Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-018-1430-2

A Projected Primal–Dual Method for Solving Constrained Monotone Inclusions

Luis Briceño-Arias[1] · Sergio López Rivera[1]

Received: 28 May 2018 / Accepted: 29 October 2018

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.

Keywords Accelerated schemes · Constrained convex optimization · Monotone operator theory · Proximity operator · Splitting algorithms

Mathematics Subject Classification 47H05 · 65K05 · 65K15 · 90C25

Communicated by Juan-Enrique Martínez Legaz.

✉ Luis Briceño-Arias
luis.briceno@usm.cl

Sergio López Rivera
sergio.lopez.12@sansano.usm.cl

1 Departamento de Matemática, Universidad Técnica Federico Santa María, Santiago, Chile

Published online: 01 November 2018

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 relevance of monotone inclusions and convex optimization is justified via the increasing number of applications in several fields of engineering and applied mathematics as image processing, evolution inclusions, variational inequalities, learning, partial differential equations, mean field games, among others (see, e.g., [1–7] and references therein). 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 [8,9]. 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 [10–16] 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., [4]), 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 Sect. 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 [11] for strongly monotone inclusions. A similar approach in the case without a priori information is used in [18] 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 generalizes the algorithms proposed in [10,12] and [16] 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 [12], complementing the ergodic rates obtained in the case without projection in [15] 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 Sect. 2. We propose our algorithm and the main results in Sect. 3, together with connections with existing methods in the literature. In Sect. 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 Sect. 5. We finish with some conclusions in Sect. 6.

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, $J_M = (\mathrm{Id} + M)^{-1}$ for its resolvent, and $\square$ stands for the parallel sum as in [19]. 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 sense 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 [19]. 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 $f \,\square\, g \colon x \mapsto \inf_{y \in \mathcal{H}} \big(f(y) + g(x - y)\big)$ 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 $\partial(f \,\square\, g) = (\partial f) \,\square\, (\partial g)$ [19, 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 [19] for definitions and further results in monotone operator theory and convex optimization.

3 Problem and Main Results

We consider the following problem.

Problem 3.1 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

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

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

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 3.1 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 includes a projection onto X which helps to the performance of the method as stated in Sect. 5.

When $\rho = \chi = 0$, $V = \mathcal{G}$ and $T = \operatorname{Id}$, ($\mathcal{P}_0$)-($\mathcal{D}_0$) can be solved by using [20, Theorem 4.2] or [16, Theorem 5]. In the last method, inertial terms are also included. In the case when $\ell^* = 0$, the algorithm in [10] can be used and if $\ell^* = h = 0$, ($\mathcal{P}_0$)-($\mathcal{D}_0$) can be solved by [12,21] or a version of [12] with line-search proposed in [22]. In [12], 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 [4, Theorem 3.1]. When $\rho > 0$ or $\chi > 0$, ergodic convergence rates are derived in [15] 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 3.1, set $T = \operatorname{Id}$, and, for every $i \in \{1, \ldots, m\}$, 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

$$\begin{aligned} &V = \mathcal{G} = G_1 \oplus \cdots \oplus G_m \\ &L \colon \mathcal{H} \to \mathcal{G} \colon x \mapsto (L_1 x, \ldots, L_m x) \end{aligned}$$

$$B\colon (u_1,\ldots,u_m)\mapsto \omega_1 B_1 u_1\times\cdots\times\omega_m B_m u_m$$
$$D\colon (u_1,\ldots,u_m)\mapsto \omega_1 D_1 u_1\times\cdots\times\omega_m D_m u_m.$$

Then, Problem 3.1 reduces to [20, Problem 1.1] (see also [11, Problem 1.1]). We prefer to set $m=1$ for simplicity. In [20], previous problem is solved when C is monotone and Lipschitz by applying the method in [8] to the product primal–dual space. Accelerated versions of previous algorithm under strong monotonicity are proposed in [23] and a different approach for exploiting strong monotonicity is used in [18] in this case. An inertial version in the previous context is developed in [24]. The cocoercivity of C is exploited in [11], where an algorithm is proposed for solving Problem 3.1 when $\rho=\chi=0$, $V=\mathcal{G}$ and $T=\operatorname{Id}$. In the following theorem, we provide an algorithm for solving Problem 3.1 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) [12, Section 5.1]. Finally, we generalize [12, Section 5.2] for obtaining linear convergence when $\rho>0$ and $\chi>0$.

Theorem 3.1 *Let $\gamma_0\in]0,2\delta[$ and $\tau_0\in]0,2\beta[$ be such that*

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

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

$$(\forall k\in\mathbb{N})\quad\left\lfloor\begin{aligned}\eta^{k+1}&=J_{\gamma_k B^{-1}}(u^k+\gamma_k(L\bar{x}^k-D^{-1}u^k))\\u^{k+1}&=P_V\,\eta^{k+1}\\p^{k+1}&=J_{\tau_k A}(x^k-\tau_k(L^*u^{k+1}+Cx^k))\\x^{k+1}&=T\,p^{k+1}\\\bar{x}^{k+1}&=x^{k+1}+\theta_k(p^{k+1}-x^k).\end{aligned}\right.\tag{2}$$

Then, the following hold.

1. *For every $k\geq 1$ and for every solution $(\hat{x},\hat{u})$ to Problem 3.1, we have*

$$\begin{aligned}\frac{\|x^k-\hat{x}\|^2}{\tau_k}+\frac{\|u^k-\hat{u}\|^2}{\gamma_k}&\geq(2\rho\tau_k+1)\frac{\|p^{k+1}-\hat{x}\|^2}{\tau_k}+\left\|p^{k+1}-x^k\right\|^2\left(\frac{1}{\tau_k}-\frac{1}{2\beta}\right)\\&+(2\chi\gamma_k+1)\frac{\|\eta^{k+1}-\hat{u}\|^2}{\gamma_k}+\|\eta^{k+1}-u^k\|^2\left(\frac{1}{\gamma_k}-\frac{1}{2\delta}\right)\\&+2\left\langle L(p^{k+1}-x^k)\mid\eta^{k+1}-\hat{u}\right\rangle-2\theta_{k-1}\left\langle L(p^k-x^{k-1})\mid\eta^k-\hat{u}\right\rangle\\&-2\theta_{k-1}\|L\|\|p^k-x^{k-1}\|\|\eta^{k+1}-u^k\|.\end{aligned}\tag{3}$$

2. *Suppose that $(\rho,\chi)=(0,0)$, set $\theta_k\equiv 1$, $\tau_k\equiv\tau$, $\gamma_k\equiv\gamma$ and assume that (1) holds with strict inequality. Then, there exists a solution $(\bar{x},\bar{u})$ to Problem 3.1 such that $x^k\rightharpoonup\bar{x}$ and $u^k\rightharpoonup\bar{u}$.*

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

$$(\forall k \in \mathbb{N}) \quad \theta_k = \frac{1}{\sqrt{1 + 2\rho\tau_k}}, \quad \tau_{k+1} = \theta_k \tau_k, \quad \gamma_{k+1} = \gamma_k/\theta_k, \tag{4}$$

and we assume that (1) *holds with equality, we obtain, for every solution* $(\hat{x}, \hat{u})$ *to Problem 3.1,* $(\forall \varepsilon > 0)(\exists N_0 \in \mathbb{N})(\forall k \geq N_0)$

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

4. *Suppose that* $\rho > 0$ *and* $\chi > 0$ *and define*

$$\mu = \frac{2\sqrt{\rho\chi}}{\|L\|} \quad \textit{and} \quad \alpha = \min\left\{ \frac{\mu\rho}{\rho + \frac{\mu}{4\beta}}, \frac{\mu\chi}{\chi + \frac{\mu}{4\delta}} \right\}. \tag{5}$$

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

$$\tau = \frac{2\beta\mu}{\mu + 4\beta\rho} \quad \textit{and} \quad \gamma = \frac{2\mu\delta}{\mu + 4\delta\chi}, \tag{6}$$

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

Proof 1: Fix $k \in \mathbb{N}$ and let $(\hat{x}, \hat{u})$ be a solution to Problem 3.1. 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) that

$$\begin{aligned} &\frac{x^k - p^{k+1}}{\tau_k} - L^* u^{k+1} - C x^k \in A p^{k+1} \\ &\frac{u^k - \eta^{k+1}}{\gamma_k} + L\bar{x}^k - D^{-1} u^k \in B^{-1} \eta^{k+1}. \end{aligned} \tag{7}$$

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

$$\left\langle \frac{x^k - p^{k+1}}{\tau_k} - L^*(u^{k+1} - \hat{u}) \,\middle|\, p^{k+1} - \hat{x} \right\rangle + \left\langle \frac{u^k - \eta^{k+1}}{\gamma_k} + L(\bar{x}^k - \hat{x}) \,\middle|\, \eta^{k+1} - \hat{u} \right\rangle$$
$$-\left\langle C x^k - C\hat{x} \mid p^{k+1} - \hat{x} \right\rangle - \left\langle D^{-1} u^k - D^{-1}\hat{u} \mid \eta^{k+1} - \hat{u} \right\rangle$$

$$\geq \rho \left\| p^{k+1} - \hat{x} \right\|^2 + \chi \|\eta^{k+1} - \hat{u}\|^2. \tag{8}$$

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, $\langle D^{-1}u^k - D^{-1}\hat{u} \mid \eta^{k+1} - \hat{u} \rangle \geq -\frac{\|\eta^{k+1} - u^k\|^2}{4\delta}$. Hence, by using [19, Lemma 2.12(i)] in (8), we deduce

$$\begin{aligned}\frac{\|x^k - \hat{x}\|^2}{\tau_k} + \frac{\|u^k - \hat{u}\|^2}{\gamma_k} &\geq \left(2\rho + \frac{1}{\tau_k}\right) \left\| p^{k+1} - \hat{x} \right\|^2 + \left(2\chi + \frac{1}{\gamma_k}\right) \|\eta^{k+1} - \hat{u}\|^2 \\ &+ 2\left[\left\langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u} \right\rangle - \left\langle L(\bar{x}^k - \hat{x}) \mid \eta^{k+1} - \hat{u} \right\rangle\right] \\ &+ \|\eta^{k+1} - u^k\|^2 \left(\frac{1}{\gamma_k} - \frac{1}{2\delta}\right) + \|p^{k+1} - x^k\|^2 \left(\frac{1}{\tau_k} - \frac{1}{2\beta}\right).\end{aligned} \tag{9}$$

Moreover, (2), $\operatorname{ran}(L) \subset V$ and $(u^k - \eta^k)_{k \geq 1} \subset V^\perp$ yield, for every $k \geq 1$,

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

which, together with (9), yield (3).

2: For every $k \in \mathbb{N}$, it follows from Theorem 3.1(1), $\rho = \chi = 0$, $\theta_k \equiv 1$, $\tau_k \equiv \tau$, $\gamma_k \equiv \gamma$

$$\frac{\|p^k - \hat{x}\|^2}{\tau} + \frac{\|\eta^k - \hat{u}\|^2}{\gamma} \geq \frac{\|p^{k+1} - \hat{x}\|^2}{\tau} + \left(\frac{1 - \alpha}{\alpha}\right) \frac{\|x^k - p^k\|^2}{\tau} + \frac{\|u^k - \eta^k\|^2}{\gamma}$$

$$
\begin{aligned}
&+ \frac{\|\eta^{k+1}-\hat{u}\|^2}{\gamma} + \|p^{k+1}-x^k\|^2\left(\frac{1}{\tau}-\frac{1}{2\beta}\right) + \|\eta^{k+1}-u^k\|^2\left(\frac{1}{\gamma}-\frac{1}{2\delta}\right)\\
&+ 2\Big\langle L(p^{k+1}-x^k) \mid \eta^{k+1}-\hat{u}\Big\rangle - 2\Big\langle L(p^k-x^{k-1}) \mid \eta^k-\hat{u}\Big\rangle\\
&- 2\|L\|\|p^k-x^{k-1}\|\|\eta^{k+1}-u^k\|\\
\geq\ & \frac{\|p^{k+1}-\hat{x}\|^2}{\tau} + \frac{\|u^k-\eta^k\|^2}{\gamma} + \|\eta^{k+1}-u^k\|^2\left(\frac{1}{\gamma}-\frac{1}{2\delta}-\frac{1}{\nu}\right)\\
&+ \left(\frac{1-\alpha}{\alpha}\right)\frac{\|x^k-p^k\|^2}{\tau} + \frac{\|\eta^{k+1}-\hat{u}\|^2}{\gamma} + \|p^{k+1}-x^k\|^2\left(\frac{1}{\tau}-\frac{1}{2\beta}\right)\\
&+ 2\Big\langle L(p^{k+1}-x^k) \mid \eta^{k+1}-\hat{u}\Big\rangle - 2\Big\langle L(p^k-x^{k-1}) \mid \eta^k-\hat{u}\Big\rangle\\
&- \nu\|L\|^2\|p^k-x^{k-1}\|^2, \qquad (10)
\end{aligned}
$$

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 (10) we have

$$
\begin{aligned}
\Upsilon_k + \frac{\|p^k-\hat{x}\|^2}{\tau} &\geq \Upsilon_{k+1} + \frac{\|p^{k+1}-\hat{x}\|^2}{\tau} + \left(\frac{1-\alpha}{\alpha}\right)\frac{\|x^k-p^k\|^2}{\tau} + \frac{\|u^k-\eta^k\|^2}{\gamma}\\
&+ \varepsilon\|\eta^{k+1}-u^k\|^2 + \nu\varepsilon\left(\frac{1}{\tau}-\frac{1}{2\beta}\right)\|p^k-x^{k-1}\|^2, \qquad (11)
\end{aligned}
$$

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

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

Note that from (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 (11) we deduce that $(\Upsilon_k + \|p^k-\hat{x}\|^2/\tau)_{k\in\mathbb{N}}$ is a Féjer sequence. We deduce from [19, Lemma 5.31] that $(\eta^k)_{k\in\mathbb{N}}$ and $(p^k)_{k\in\mathbb{N}}$ are bounded,

$$
x^k - p^k \to 0, \quad u^k - \eta^k \to 0, \quad \eta^{k+1} - u^k \to 0, \quad \text{and} \quad p^k - x^{k-1} \to 0. \qquad (12)
$$

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 (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 [19, Example 20.29] and, therefore, they have weak–strong closed graphs [19, Proposition 20.38]. Hence, it follows from (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, (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 [21, Proposition 2.7(iii)], $\boldsymbol{Q}\colon(p,\eta)\mapsto(Cp,D^{-1}\eta)$ is $\min\{\beta,\delta\}$−cocoercive, and

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

It follows from [19, Corollary 25.5] that $\boldsymbol{M}+\boldsymbol{Q}$ is maximally monotone, and since (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 3.1, and the result follows.

3: Fix $k\in\mathbb{N}$. Since $\rho>0$, $\delta=+\infty$, $\chi=0$, we obtain from Theorem 3.1(1)

$$\begin{aligned}
\frac{\|x^k-\hat{x}\|^2}{\tau_k}+\frac{\|u^k-\hat{u}\|^2}{\gamma_k} &\geq (2\rho\tau_k+1)\frac{\tau_{k+1}}{\tau_k}\frac{\|p^{k+1}-\hat{x}\|^2}{\tau_{k+1}}+\frac{\gamma_{k+1}}{\gamma_k}\frac{\|\eta^{k+1}-\hat{u}\|^2}{\gamma_{k+1}}\\
&+2\left\langle L(p^{k+1}-x^k)\mid \eta^{k+1}-\hat{u}\right\rangle-2\theta_{k-1}\left\langle L(p^k-x^{k-1})\mid\eta^k-\hat{u}\right\rangle\\
&+\left\|p^{k+1}-x^k\right\|^2\left(\frac{1}{\tau_k}-\frac{1}{2\beta}\right)-\theta_{k-1}^2\gamma_k\|L\|^2\|p^k-x^{k-1}\|^2,
\end{aligned}\tag{13}$$

where we use $2ab\leq a^2/\gamma+\gamma b^2$. Moreover, it follows from (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 (13), yields

$$\begin{aligned}
\frac{\|x^k-\hat{x}\|^2}{\tau_k}+\frac{\|u^k-\hat{u}\|^2}{\gamma_k} &\geq \frac{1}{\theta_k}\left(\frac{\|p^{k+1}-\hat{x}\|^2}{\tau_{k+1}}+\frac{\|\eta^{k+1}-\hat{u}\|^2}{\gamma_{k+1}}\right)\\
&+2\left\langle L(p^{k+1}-x^k)\mid\eta^{k+1}-\hat{u}\right\rangle-2\theta_{k-1}\left\langle L(p^k-x^{k-1})\mid\eta^k-\hat{u}\right\rangle\\
&+\left\|p^{k+1}-x^k\right\|^2\left(\frac{1}{\tau_k}-\frac{1}{2\beta}\right)-\theta_{k-1}^2\gamma_k\|L\|^2\|p^k-x^{k-1}\|^2.
\end{aligned}\tag{14}$$

Now define

$$(\forall k \in \mathbb{N}) \quad \Delta_k = \frac{\|x^k - \hat{x}\|^2}{\tau_k} + \frac{\|u^k - \hat{u}\|^2}{\gamma_k}. \tag{15}$$

Dividing (14) by τ_k and using $\theta_k \tau_k = \tau_{k+1}$, we obtain from the nonexpansivity of P_V and T that

$$\begin{aligned}\frac{\Delta_k}{\tau_k} \geq \frac{\Delta_{k+1}}{\tau_{k+1}} + \frac{2}{\tau_k}\left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - \frac{2}{\tau_{k-1}}\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle \\ + \frac{\|p^{k+1} - x^k\|^2}{\tau_k^2}\left(1 - \frac{\tau_k}{2\beta}\right) - \gamma_k \tau_k \|L\|^2 \frac{\|p^k - x^{k-1}\|^2}{\tau_{k-1}^2}.\end{aligned} \tag{16}$$

In addition, (1) with equality reduces to

$$\|L\|^2 = \left(\frac{1}{\tau_0} - \frac{1}{2\beta}\right)\frac{1}{\gamma_0} \quad \Leftrightarrow \quad \gamma_0 \tau_0 \|L\|^2 = \left(1 - \frac{\tau_0}{2\beta}\right). \tag{17}$$

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 (4)), we have from (17) that

$$\gamma_k \tau_k \|L\|^2 = \gamma_0 \tau_0 \|L\|^2 = \left(1 - \frac{\tau_0}{2\beta}\right) \leq \left(1 - \frac{\tau_{k-1}}{2\beta}\right), \tag{18}$$

and (16) yields

$$\begin{aligned}\frac{\Delta_k}{\tau_k} \geq \frac{\Delta_{k+1}}{\tau_{k+1}} + \frac{\|p^{k+1} - x^k\|^2}{\tau_k^2}\left(1 - \frac{\tau_k}{2\beta}\right) - \frac{\|p^k - x^{k-1}\|^2}{\tau_{k-1}^2}\left(1 - \frac{\tau_{k-1}}{2\beta}\right) \\ + \frac{2}{\tau_k}\left\langle L(p^{k+1} - x^k) \mid \eta^{k+1} - \hat{u}\right\rangle - \frac{2}{\tau_{k-1}}\left\langle L(p^k - x^{k-1}) \mid \eta^k - \hat{u}\right\rangle.\end{aligned} \tag{19}$$

Now fix $N \geq 1$. By adding from $k = 0$ to $k = N - 1$ in (19), 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

$$\begin{aligned}\frac{\Delta_0}{\tau_0} &\geq \frac{\Delta_N}{\tau_N} + \frac{\|p^N - x^{N-1}\|^2}{\tau_{N-1}^2}\left(1 - \frac{\tau_{N-1}}{2\beta}\right) + \frac{2}{\tau_{N-1}}\left\langle L(p^N - x^{N-1}) \mid u^N - \hat{u}\right\rangle \\ &\geq \frac{\Delta_N}{\tau_N} - \frac{\|L\|^2}{(1 - \frac{\tau_{N-1}}{2\beta})}\left\|u^N - \hat{u}\right\|^2 \\ &= \frac{1}{\tau_N}\left(\Delta_N - \frac{\gamma_N \tau_N \|L\|^2}{\left(1 - \frac{\tau_{N-1}}{2\beta}\right)}\frac{\|u^N - \hat{u}\|^2}{\gamma_N}\right) \geq \frac{\|x^N - \hat{x}\|^2}{\tau_N^2},\end{aligned} \tag{20}$$

where the last inequality follows from (18) and (15). Multiplying (20) by τ_N^2 and using (17), we conclude that

$$\left\|x^N - \hat{x}\right\|^2 \leq \tau_N^2 \left(\frac{\left\|x^0 - \hat{x}\right\|^2}{\tau_0^2} + \frac{\|L\|^2}{\left(1 - \frac{\tau_0}{2\beta}\right)} \left\|u^0 - \hat{u}\right\|^2 \right). \tag{21}$$

The result follows from $\lim_{N\to\infty} N\rho\tau_N = 1$ [12, Corollary 1].

4: Fix $k \in \mathbb{N}$. Note that (6) yields $\left(\frac{1}{\tau} - \frac{1}{2\beta}\right)\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right) = \|L\|^2$. From (9), $u^{k+1} = P_V \eta^{k+1}$ and $\operatorname{ran} L \subset V$, we have

$$\begin{aligned}
\frac{\left\|u^k - \hat{u}\right\|^2}{2\gamma} + \frac{\left\|x^k - \hat{x}\right\|^2}{2\tau} &\geq (2\rho\tau + 1)\frac{\|p^{k+1} - \hat{x}\|^2}{2\tau} + (2\chi\gamma + 1)\frac{\|\eta^{k+1} - \hat{u}\|^2}{2\gamma} \\
&\quad + \frac{\left\|p^{k+1} - x^k\right\|^2}{2}\left(\frac{1}{\tau} - \frac{1}{2\beta}\right) + \frac{\|\eta^{k+1} - u^k\|^2}{2}\left(\frac{1}{\gamma} - \frac{1}{2\delta}\right) \\
&\quad + \left\langle L(p^{k+1} - \overline{x}^k) \mid \eta^{k+1} - \hat{u}\right\rangle.
\end{aligned} \tag{22}$$

Hence, by defining

$$(\forall k \in \mathbb{N}) \quad \Omega_k := \left(\chi + \frac{\mu}{4\delta}\right)\|u^k - \hat{u}\|^2 + \left(\rho + \frac{\mu}{4\beta}\right)\|x^k - \hat{x}\|^2, \tag{23}$$

multiplying (22) by μ and using (5), (6), $u^{k+1} = P_V \eta^{k+1}$, $\operatorname{ran} L \subset V$, and the nonexpansivity of T and P_V, we have

$$\begin{aligned}
\Omega_k &\geq \Omega_{k+1} + \mu\rho\left\|p^{k+1} - \hat{x}\right\|^2 + \mu\chi\left\|\eta^{k+1} - \hat{u}\right\|^2 + \rho\left\|p^{k+1} - x^k\right\|^2 \\
&\quad + \chi\left\|\eta^{k+1} - u^k\right\|^2 + \mu\left\langle L(p^{k+1} - \overline{x}^k) \mid \eta^{k+1} - \hat{u}\right\rangle \\
&\geq (1 + \alpha)\Omega_{k+1} + \rho\left\|p^{k+1} - x^k\right\|^2 + \chi\left\|u^{k+1} - u^k\right\|^2 \\
&\quad + \mu\left\langle L(p^{k+1} - \overline{x}^k) \mid u^{k+1} - \hat{u}\right\rangle.
\end{aligned} \tag{24}$$

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

$$\begin{aligned}
&\mu\left\langle L(p^{k+1} - \overline{x}^k) \mid u^{k+1} - \hat{u}\right\rangle \\
&\quad = \mu\left\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\right\rangle - \mu\theta\left\langle L(p^k - x^{k-1}) \mid u^{k+1} - \hat{u}\right\rangle \\
&\quad = \mu\left\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\right\rangle - \omega\mu\left\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\right\rangle \\
&\qquad - \omega\mu\left\langle L(p^k - x^{k-1}) \mid u^{k+1} - u^k\right\rangle \\
&\qquad - (\theta - \omega)\mu\left\langle L(p^k - x^{k-1}) \mid u^{k+1} - \hat{u}\right\rangle \\
&\quad \geq \mu\left\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\right\rangle - \omega\mu\left\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\right\rangle
\end{aligned}$$

$$
\begin{aligned}
&- \omega\mu \|L\| \left(\frac{\lambda \left\| p^k - x^{k-1} \right\|^2}{2} + \frac{\left\| u^{k+1} - u^k \right\|^2}{2\lambda} \right) \\
&- (\theta - \omega)\mu \|L\| \left(\frac{\lambda \left\| p^k - x^{k-1} \right\|^2}{2} + \frac{\left\| u^{k+1} - \hat{u} \right\|^2}{2\lambda} \right) \\
= {}& \mu \left\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u} \right\rangle - \omega\mu \left\langle L(p^k - x^{k-1}) \mid u^k - \hat{u} \right\rangle \\
&- \mu\theta\lambda \|L\| \frac{\left\| p^k - x^{k-1} \right\|^2}{2} - \frac{\omega\mu \|L\| \left\| u^{k+1} - u^k \right\|^2}{2\lambda} \\
&- (\theta - \omega)\mu \|L\| \frac{\left\| u^{k+1} - \hat{u} \right\|^2}{2\lambda}.
\end{aligned} \tag{25}
$$

By choosing $\lambda = \omega\sqrt{\frac{\rho}{\chi}}$, from (24), (25) and (5), we obtain

$$
\begin{aligned}
\Omega_k \geq {}& \frac{\Omega_{k+1}}{\omega} + \left(1 + \alpha - \frac{1}{\omega}\right) \Omega_{k+1} + \rho \left\| p^{k+1} - x^k \right\|^2 \\
&+ \mu \left\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u} \right\rangle - \omega\mu \left\langle L(p^k - x^{k-1}) \mid u^k - \hat{u} \right\rangle \\
&- \omega\theta\rho \left\| p^k - x^{k-1} \right\|^2 - \left(\frac{\theta - \omega}{\omega} \right) \chi \left\| u^{k+1} - \hat{u} \right\|^2.
\end{aligned} \tag{26}
$$

Since $\theta \in](1+\alpha)^{-1}, 1]$, by setting $\omega = \frac{1+\theta}{2+\alpha} \in](1+\alpha)^{-1}, \theta[$, we have $1+\alpha-\frac{1}{\omega} = \frac{\theta - \omega}{\omega} > 0$. Hence, since (23) yields $\Omega_{k+1} \geq \chi \left\| u^{k+1} - \hat{u} \right\|^2$, from (26) and $\theta \leq 1$ we have

$$
\begin{aligned}
\Omega_k \geq {}& \frac{\Omega_{k+1}}{\omega} + \rho \left\| p^{k+1} - x^k \right\|^2 - \omega\rho \left\| p^k - x^{k-1} \right\|^2 \\
&+ \mu \left\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u} \right\rangle - \omega\mu \left\langle L(p^k - x^{k-1}) \mid u^k - \hat{u} \right\rangle.
\end{aligned} \tag{27}
$$

Moreover, using $p^0 = x^0 =: x^{-1}$, multiplying (27) by ω^{-k} and adding from $k = 0$ to $k = N - 1$, we conclude from the definition of μ that

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

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 3.1 1. Note that condition (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.

2. 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 [18]. The choice of step sizes in [18] coincides when $C=0$, but the sequence $(\theta_k)_{k\in\mathbb{N}}$ is strictly larger than 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 [18], leading to a better estimation when C is cocoercive.
3. Condition (1) is weaker than the condition needed in [11]. 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 (1). Our condition is strictly weaker, as it can be seen in Fig. 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

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

4. 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

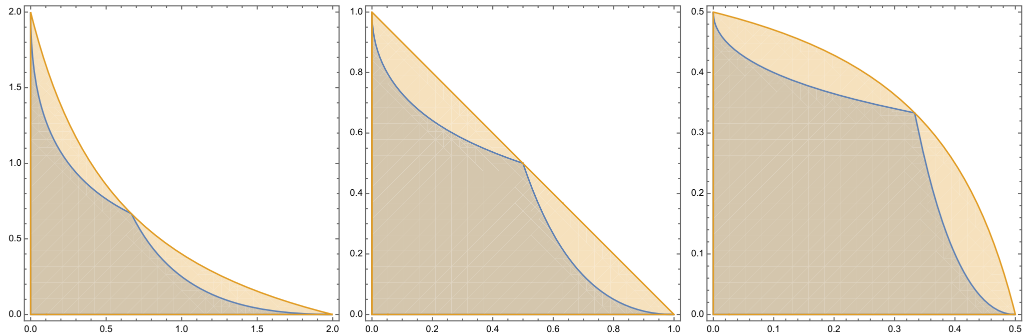

Fig. 1 We plot regions R_b in blue and S_b in orange. Left: case $b = 1$, Center: case $b = 1/2$, Right: case $b = 1/4$. Note that in the case $\tau = \gamma$ the regions coincide

[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 $M : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximally monotone and $N : \mathcal{H} \to \mathcal{H}$ 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 3.1 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.

5. The method proposed in [23] is an accelerated version of the method proposed in [20] 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 3.1 simpler.
6. Theorem 3.1(3) generalizes the acceleration scheme proposed in [12] to monotone inclusions with a priori information, and we obtain an accelerated version of the methods in [11] 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.
7. In the context of primal–dual problem ($\mathcal{P}_0$)-($\mathcal{D}_0$), (2) reduces to

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \eta^{k+1} = \operatorname{prox}_{\gamma_k g^*}(u^k + \gamma_k(L\bar{x}^k - \nabla \ell^*(u^k))) \\ u^{k+1} = P_V\, \eta^{k+1} \\ p^{k+1} = \operatorname{prox}_{\tau_k f}\left(x^k - \tau_k\left(L^* u^{k+1} + \nabla h(x^k)\right)\right) \\ x^{k+1} = T\, p^{k+1} \\ \bar{x}^{k+1} = x^{k+1} + \theta_k(p^{k+1} - x^k), \end{array} \right. \tag{28}
$$

and our conditions on the parameters coincide with [15,16]. Without strong convexity of f and g^*, we deduce from Theorem 3.1(2) the weak convergence of the sequences generated by (28), generalizing results in [4,10,16]. When f or g^* is strongly convex, Theorem 3.1(3) yields an accelerated and projected version of [10]. When $V = \mathcal{G}$ and $T = \operatorname{Id}$, this result complements the ergodic convergence rates obtained in [15] and generalizes [12]. When $\ell^* = 0$, $V = \mathcal{G}$, $T = \operatorname{Id}$, and f

and g^* are strongly convex, Theorem 3.1(4) yields non-ergodic linear convergence of [10], complementing the ergodic linear convergence in [15]. The advantage of the algorithm (28) with respect to [10,12] 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 [4] and in Sect. 5, in which we provide some numerical simulations.

4 Application to Constrained Convex Optimization

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

Problem 4.1 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

$$\min_{x \in \mathbb{R}^N} f(x) \quad \text{s.t.} \quad Rx = c \quad Sx = d, \tag{29}$$

under the assumption that solutions exist.

Note that (29) 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 $\operatorname{prox}_{\gamma\iota^*_{\{b\}}} = \operatorname{Id} - \gamma b$ [19, Proposition 24.8(ix)], the method proposed in [12, Algorithm 1] in this case reads: given $x^0 = \bar{x}^0 \in \mathcal{H}$ and $u^0 \in \mathcal{G}$,

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = u^k + \gamma(L\bar{x}^k - b) \\ x^{k+1} = \operatorname{prox}_{\tau f}(x^k - \tau L^* u^{k+1}) \\ \bar{x}^{k+1} = 2x^{k+1} - x^k, \end{array} \right. \tag{30}$$

where $\gamma\tau\|L\|^2 < 1$. The constraint is imposed via the Lagrange multiplier update in the first step of (30). 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 [4,17]). If it is difficult to compute $P_{L^{-1}b}$ but we can project onto $R^{-1}c$, we can rewrite (29) 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 3.1, (28) and $P_{R^{-1}c}\colon x \mapsto x - R^*(RR^*)^{-1}(Rx - c)$, when RR^* is invertible.

Corollary 4.1 *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*

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

$m = 1, n = 100$	$e = 10^{-4}$		$e = 5 \times 10^{-5}$		$e = 10^{-5}$	
	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

Table 2 Average time and number of iterations when $m = 10$ for obtaining $r_k < e$

$m = 10, n = 100$	$e = 10^{-4}$		$e = 5 \times 10^{-5}$		$e = 10^{-5}$	
	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

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} u^{k+1} & = u^k + \gamma(L\bar{x}^k - b) \\ p^{k+1} & = \mathrm{prox}_{\tau f}(x^k - \tau L^* u^{k+1}) \\ x^{k+1} & = p^{k+1} - R^*(RR^*)^{-1}(Rp^{k+1} - c) \\ \bar{x}^{k+1} & = x^{k+1} + p^{k+1} - x^k. \end{array} \right. \tag{31}$$

Then, there exists a solution $\hat{x}$ to Problem 4.1 and an associated multiplier $\hat{u}$ such that $x^k \to \hat{x}$ and $u^k \to \hat{u}$.

5 Numerical Experiences

In this section, we consider some particular instances of Problem 4.1. 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 (31) 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 (31) and (30), respectively.

Test 1 In Problem 4.1, Table 1 shows the efficiency of CP and PCP for the case $m = 1$ and $n = 100$. 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 $100 \cdot (x_{\mathrm{CP}} - x_{\mathrm{PCP}})/x_{\mathrm{CP}}$, slightly increases.

Test 2 In Problem 4.1, Table 2 shows the efficiency of CP and PCP for the case $m = 10$ and $n = 100$. In this case, there are clear differences between both algorithms

Table 3 Average time and number of iterations when $m = 30$ for obtaining $r_k < e$

$m = 30, n = 100$	$e = 10^{-4}$		$e = 5 \times 10^{-5}$		$e = 10^{-5}$	
	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

Table 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
$e = 5 \times 10^{-5}$	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

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 4.1, Table 3 shows the efficiency of CP and PCP for the case $m = 30$ and $n = 100$. 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 4 summarizes the percentage of improvements for each test. 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.

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 [12].

Acknowledgements The authors thank the two anonymous referees which significantly helped to improve the quality of this manuscript. In addition, the authors thank the “Programa de financiamiento basal” from CMM–Universidad de Chile and the project DGIP-UTFSM PI-M-18.14 from Universidad Técnica Federico Santa María.

References

1. 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)
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 strongly convergent primal–dual method for nonoverlapping domain decomposition. Numer. Math. **133**, 433–470 (2016)
4. 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)
5. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
6. 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, pp. 299–331. North-Holland, Amsterdam (1983)
7. Mercier, B.: Topics in Finite Element Solution of Elliptic Problems. Lectures on Mathematics, vol. 63. Tata Institute of Fundamental Research, Bombay (1979)
8. Tseng, P.: A modified forward–backward splitting method for maximal monotone mappings. SIAM J. Control Optim. **38**, 431–446 (2000)
9. Briceño-Arias, L.M., Davis, D.: Forward-backward-half forward algorithm for solving monotone inclusions. SIAM J. Optim. **28**, 2839–2871 (2018)
10. 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)
11. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**, 667–681 (2013)
12. 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)
13. 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)
14. 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)
15. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal–dual algorithm. Math. Program. **159**, 253–287 (2016)
16. Lorenz, D.A., Pock, T.: An inertial forward–backward algorithm for monotone inclusions. J. Math. Imaging Vis. **51**, 311–325 (2015)
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. arXiv:1802.07902
18. 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)
19. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer, New York (2017)
20. 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)
21. 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)
22. Malitsky, Y., Pock, T.: A first-order primal–dual algorithm with linesearch. SIAM J. Optim. **28**, 411–432 (2018)
23. Boţ, R.I., Hendrich, C.: Convergence analysis for a primal–dual monotone + skew splitting algorithm with applications to total variation minimization. J. Math. Imaging Vis. **49**, 551–568 (2014)
24. 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)
25. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)