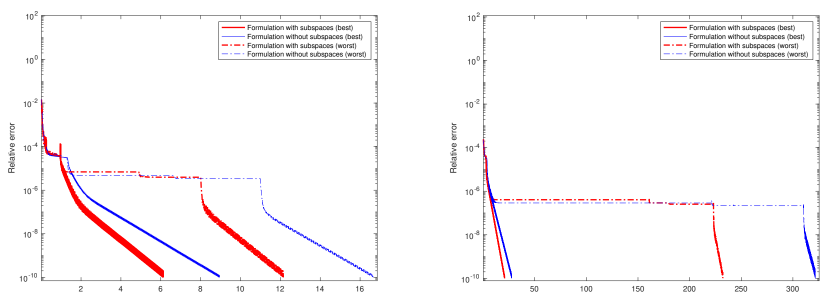

Figure 3.4: Relative error in Network 2 (semilog) with $\lvert\Xi\rvert = 3$ (left) and $\lvert\Xi\rvert = 10$ (right).

3.2.5. Conclusions

We propose a primal-dual method with partial inverse for solving constrained composite monotone inclusions involving a normal cone to a closed vector subspace. When the monotone operators are subdifferentials of convex functions, our method solves composite convex optimization problems over closed vector subspaces. We also incorporate a priori information on the solution of the monotone inclusion, which produces an additional projection step in the primal-dual algorithm. Either this projection or our vector subspace approach produces significant gains in numerical efficiency with respect to the available methods in the literature.

Chapter 4

Conclusions and perspectives

In Chapter 2, we provide a projected primal-dual splitting method for solving composite monotone inclusions with a priori information on solutions, generalizing various methods in the literature [9, 20, 29, 48]. Moreover, in the presence of strong monotonicity, we prove accelerated convergence of the primal iterates generated by our method, generalizing [20, Theorem 2] and giving an accelerated version of the method proposed in [48] for strongly monotone inclusions. Finally, we demonstrate linear convergence in the fully strongly monotone case, generalizing [20, Theorem 3] and complementing the ergodic linear convergence in the case without projection in [21].

The importance of the a priori information set is illustrated via the minimization problem of the norm of ℓ^1 with affine constraints, in which the proposed method outperforms [20]. In fact, the greater the selection of constraints on which it is possible to project, the greater the percentage of improvement of the proposed method with respect to the Chambolle-Pock's method [20]. In addition, the percentage of improvement is greater as the tolerance for the methods increases, reaching up to $69,3\%$ of improvement on the computational time.

In Chapter 3, we provide a primal-dual method of partial inverses to solve monotone inclusions with vector subspaces and a priori information on solutions, generalizing the proposed method in Chapter 2 when the step sizes are fixed.

The efficiency of the proposed method is illustrated in two non-differentiable convex optimization problems. The first is the constrained LASSO [34], in which we compare our method without a priori information with respect to the methods in [12] and [48]. We conclude that when the vector subspace constraint is a larger set, the improvement on the computational time is larger with respect to [48]. This behavior is due to the fact that our algorithm requires calculating the projection on the vector subspace, which is more expensive in time as the subspace is smaller. The second application is the constrained ℓ^1 minimization problem, in which we use the a priori information. We observe that the proposed method provides an improvement up to 43% in computational time with respect [20] and the improvement is larger when the vector subspace is larger. Moreover, as is perceived in Chapter 2, by including the a priori information via the projection onto a selection of affine linear constraints, we observe an important reduction on the number of iterations with respect to the algorithm without such information.

As an engineering application, we solve the arc-capacity expansion problem in a network. We conclude that by introducing a proper vector subspace in the formulation, there is up to 26.75%

of improvement with respect to the formulation without subspaces. In general, the percentage of improvement is greater in the larger network and when there are more scenarios for the problem.

As a perspective we will further study the accelerated convergence and linear convergence of the primal-dual splitting method of partial inverses proposed in Chapter 3. An possible hypothesis to obtain these types of convergence is the strong monotonicity of the maximally monotone operators involved in the problem, varying the step sizes as the algorithm in [20].

An interesting generalization to study is the case when the operator C in Problem 1.2.1 is monotone and Lipschitz. This allows us to generalize the inclusion problem studied in [26], as well as the problem in [13], which include a vector subspace in the inclusion.

Another possible research topic is to consider non-monotone operators in the inclusion or with weaker hypotheses than monotony, such as hypomonotone operators.

Finally, note that the a priori information has been used in the context of convex optimization with affine linear equality constraints. However, we can also consider convex non linear constraints of the form $\{x \,|\, g(x) \leq \alpha\}$ for a convex l.s.c. function g. In the case when $P_{\{g \leq \alpha\}}$ cannot be calculated efficiently, the study of other techniques for solving this problem are part of further research.

Bibliography

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

[2] Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Alternating proximal algorithms for weakly coupled convex minimization problems—applications to dynamical games and PDE's. J. Convex Anal. **15**, 485-506 (2008)

[3] Attouch, H., Briceño-Arias, L.M., Combettes, P.L.: A parallel splitting method for coupled monotone inclusions. SIAM J. Control Optim. **48**, 3246-3270 (2010)

[4] Aujol, J.-F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition - modeling, algorithms, and parameter selection. Int. J. Comput. Vis. **67**, 111-136 (2006)

[5] Baillon, J.-B., Haddad, G.: Quelques propri*été*s des op*é*rateurs angle-born*é*s et n-cycliquement monotones. Israel J. Math. **26**, 137–150 (1977)

[6] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd ed. Springer, New York (2017)

[7] Beckmann, M.J., McGuire, C.B., Winsten, C.B.: Studies in the economics of transportation. Yale University Press, New Haven (1956)

[8] Boţ, R. I., Csetnek, E.R.: An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems. Numer. Algorithms. **71**, 519-540 (2016)

[9] Boţ, R. I., Csetnek, E.R., Heinrich, A., Hendrich, C.: On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. Math. Program. **150**, 251-279 (2015)

[10] Boţ, R. I., Hendrich, C.: Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. J. Math. Imaging Vision. **49**, 551-568 (2014)

[11] Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential equations. Springer, New York (2010)

[12] Briceño-Arias, L.M.: Forward-Douglas–Rachford splitting and forward partial inverse method for solving monotone inclusions. Optimization: A Journal of Mathematical Programming and Operations Research. **64**(5), 1239-1261 (2013)

[13] Briceño-Arias, L.M.: Forward–Partial Inverse–Forward Splitting for Solving Monotone Inclusions. J. Optim. Theory Appl. **166**(2), 391-413 (2015)

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

[15] Briceño-Arias, L.M., Davis, D.: Forward-backward-half forward algorithm for solving monotone inclusions. SIAM J. Optim. **28**(4), 2839-2871 (2018)

[16] Briceño-Arias, L.M., Deride, J., López-Rivera, S.: Primal-dual partial inverse splitting for constrained monotone inclusions. arXiv: 2007.01983

[17] Briceño-Arias, L.M., Kalise, D., Kobeissi, Z., Laurière, M., González, A.M., Silva, F.J.: On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings. ESAIM: Proceedings and Surveys. **65**, 330-348 (2019)

[18] Briceño-Arias, L.M., Kalise, D., Silva, F.J.: Proximal methods for stationary Mean Field Games with local couplings. SIAM J. Control Optim. **56**(2), 801-836 (2018)

[19] Briceño-Arias, L.M., López Rivera, S.: A projected primal-dual method for solving constrained monotone inclusions. J. Optim. Theory Appl. **180**(3), 907-924 (2019)

[20] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120-145 (2011)

[21] Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**, 253-287 (2016)

[22] Chen, X., Pong, T.K., Wets, R. J.-B.: Two-stage stochastic variational inequalities: an erm-solution procedure. Mathematical Programming. **165**(1), 71–111 (2017)

[23] Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**, 167–188 (1997)

[24] Combettes, P.L.: Quasi-Fejérian analysis of some optimization algorithms. In Inherently parallel algorithms in feasibility and optimization and their applications. Stud. Comput. Math. **8**, 115-152 (2001)

[25] Combettes, P. L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization. **53**, 475-504 (2004)

[26] Combettes, P. L., Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued Var. Anal. **20**, 307-330 (2012)

[27] Cominetti, R., Mascarenhas, W.F., Silva, P.J.S.: A Newton's method for the continuous quadratic knapsack problem. Mathematical Programming. **6**(2), 151–169 (2014)

[28] Cominetti, R. , Peypouquet, J., Sorin, S.: Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization. J. Differ. Equations **245**, 3753-3763 (2008)

[29] Condat, L.: A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**, 460-479 (2013)

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

[31] Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**, 1015-1046 (2010)

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

[33] Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M.,

Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary Value Problems, 299-331. North-Holland, Amsterdam (1983)

[34] Gaines, B. R., Kim, J., Zhou, H.: Algorithms for fitting the constrained lasso. Journal of Computational and Graphical Statistics. **27**(4), 861-871 (2018)

[35] He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. **5**, 119-149 (2012)

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

[37] Lorenz, D.A., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vision. **51**, 311-325 (2015)

[38] Malitsky, Y., Pock, T.: A first-order primal-dual algorithm with linesearch. SIAM J. Optim. **28**, 411-432 (2018)

[39] Martinet, B.: Regularisation d'indquations variationelles par approximations successives. Rev. Francaise Inf. Rech. Oper. 154-159 (1970)

[40] Mercier, B.: Topics in Finite Element Solution of Elliptic Problems (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay (1979)

[41] Mercier B.: Inéquations Variationnelles de la Mécanique (Publications Mathématiques d'Orsay, no. 80.01). Orsay, France: Université de Paris-XI (1980)

[42] Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Amer. Math. Soc. **73**, 591–597 (1967)

[43] Peypouquet, J., Sorin, S.: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. J. Convex Anal. **17**, 1113-1163 (2010)

[44] Rockafellar, R. T.: Monotone Operators and the Proximal Point Algorithm. SIAM Journal on Control and Optimization. **14**(5), 877–898 (1976)

[45] Sheffi, Y.: Urban transportation networks: equilibrium analysis with mathematical programming methods. Englewood Cliffs (NJ): Prentice-Hall, (1985)

[46] Spingarn J.E.: Partial inverse of a monotone operator. Appl. Math. Optim. **10**, 247-265 (1983)

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

[48] Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**, 667-681 (2011)