# A first-order stochastic primal-dual algorithm with correction step

Lorenzo Rosasco[1,2], Silvia Villa[1] and Bằng Công Vũ[1]

[1] LCSL, Istituto Italiano di Tecnologia and Massachusetts Institute of Technology
Bldg. 46-5155, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[2] DIBRIS, Università di Genova, Via Dodecaneso 35
16146 Genova, Italy
lrosasco@mit.edu; silvia.villa@iit.it; cong.bang@iit.it

November 8, 2017

## Abstract

We investigate the convergence properties of a stochastic primal-dual splitting algorithm for solving structured monotone inclusions involving the sum of a cocoercive operator and a composite monotone operator. The proposed method is the stochastic extension to monotone inclusions of a proximal method studied in [26, 35] for saddle point problems. It consists in a forward step determined by the stochastic evaluation of the cocoercive operator, a backward step in the dual variables involving the resolvent of the monotone operator, and an additional forward step using the stochastic evaluation of the cocoercive introduced in the first step. We prove weak almost sure convergence of the iterates by showing that the primal-dual sequence generated by the method is stochastic quasi Fejér-monotone with respect to the set of zeros of the considered primal and dual inclusions. Additional results on ergodic convergence in expectation are considered for the special case of saddle point models.

**Keywords:** monotone inclusion, maximal monotone operator, operator splitting, cocoercive operator, composite operator, duality, stochastic errors, primal-dual algorithm

**Mathematics Subject Classifications (2010)**: 47H05, 49M29, 49M27, 90C25

## 1 Introduction

This paper is concerned with the algorithmic solution, in a stochastic setting, of structured monotone inclusions defined by the sum of a cocoercive operator and a monotone operator composed with a linear transformation and its adjoint. This problem arises in many applications, such as variational inequalities and equilibrium problems [29], signal and image processing [21, 23], game theory [10], and statistical learning [25, 27, 39, 50, 60]. A necessarily incomplete list of related works include [8, 9, 11, 12, 15, 14, 22, 24, 43, 55].

1

For monotone inclusions with the considered structure, key is the joint solution of the primal problem and its associated dual form. Indeed, primal-dual schemes have several advantages, since they do not require the inversion of any of the involved linear functions, and independently *activate* each of the monotone operators. The single-valued operators is involved in the forward step, while the set-valued operator appears in the backward steps and requires the computation of the resolvent [15, 55].

Recently, stochastic versions of splitting methods for monotone inclusions have been studied. This is relevant to consider practical situations where operators are known only through measurements subject to random noise, or when the computation of a stochastic estimate is cheaper than the evaluation of the operator itself. Among the many approaches, we mention stochastic forward-backward splitting [19, 51, 7], stochastic Douglas-Rachford [19], and stochastic versions of primal-dual methods as in [6, 19, 45, 52]. These works found natural applications in stochastic optimization [19, 51] and machine learning [27, 53].

In this paper we focus on a primal dual splitting method, that generalizes to monotone inclusions the so called proximal alternating predictor-corrector algorithm (PAPC), proposed independently in [35] for regularized least squares minimization problems and in [26] for saddle point problems. In particular, we allow for stochastic errors: the proposed algorithm requires the computation of the resolvent of the maximal monotone operators at each step, and it uses a stochastic approximation of the cocoercive operator. The update of the primal variable requires an extra step (called a correction step in [26]). In this respect, the algorithm resembles the extragradient method proposed in [41], but differs from it since it does not require extra evaluations of the involved operators. Indeed, the correction step does not impact the complexity of the method since it requires only matrix/vector multiplications. The considered algorithm is also related to the general inertial splitting scheme [52, Algorithm 4.41] for solving multivariate monotone inclusions. In contrast to [52, Algorithm 4.41], we allow for an additional projection step on a set of linear constraints, and consider variable step-sizes. Our analysis establishes almost sure weak convergence of the iterates generated by the method, and ergodic convergence under more general conditions on the error when saddle points problems are considered. The analysis differs from the one in [26, 35], and is based on variable metric stochastic quasi-Fejer sequences [61].

The paper is organized as follows: in Section 2 we introduce notation and preliminaries. We then present the problem, the algorithm, and we prove almost sure convergence in Section 3. In Section 4 we focus on minimization problems and their saddle point formulations. For this special case, we prove ergodic convergence in expectation of the duality gap under more general conditions on the errors and the step-sizes. Finally, in Section 5, we consider the case of sum of composite inclusions and we show how to apply the general scheme to this case, using the product space reformulation [9]. As a corollary, we obtain convergence results for structured minimization problems.

**Problem 1.1** Let $\beta \in ]0, +\infty[$, let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $B \colon \mathcal{H} \to \mathcal{H}$ be a $\beta$-cocoercive operator, let $V$ be a closed vector subspace of $\mathcal{H}$, let $A \colon \mathcal{G} \to 2^{\mathcal{G}}$ be a maximally monotone operator, and let $L \colon \mathcal{H} \to \mathcal{G}$ be a bounded linear operator. The normal cone operator to $V$ is denoted by $N_V$. Denote by $\mathcal{P}$ the set of all points $\overline{x} \in \mathcal{H}$ such that

$$0 \in B\overline{x} + L^* A(L\overline{x}) + N_V \overline{x}, \tag{1.1}$$

and $\mathcal{D}$ the set of all $\overline{v} \in \mathcal{G}$ such that

$$0 \in -L(B + N_V)^{-1}(-L^*\overline{v}) + A^{-1}\overline{v}, \tag{1.2}$$

The problem is to find a point $(\overline{x}, \overline{v})$ in $\mathcal{P} \times \mathcal{D}$.

Note that since $V$ is a closed vector subspace,

$$N_V x = \begin{cases} V^\perp & \text{if } x \in V \\ \varnothing & \text{otherwise} . \end{cases}$$

## 2 Notation and preliminary results

Throughout, $\mathcal{H}$ is a real separable Hilbert space. We denote by $\langle \cdot \mid \cdot \rangle$ and $\| \cdot \|$ the scalar product and the associated norm of $\mathcal{H}$. The symbols $\rightharpoonup$ and $\rightarrow$ denote weak and strong convergence, respectively. We denote by $\ell_+^1(\mathbb{N})$ the set of summable sequences in $[0, +\infty[$, and by $\mathcal{B}(\mathcal{H})$ the space of linear operators from $\mathcal{H}$ into itself. Let $U \in \mathcal{B}(\mathcal{H})$ be self-adjoint and strongly positive, i.e.

$$(\exists \chi \in \, ]0, +\infty[)(\forall x \in \mathcal{H}) \quad \langle Ux \mid x \rangle \geq \chi \|x\|^2. \tag{2.1}$$

We define a scalar product and a norm respectively by

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x \mid y \rangle_U = \langle Ux \mid y \rangle \quad \text{and} \quad \|x\|_U = \sqrt{\langle Ux \mid x \rangle}.$$

Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain and the graph of $A$ are defined by

$$\text{dom } A = \{ x \in \mathcal{H} \mid Ax \neq \varnothing \} \text{ and } \text{gra } A = \{ (x, u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax \}.$$

The set of zeros of $A$ is $\text{zer } A = \{ x \in \mathcal{H} \mid 0 \in Ax \}$ and the range of $A$ is $\text{ran } A = A(\mathcal{H})$. The inverse of $A$ is $A^{-1} \colon \mathcal{H} \to 2^{\mathcal{H}} \colon u \mapsto \{ x \in \mathcal{H} \mid u \in Ax \}$. The resolvent of $A$ is

$$J_A = (\text{Id} + A)^{-1}, \tag{2.2}$$

where Id denotes the identity operator of $\mathcal{H}$. Moreover, $A$ is monotone if

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

and maximally so, if there exists no monotone operator $\widetilde{A} \colon \mathcal{H} \to \mathcal{H}$ such that $\text{gra } A \subset \text{gra } \widetilde{A} \neq \text{gra } A$. If $A$ is monotone, then $J_A$ is single-valued and nonexpansive, and, in addition, if $A$ is maximally monotone, then $\text{dom } J_A = \mathcal{H}$.

Let $\Gamma_0(\mathcal{H})$ be the class of proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$. For any self-adjoint strongly positive operator $U \in \mathcal{B}(\mathcal{H})$ and $f \in \Gamma_0(\mathcal{H})$, we define

$$\text{prox}_f^U \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{y \in \mathcal{H}}{\text{argmin}} \, \big( f(y) + \frac{1}{2} \|x - y\|_U^2 \big), \tag{2.3}$$

and

$$\text{prox}_f \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{y \in \mathcal{H}}{\text{argmin}} \, \big( f(y) + \frac{1}{2} \|x - y\|^2 \big).$$

It holds $\text{prox}_f^U = J_{U^{-1} \partial f}$, and $\text{prox}_f = J_{\partial f}$ coincides with the classical definition of proximity operator in [38]. Moreover, let $x \in \mathcal{H}$ and set $p = \text{prox}_f^U x$. Then

$$(\forall y \in \mathcal{H}) \quad f(p) - f(y) \leq \langle y - p \mid U(p - x) \rangle . \tag{2.4}$$

3

The conjugate function of $f$ is

$$f^* \colon a \mapsto \sup_{x \in \mathcal{H}} \big( \langle a \mid x \rangle - f(x) \big).$$

Note that,

$$(\forall f \in \Gamma_0(\mathcal{H}))(\forall (x,y) \in \mathcal{H}^2) \quad y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y),$$

or equivalently,

$$(\forall f \in \Gamma_0(\mathcal{H})) \quad (\partial f)^{-1} = \partial f^*. \tag{2.5}$$

The strong relative interior of a subset $C$ of $\mathcal{H}$, denoted by $\operatorname{sri} C$, is the set of points $x \in C$ such that the cone generated by $-x + C$ is a closed vector subspace of $\mathcal{H}$. We refer to [3] for an account of the main results of convex analysis and monotone operator theory.

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. A $\mathcal{H}$-valued random variable is a measurable (strong and weak measurability coincide since $\mathcal{H}$ is separable) function $X \colon \Omega \to \mathcal{H}$, where $\mathcal{H}$ is endowed with the Borel $\sigma$-algebra. We denote by $\sigma(X)$ the $\sigma$-field generated by $X$. The expectation of a random variable $X$ is denoted by $\mathsf{E}[X]$. The conditional expectation of $X$ given a $\sigma$-field $\mathcal{A} \subset \mathcal{F}$ is denoted by $\mathsf{E}[X|\mathcal{A}]$. Given a random variable $Y \colon \Omega \to \mathcal{H}$, the conditional expectation of $X$ given $Y$ is denoted by $\mathsf{E}[X|Y]$. See [34] for more details on probability theory in Hilbert spaces. A $\mathcal{H}$-valued random process is a sequence $(x_n)_{n \in \mathbb{N}}$ of $\mathcal{H}$-valued random variables. The abbreviation a.s. stands for "almost surely".

**Lemma 2.1** [54, Theorem 1] *Let $(\mathcal{F}_n)_{n \in \mathbb{N}}$ be an increasing sequence of sub-$\sigma$-algebras of $\mathcal{F}$, let $(z_n)_{n \in \mathbb{N}}$, $(\xi_n)_{n \in \mathbb{N}}$, $(\zeta_n)_{n \in \mathbb{N}}$ and $(t_n)_{n \in \mathbb{N}}$ be $[0, +\infty[$-valued random sequences such that, for every $n \in \mathbb{N}$, $z_n$, $\xi_n$, $\zeta_n$, and $t_n$ are $\mathcal{F}_n$-measurable. Assume moreover that $\sum_{n \in \mathbb{N}} t_n < +\infty$, $\sum_{n \in \mathbb{N}} \zeta_n < +\infty$ a.s., and*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}[z_{n+1}|\mathcal{F}_n] \leq (1 + t_n)z_n + \zeta_n - \xi_n \quad a.s.. \tag{2.6}$$

*Then $(z_n)_{n \in \mathbb{N}}$ converges a.s. and $(\xi_n)_{n \in \mathbb{N}}$ is summable a.s..*

The following lemma is a special case of [61, Proposition 2.4].

**Lemma 2.2** *Let $C$ be a non-empty closed subset of $\mathcal{H}$, let $\alpha \in \,]0, \infty[$, let $W \in \mathcal{B}(\mathcal{H})$ and $(W_n)_{n \in \mathbb{N}} \subset \mathcal{B}(\mathcal{H})$ be self-adjoint and strongly positive operators with constant $\alpha$, such that $W_n \to W$ pointwise, and let $(x_n)_{n \in \mathbb{N}}$ be a $\mathcal{H}$-valued random process. Suppose that, for every $x \in C$, there exist $[0, +\infty[$-valued random sequences $(\xi_n(x))_{n \in \mathbb{N}}$, $(\zeta_n(x))_{n \in \mathbb{N}}$ and $(t_n(x))_{n \in \mathbb{N}}$ such that, for every $n \in \mathbb{N}$, $\xi_n(x)$, $\zeta_n(x)$ and $t_n(x)$ are $\mathcal{F}_n$-measurable, $(\zeta_n(x))_{n \in \mathbb{N}}$ and $(t_n(x))_{n \in \mathbb{N}}$ are summable a.s., and*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}[\|x_{n+1} - x\|_{W_{n+1}}|\mathcal{F}_n] \leq (1 + t_n(x))\|x_n - x\|_{W_n} + \zeta_n(x) - \xi_n(x) \qquad a.s. \tag{2.7}$$

*Then the following hold.*

(i) *$(x_n)_{n \in \mathbb{N}}$ is bounded a.s. and $(\xi_n(x))_{n \in \mathbb{N}}$ is summable a.s.*

(ii) *There exists $\widetilde{\Omega} \subset \Omega$ such that $\mathsf{P}(\widetilde{\Omega}) = 1$ and, for every $\omega \in \widetilde{\Omega}$ and $x \in C$, $(\|x_n(\omega) - x\|)_{n \in \mathbb{N}}$ converges.*

4

(iii) *Suppose that the set of weak cluster points of $(x_n)_{n \in \mathbb{N}}$ is a subset of $C$ a.s. Then $(x_n)_{n \in \mathbb{N}}$ converges weakly a.s. to a $C$-valued random vector.*

**Lemma 2.3** [18, Lemma 3.7] *Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $U \in \mathcal{B}(\mathcal{H})$ be self-adjoint and strongly positive, and let $\mathcal{G}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x, y) \mapsto \langle x \mid y \rangle_{U^{-1}}$. Then, the following hold.*

(i) *$UA \colon \mathcal{G} \to 2^{\mathcal{G}}$ is maximally monotone.*

(ii) *$J_{UA} \colon \mathcal{G} \to \mathcal{G}$ is firmly nonexpansive.*

# 3 Algorithm and almost sure convergence

In this section we state our main results. We introduce the extension to monotone inclusions of the primal-dual algorithm in [26, 35], allowing for stochastic errors in the evaluation of the operator $B$ and we prove almost sure weak convergence of the iterates.

**Algorithm 3.1** Let $(\gamma_n)_{n \in \mathbb{N}}$ and $(\tau_n)_{n \in \mathbb{N}}$ be sequences of strictly positive real numbers, let $U$ be a self adjoint positive definite on $\mathcal{G}$, let $(r_n)_{n \in \mathbb{N}}$ be a $\mathcal{H}$-valued, square integrable random process, let $x_0$ be a $\mathcal{H}$-valued, squared integrable random variable and let $v_0$ be a $\mathcal{G}$-valued, squared integrable random variable. Iterate

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} p_n = P_V(x_n - \gamma_n(L^* v_n + r_n)) \\ v_{n+1} = J_{\frac{\tau_n}{\gamma_n} U A^{-1}}\left(v_n + \frac{\tau_n}{\gamma_n} U L p_n\right) \\ x_{n+1} = P_V(x_n - \gamma_n(L^* v_{n+1} + r_n)). \end{array} \right. \tag{3.1}$$

**Almost sure convergence.** We first establish almost sure convergence of the iterates generated by Algorithm (3.1) under suitable conditions on the parameters $(\gamma_n)_{n \in \mathbb{N}}$ and $(\tau_n)_{n \in \mathbb{N}}$ as well as on the stochastic estimates of $B$.

**Theorem 3.2** *In the setting of Problem 1.1, suppose that $\mathcal{P}$ is non empty and consider algorithm 3.1. Assume that the following conditions are satisfied for $\mathcal{F}_n = \sigma((x_k, v_k)_{0 \le k \le n})$*

(i) *$(\gamma_n)_{n \in \mathbb{N}}$ is decreasing and $(\tau_n)_{n \in \mathbb{N}}$ is increasing*

(ii) *$(\exists \tau \in \, ]0, +\infty[)$ such that $\sup_{n \in \mathbb{N}} \tau_n \le \tau$, and $(\tau U)^{-1} - L P_V L^*$ is positive definite.*

(iii) *$\gamma_0 \in \, ]0, \beta[$ and $\inf_{n \in \mathbb{N}} \gamma_n > 0$.*

(iv) *$(\forall n \in \mathbb{N}) \quad \mathsf{E}[r_n | \mathcal{F}_n] = B x_n$.*

(v) *$\sum_{n \in \mathbb{N}} \mathsf{E}[\|r_n - B x_n\|^2 | \mathcal{F}_n] < +\infty \quad \mathsf{P}$-a.s.*

*Then the following hold for some random vector $(x, v)$, $\mathcal{P} \times \mathcal{D}$-valued a.s.*

(i) *$(x_n)_{n \in \mathbb{N}}$ converges weakly to $x$ and $(v_n)_{n \in \mathbb{N}}$ converges weakly to $v$ a.s.*

(ii) $\sum_{n\in\mathbb{N}} \|Bx_n - Bx\|^2 < +\infty$ a.s.

*Proof.* Let $\boldsymbol{S}$ be the set of all points $(\overline{x}, \overline{v}) \in \mathcal{H} \times \mathcal{G}$ such that

$$-L^*\overline{v} \in B\overline{x} + N_V\overline{x} \quad \text{and} \quad L\overline{x} \in A^{-1}\overline{v}. \tag{3.2}$$

Then $\boldsymbol{S}$ is non empty and it is a closed convex subset of $\mathcal{H} \times \mathcal{G}$ [9, Proposition 2.8] contained in $\mathcal{P} \times \mathcal{D}$. Since $A^{-1}$ is maximally monotone, it follows from Lemma 2.3 that $(\tau_n/\gamma_n)UA^{-1}$ is maximally monotone with respect to $\langle \cdot \mid \cdot \rangle_{(\gamma_n/\tau_n)U^{-1}}$. Hence, $J_{(\tau_n/\gamma_n)UA^{-1}}$ is firmly nonexpansive with respect to the norm $\|\cdot\|_{(\gamma_n/\tau_n)U^{-1}}$. Moreover, we also derive from (3.2) that

$$(\forall n \in \mathbb{N}) \qquad \overline{v} = J_{\frac{\tau_n}{\gamma_n}UA^{-1}}\left(\overline{v} + \frac{\tau_n}{\gamma_n}UL\overline{x}\right). \tag{3.3}$$

Therefore,

$$\gamma_n\|v_{n+1} - \overline{v}\|^2_{(\tau_n U)^{-1}}$$
$$\leq \gamma_n\|v_n - \overline{v} + \frac{\tau_n}{\gamma_n}UL(p_n - \overline{x})\|^2_{(\tau_n U)^{-1}} - \gamma_n\|v_n - v_{n+1} + \frac{\tau_n}{\gamma_n}UL(p_n - \overline{x})\|^2_{(\tau_n U)^{-1}}. \tag{3.4}$$

We have

$$\gamma_n\|v_n - \overline{v} + \frac{\tau_n}{\gamma_n}UL(p_n - \overline{x})\|^2_{(\tau_n U)^{-1}}$$
$$= \gamma_n\|v_n - \overline{v}\|^2_{(\tau_n U)^{-1}} + 2\langle L^*(v_n - \overline{v}) \mid p_n - \overline{x}\rangle + \frac{1}{\gamma_n}\|L(p_n - \overline{x})\|^2_{(\tau_n U)} \tag{3.5}$$

and

$$\gamma_n\|v_n - v_{n+1} + \frac{\tau_n}{\gamma_n}UL(p_n - \overline{x})\|^2_{(\tau_n U)^{-1}}$$
$$= \gamma_n\|v_{n+1} - v_n\|^2_{(\tau_n U)^{-1}} + 2\langle L^*(v_n - v_{n+1}) \mid p_n - \overline{x}\rangle + \frac{1}{\gamma_n}\|L(p_n - \overline{x})\|^2_{(\tau_n U)}. \tag{3.6}$$

By inserting the last two equalities into (3.4), we obtain

$$\gamma_n\|v_{n+1} - \overline{v}\|^2_{(\tau_n U)^{-1}} \leq \gamma_n\|v_n - \overline{v}\|^2_{(\tau_n U)^{-1}} - \gamma_n\|v_{n+1} - v_n\|^2_{(\tau_n U)^{-1}} + 2\langle L^*(v_{n+1} - \overline{v}) \mid p_n - \overline{x}\rangle. \tag{3.7}$$

Let us now estimate the last term in (3.7). Since $P_V$ is self-adjoint, $P_V P_V = P_V$, and $x_{n+1} - \overline{x} \in V$, we have

$$\langle L^*(v_{n+1} - \overline{v}) \mid p_n - \overline{x}\rangle = \langle L^*(v_{n+1} - \overline{v}) \mid x_{n+1} - \overline{x}\rangle + \langle L^*(v_{n+1} - \overline{v}) \mid p_n - x_{n+1}\rangle$$
$$= \langle L^*(v_{n+1} - \overline{v}) \mid P_V(x_{n+1} - \overline{x})\rangle + \gamma_n\langle L^*(v_{n+1} - \overline{v}) \mid P_V L^*(v_{n+1} - v_n)\rangle$$
$$= \langle P_V L^*(v_{n+1} - \overline{v}) \mid x_{n+1} - \overline{x}\rangle + \gamma_n\langle P_V L^*(v_{n+1} - \overline{v}) \mid P_V L^*(v_{n+1} - v_n)\rangle. \tag{3.8}$$

Upon setting $M = LP_V L^*$, we have

$$2\gamma_n\langle(v_{n+1} - \overline{v}) \mid M(v_{n+1} - v_n)\rangle = \gamma_n\|v_{n+1} - \overline{v}\|^2_M + \gamma_n\|v_{n+1} - v_n\|^2_M - \gamma_n\|v_n - \overline{v}\|^2_M \tag{3.9}$$

6

Set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \chi_n & = 2\left\langle P_V L^*(v_{n+1} - \overline{v}) \mid x_{n+1} - \overline{x}\right\rangle, \\ \zeta_n & = 2\left\langle r_n - Bx_n \mid x_{n+1} - \overline{x}\right\rangle, \end{cases} \tag{3.10}$$

and note that, by (3.7),(3.8), and (3.9)

$$\gamma_n \|v_{n+1} - \overline{v}\|^2_{(\tau_n U)^{-1} - M} \le \gamma_n \|v_n - \overline{v}\|^2_{(\tau_n U)^{-1} - M} - \gamma_n \|v_{n+1} - v_n\|^2_{(\tau_n U)^{-1} - M} + \chi_n \tag{3.11}$$

Equation (3.2) yields

$$P_V L^* v_{n+1} = \frac{1}{\gamma_n}\left(P_V x_n - x_{n+1}\right) - P_V r_n$$
$$-P_V L^* \overline{v} = P_V B\overline{x}. \tag{3.12}$$

Using the monotonicity and the cocoercivity of $B$, and noting that $x_{n+1} - \overline{x} \in V$, (3.12) yields

$$\chi_n + \zeta_n = \frac{2}{\gamma_n}\left\langle P_V x_n - x_{n+1} \mid x_{n+1} - \overline{x}\right\rangle - 2\left\langle P_V r_n - r_n \mid x_{n+1} - \overline{x}\right\rangle - 2\left\langle Bx_n - P_V B\overline{x} \mid x_{n+1} - \overline{x}\right\rangle$$

$$= \frac{2}{\gamma_n}\left\langle x_n - x_{n+1} \mid x_{n+1} - \overline{x}\right\rangle - 2\left\langle Bx_n - B\overline{x} \mid x_n - \overline{x}\right\rangle - 2\left\langle Bx_n - B\overline{x} \mid x_{n+1} - x_n\right\rangle$$

$$\le \frac{2}{\gamma_n}\left\langle x_n - x_{n+1} \mid x_{n+1} - \overline{x}\right\rangle + 2\left\langle Bx_n - B\overline{x} \mid x_n - x_{n+1}\right\rangle - 2\beta\|Bx_n - B\overline{x}\|^2$$

$$\le \frac{2}{\gamma_n}\left\langle x_n - x_{n+1} \mid x_{n+1} - \overline{x}\right\rangle - 2\beta\|Bx_n - B\overline{x}\|^2 + \beta\|Bx_n - B\overline{x}\|^2 + \beta^{-1}\|x_n - x_{n+1}\|^2$$

$$\le \frac{1}{\gamma_n}\left(\|x_n - \overline{x}\|^2 - \|x_{n+1} - \overline{x}\|^2\right) - \left(\frac{1}{\gamma_n} - \frac{1}{\beta}\right)\|x_n - x_{n+1}\|^2 - \beta\|Bx_n - B\overline{x}\|^2. \tag{3.13}$$

Set $R_n = \gamma_n^2\left((\tau_n U)^{-1} - M\right)$ and define

$$|||\cdot|||_n^2 \colon \mathcal{H} \times \mathcal{G} \ni (x, v) \mapsto \|x\|^2 + \|v\|^2_{R_n} \quad \text{and} \quad \overline{\boldsymbol{x}} = (\overline{x}, \overline{v}), (\forall n \in \mathbb{N}) \quad \boldsymbol{x}_n = (x_n, v_n). \tag{3.14}$$

We derive from (3.11), (3.13), and (3.14) that

$$|||\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}|||_n^2 \le |||\boldsymbol{x}_n - \overline{\boldsymbol{x}}|||_n^2 - \left(1 - \frac{\gamma_n}{\beta}\right)\|x_{n+1} - x_n\|^2 - \|v_{n+1} - v_n\|^2_{R_n} - \gamma_n \zeta_n - \gamma_n \beta\|Bx_n - B\overline{x}\|^2. \tag{3.15}$$

Let us set

$$\begin{cases} \overline{p}_n = P_V(x_n - \gamma_n(L^* v_n + Bx_n)) \\ \overline{v}_{n+1} = J_{(\tau_n/\gamma_n)UA^{-1}}(v_n + \frac{\tau_n}{\gamma_n}UL\overline{p}_n) \\ \overline{x}_{n+1} = P_V(x_n - \gamma_n(L^* \overline{v}_{n+1} + Bx_n)). \end{cases} \tag{3.16}$$

It follows from the nonexpansiveness of $J_{(\tau_n/\gamma_n)UA^{-1}}$ with respect to $\|\cdot\|_{(\gamma_n/\tau_n)U^{-1}}$ and the nonexpansiveness of $P_V$ that

$$\|x_{n+1} - \overline{x}_{n+1}\| \le \gamma_n\|L^*(v_{n+1} - \overline{v}_{n+1}) + r_n - Bx_n\|$$

$$\le \gamma_n\left(\|L\|\|\overline{v}_{n+1} - v_{n+1}\| + \|r_n - Bx_n\|\right)$$

$$\le \gamma_n\left((\tau_n/\gamma_n)^{1/2}\|L\|^2\|U\|^{1/2}\|\overline{p}_n - p_n\| + \|r_n - Bx_n\|\right)$$

$$\le \gamma_n\left((\tau_n \gamma_n)^{1/2}\|U\|^{1/2}\|L\|^2\|r_n - Bx_n\| + \|r_n - Bx_n\|\right).$$

$$\le \gamma_n((\tau_n \gamma_n)^{1/2}\|U\|^{1/2}\|L\|^2 + 1)\|r_n - Bx_n\|. \tag{3.17}$$

Hence, upon setting $\sigma_n = \gamma_n((\tau_n\gamma_n)^{1/2}\|U\|^{1/2}\|L\|^2 + 1)$,

$$(\forall n \in \mathbb{N}) \quad -\zeta_n = -2\langle r_n - Bx_n \mid \overline{x}_{n+1} - \overline{x}\rangle - 2\langle r_n - Bx_n \mid x_{n+1} - \overline{x}_{n+1}\rangle$$
$$\leq -2\langle r_n - Bx_n \mid \overline{x}_{n+1} - \overline{x}\rangle + 2\|r_n - Bx_n\|\|x_{n+1} - \overline{x}_{n+1}\|$$
$$\leq -2\langle r_n - Bx_n \mid \overline{x}_{n+1} - \overline{x}\rangle + \sigma_n\|r_n - Bx_n\|^2, \tag{3.18}$$

Since $(\gamma_n)_{n\in\mathbb{N}}$ is decreasing and $(\tau_n)_{n\in\mathbb{N}}$ is increasing, we derive from inequality (3.15) that

$$\||\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\||_{n+1}^2 \leq \||\boldsymbol{x}_n - \overline{\boldsymbol{x}}\||_n^2 - \left(1 - \frac{\gamma_n}{\beta}\right)\|x_{n+1} - x_n\|^2 - \|v_{n+1} - v_n\|_{R_n}^2 - \gamma_n\beta\|Bx_n - B\overline{x}\|^2$$
$$- 2\gamma_n\langle r_n - Bx_n \mid \overline{x}_{n+1} - \overline{x}\rangle + \gamma_n\sigma_n\|r_n - Bx_n\|^2. \tag{3.19}$$

Since $L$, $B$, $J_{(\tau_n/\gamma_n)UA^{-1}}$, and $P_V$ are continuous, and $x_n$ is $\mathcal{F}_n$-measurable, $\overline{x}_{n+1} - \overline{x}$ is $\mathcal{F}_n$-measurable, and by (iii), we obtain

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}[\langle r_n - Bx_n \mid \overline{x}_{n+1} - \overline{x}\rangle | \mathcal{F}_n] = \langle \mathsf{E}[r_n - Bx_n | \mathcal{F}_n] \mid \overline{x}_{n+1} - \overline{x}\rangle = 0. \tag{3.20}$$

Therefore, by taking conditional expectation with respect to $\mathcal{F}_n$ of both sides of (3.19), we obtain

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}[\||\boldsymbol{x}_{n+1} - \overline{\boldsymbol{x}}\||_{n+1}^2 | \mathcal{F}_n] \leq \||\boldsymbol{x}_n - \overline{\boldsymbol{x}}\||_n^2 - \mathsf{E}\left[\left(1 - \frac{\gamma_n}{\beta}\right)\|x_{n+1} - x_n\|^2 + \|v_{n+1} - v_n\|_{R_n}^2 \Big| \mathcal{F}_n\right]$$
$$\gamma_n\sigma_n\mathsf{E}[\|r_n - Bx_n\|^2 | \mathcal{F}_n] - \gamma_n\beta\|Bx_n - B\overline{x}\|^2. \tag{3.21}$$

Since the sequence $(\gamma_n\sigma_n\mathsf{E}[\|r_n - Bx_n\|^2 | \mathcal{F}_n])_{n\in\mathbb{N}}$ is summable a.s by assumptions (ii),(iii) and (v), in view of (2.7), (3.21) shows that $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is a stochastic quasi-Fejér monotone sequence with respect to the target set $\boldsymbol{\mathcal{S}}$. Therefore, it follows from Lemma 2.2 and condition (iii) that

$$\begin{cases} (\boldsymbol{x}_n)_{n\in\mathbb{N}} \text{ is bounded a.s.} \\ \sum_{n\in\mathbb{N}} \mathsf{E}\left[\left(1 - \frac{\gamma_n}{\beta}\right)\|x_{n+1} - x_n\|^2 + \|v_{n+1} - v_n\|_{R_n}^2 \big| \mathcal{F}_n\right] < +\infty \text{ a.s.} \\ \sum_{n\in\mathbb{N}} \|Bx_n - B\overline{x}\|^2 < +\infty \text{ a.s.} \end{cases} \tag{3.22}$$

Since $L, B, P_V$ and $J_{(\tau_n/\gamma_n)UA^{-1}}$ are continuous, for every $n \in \mathbb{N}$, $\overline{x}_{n+1}$, $\overline{v}_{n+1}$ and $\overline{p}_n$ are $\mathcal{F}_n$-measurable. Therefore, for every $n \in \mathbb{N}$,

$$\|\overline{x}_{n+1} - x_n\|^2 + \|\overline{v}_{n+1} - v_n\|_{U^{-1}}^2 = \mathsf{E}[\|\overline{x}_{n+1} - x_n\|^2 | \mathcal{F}_n] + \mathsf{E}[\|\overline{v}_{n+1} - v_n\|_{U^{-1}}^2 | \mathcal{F}_n]$$
$$\leq 2\Big(\mathsf{E}[\|\overline{x}_{n+1} - x_{n+1}\|^2 | \mathcal{F}_n] + \mathsf{E}[\|x_{n+1} - x_n\|^2 | \mathcal{F}_n]$$
$$+ \mathsf{E}[\|\overline{v}_{n+1} - v_{n+1}\|_{U^{-1}}^2 | \mathcal{F}_n] + \mathsf{E}[\|v_{n+1} - v_n\|_{U^{-1}}^2 | \mathcal{F}_n]\Big). \tag{3.23}$$

Note that, since $(\tau_n/\gamma_n)A^{-1}$ is maximally monotone, $J_{U(\tau_n/\gamma_n)A^{-1}}$ is firmly nonexpansive with respect to $\langle \cdot \mid \cdot \rangle_{U^{-1}}$ by Lemma 2.3. Thus,

$$\mathsf{E}[\|\overline{v}_{n+1} - v_{n+1}\|_{U^{-1}}^2 | \mathcal{F}_n] \leq (\tau_n/\gamma_n)\mathsf{E}[\|UL(\overline{p}_n - p_n)\|_{U^{-1}}^2 | \mathcal{F}_n]$$
$$\leq \tau\gamma_0\mathsf{E}[\|ULP_V(r_n - Bx_n)\|_{U^{-1}}^2 | \mathcal{F}_n] \to 0, \tag{3.24}$$

8

and hence

$$\mathsf{E}[\|\overline{x}_{n+1} - x_{n+1}\|^2|\mathcal{F}_n] \leq 2\gamma_0^2\mathsf{E}[\|P_V L^*(\overline{v}_{n+1} - v_{n+1})\|^2|\mathcal{F}_n] + 2\gamma_0^2\mathsf{E}[\|P_V(r_n - Bx_n)\|^2|\mathcal{F}_n] \to 0.$$
(3.25)

Now using (3.22), (3.24), and (3.25), we derive from (3.23) that

$$\overline{x}_{n+1} - x_n \to 0 \quad \text{and} \quad \overline{v}_{n+1} - v_n \to 0, \quad \text{and} \quad \overline{x}_{n+1} - \overline{p}_n \to 0 \qquad \text{a.s.} \tag{3.26}$$

Moreover, it follows from the third line of (3.22) that

$$Bx_n \to B\overline{x} \quad \text{a.s.,} \tag{3.27}$$

and from (3.16) that

$$U^{-1}(v_n - \overline{v}_{n+1}) \in \frac{\tau_n}{\gamma_n}\left(A^{-1}\overline{v}_{n+1} - L\overline{p}_n\right) \quad \text{and} \quad \gamma_n^{-1}(x_n - \overline{x}_{n+1}) - Bx_n \in L^*\overline{v}_{n+1} + N_V\overline{x}_{n+1}, \tag{3.28}$$

almost surely. Next, let us prove that every weak cluster point of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is in $\boldsymbol{S}$ a.s. Let $\Omega_0$ be the set of all $\omega \in \Omega$ such that $(\boldsymbol{x}_n(\omega))_{n\in\mathbb{N}}$ is bounded and (3.26),(3.27) and (3.28) are satisfied. We have $\mathsf{P}(\Omega_0) = 1$. Fix $\omega \in \Omega_0$. Let $\boldsymbol{x}(\omega) = (x(\omega), v(\omega))$ be a weak cluster point of $(\boldsymbol{x}_n(\omega))_{n\in\mathbb{N}}$. Then there exists a subsequence $(\boldsymbol{x}_{k_n}(\omega))_{n\in\mathbb{N}}$ that converges weakly to $\boldsymbol{x}(\omega)$. By (3.26)

$$\overline{x}_{k_n+1}(\omega) - x_{k_n}(\omega) \to 0 \quad \text{and} \quad \overline{v}_{k_n+1}(\omega) - v_{k_n}(\omega) \to 0, \quad \text{and} \quad \overline{x}_{k_n+1}(\omega) - \overline{p}_{k_n}(\omega) \to 0 \tag{3.29}$$

and hence by (3.27), and [3, Proposition 20.33(ii)]

$$(\overline{x}_{k_n+1}(\omega), \overline{v}_{k_n+1}(\omega), \overline{p}_{k_n}(\omega)) \rightharpoonup (x(\omega), v(\omega), x(\omega)) \quad \text{and} \quad Bx_{k_n}(\omega) \to B\overline{x} = Bx(\omega). \tag{3.30}$$

The operator $(x, p, v) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G} \mapsto (N_V(x), 0, A^{-1}v) + (L^*v, 0, -Lx)$ is maximally monotone by [3, Corollary 24.4(i)], since it is the sum of two maximally monotone operators [3, Proposition 20.23 and Example 20.30]. Hence, its graph is sequentially closed in $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$ by [3, Proposition 20.33(ii)], and we derive from (3.28) and (3.29) that

$$-Bx(\omega) = L^*v(\omega) + N_Vx(\omega) \quad \text{and} \quad Lx(\omega) \in A^{-1}v(\omega), \tag{3.31}$$

which shows that $(x(\omega), v(\omega)) \in \boldsymbol{S}$ by (3.2). Altogether, it follows Lemma 2.2(iii) that $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$, with $\boldsymbol{x}$ which is $\mathcal{P} \times \mathcal{D}$-valued a.s. $\square$

A direct corollary of the above theorem is the exact case, where no stochastic errors occur.

**Corollary 3.3** *Suppose that $\mathcal{P}$ is non empty. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a decreasing sequence of strictly positive real numbers such that $\gamma_0 < \beta$, let $(\tau_n)_{n\in\mathbb{N}}$ be an increasing sequence of strictly positive real numbers such that $\tau = \sup_{n\in\mathbb{N}} \tau_n < +\infty$, let $U$ be a self adjoint positive definite on $\mathcal{G}$ such that $(\tau U)^{-1} - LP_V L^*$ is positive definite, let $x_0 \in \mathcal{H}$ and $v_0 \in \mathcal{G}$. Iterate*

$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} p_n = P_V(x_n - \gamma_n(L^*v_n + Bx_n)) \\ v_{n+1} = J_{(\tau_n/\gamma_n)UA^{-1}}(v_n + \frac{\tau_n}{\gamma_n}ULp_n) \\ x_{n+1} = P_V(x_n - \gamma_n(L^*v_{n+1} + Bx_n)). \end{array} \right. \tag{3.32}$$

*Then the following hold for some $(x, v) \in \mathcal{P} \times \mathcal{D}$-valued.*

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly $x$ and $(v_n)_{n\in\mathbb{N}}$ converges weakly to $v$.*

(ii) $\sum_{n\in\mathbb{N}} \|Bx_n - Bx\|^2 < +\infty.$

In the following remark we comment on the features of the proposed algorithm and we discuss relations with existing work.

**Remark 3.4**

- Specific instances of the algorithm considered here has been studied in the deterministic case in two papers independently: in [26] to solve a saddle point problem, and in [35] to minimize a regularized least squares problem.

- The proposed algorithm extends the algorithm in [26, 35] in several directions. First, we consider monotone inclusions instead of saddle point problems, and, second, we allow for stochastic evaluations of the single valued cocoercive operator. Moreover, the analysis encompasses a variable step-size, in contrast to the fixed one considered in [26, 35]. Note that our algorithm gives the possibility to treat linear constraints differently from what has been proposed in [26, Section 4.2], thanks to the presence of the projection step. Concerning the results, it is worth noting that the proof of weak convergence relies on different tools than the ones in [26, 35], which are specialized to the variational case.

- As in the deterministic setting, each iteration of the algorithm consists of three steps. Note that, differently from the extra-gradient methods [33, 41], the third step does not require an additional evaluation of the operator $B$. When $V = \mathcal{H}$, and the step-sizes are constant, the algorithm is a specific instance of the one considered in [52]. Due to the variable step-size, we need to use the concept of variable metric stochastic Fejér monotonicity [18, 61].

# 4   Saddle point problems

We next prove some results on ergodic convergence of the duality gap for the case of minimization or saddle point problems.

**Problem 4.1** Let $h\colon \mathcal{H} \to \mathbb{R}$ be a convex differentiable function with a $\beta^{-1}$-Lipschitz continuous gradient, for some $\beta \in \,]0, +\infty[$, let $g \in \Gamma_0(\mathcal{G})$ and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$, let $V$ be a closed vector subspace of $\mathcal{H}$. The primal problem is to

$$\underset{x\in V}{\text{minimize}}\ h(x) + g(Lx), \tag{4.1}$$

and the dual problem is to

$$\underset{v\in\mathcal{G}}{\text{minimize}}\ (h + \iota_V)^*(-L^*v) + g^*(v), \tag{4.2}$$

Denote by $\mathcal{P}_V$ and $\mathcal{D}_V$ the set of solutions to (4.1) and (4.2), respectively.

We consider the convex-concave saddle point formulation of the above problem by setting

$$K\colon \mathcal{H} \times \mathcal{G} \to \mathbb{R} \cup \{-\infty, +\infty\}$$
$$(x, v) \mapsto h(x) + \iota_V(x) + \langle Lx \mid v \rangle - g^*(v). \tag{4.3}$$

We are interested in finding $(\overline{x}, \overline{v}) \in \mathcal{P}_V \times \mathcal{D}_V$, or equivalently (under suitable qualification conditions, see [3, Theorem 15.23 and Proposition 19.18]), a saddle point of (4.3). We will consider the following notion of approximated saddle points, extending to the stochastic case the one given in [42]. Let $\varepsilon > 0$. A $\mathcal{H} \times \mathcal{G}$ valued random variable is an $\varepsilon$- saddle point of $K$ in expectation if

$$\sup_{(\overline{x}, \overline{v}) \in \mathcal{P}_V \times \mathcal{D}_V} \mathsf{E}\left[K(z, \overline{v}) - K(\overline{x}, u)\right] \le \epsilon \tag{4.4}$$

**Algorithm 4.2** Let $(\gamma_n)_{n \in \mathbb{N}}$ be a decreasing sequence of strictly positive real numbers, let $(\tau_n)_{n \in \mathbb{N}}$ be an increasing sequence of strictly positive real numbers, let $U$ be a self adjoint positive definite linear operator on $\mathcal{G}$, let $(r_n)_{n \in \mathbb{N}}$ be a $\mathcal{H}$-valued, square integrable random process, let $x_0$ be a $V$-valued, squared integrable random variable and let $v_0$ be a $\mathcal{G}$-valued, squared integrable random variable. Iterate

$$(\forall N \in \mathbb{N}) \quad \left|
\begin{array}{l}
\text{For } n = 0, \dots, N \\
\quad \left|
\begin{array}{l}
p_n = P_V(x_n - \gamma_n(L^* v_n + r_n)) \\
v_{n+1} = \mathrm{prox}_{(\tau_n/\gamma_n)g^*}^{U^{-1}}(v_n + \frac{\tau_n}{\gamma_n} U L p_n) \\
x_{n+1} = P_V(x_n - \gamma_n(L^* v_{n+1} + r_n))
\end{array}
\right. \\
\tilde{x}_N = \left(\sum_{n=0}^N \gamma_n\right)^{-1} \sum_{n=0}^N \gamma_n x_{n+1} \\
\tilde{v}_N = \left(\sum_{n=0}^N \gamma_n\right)^{-1} \sum_{n=0}^N \gamma_n v_{n+1}.
\end{array}
\right. \tag{4.5}$$

**Theorem 4.3** *In the setting of Problem 4.1, suppose that $\mathcal{P}_V$ is non empty. Moreover, assume that the following conditions are satisfied for Algorithm 4.2, with $\mathcal{F}_n = \sigma((x_k, v_k)_{0 \le k \le n})$*

(i) $\gamma_0 \in \;]0, \beta[$.

(ii) *there exists $\tau \in \;]0, +\infty[$ such that $\sup_{n \in \mathbb{N}} \tau_n \le \tau$ and $(\tau U)^{-1} - L P_V L^*$ is positive semidefinite.*

(iii) $(\forall n \in \mathbb{N}) \; \mathsf{E}[r_n | \mathcal{F}_n] = \nabla h(x_n)$.

(iv) $c_0 = \sum_{n \in \mathbb{N}} \gamma_n^2 \mathsf{E}[\|r_n - \nabla h(x_n)\|^2] < \infty$.

*Let $x \in V$ and let $v \in \mathrm{dom}\, g^*$. Set*

$$c(x, v) = \mathsf{E}[\|x_0 - x\|^2 + \gamma_0^2 \|v_0 - v\|_{(\tau_0 U)^{-1} - L P_V L^*}^2] + 2((\tau \gamma_0 \|U\|)^{1/2} \|L\|^2 + 1) c_0.$$

*Then*

$$\mathsf{E}[K(\tilde{x}_N, v) - K(x, \tilde{v}_N)] \le \frac{c(x, v)}{2} \left(\sum_{n=0}^N \gamma_n\right)^{-1}. \tag{4.6}$$

*Proof.* The Lipschitz continuity of $\nabla h$ and the convexity of $h$ imply that

$$(\forall u \in \mathrm{dom}\, g^*)(\forall(t, y, z) \in V^3) \quad K(t, u) \le K(y, u) + \langle \nabla_x K(z, u) \mid t - y \rangle + \frac{1}{2\beta} \|t - z\|^2. \tag{4.7}$$

Inequality (4.7), with $u = v_{n+1}$, $t = x_{n+1}$, $y = x$ and $z = x_n$, yields

$$K(x_{n+1}, v_{n+1}) \leq K(x, v_{n+1}) + \langle \nabla_x K(x_n, v_{n+1}) \mid x_{n+1} - x \rangle + \frac{1}{2\beta} \|x_{n+1} - x_n\|^2$$

$$= K(x, v_{n+1}) + \langle \nabla h(x_n) + L^* v_{n+1} \mid x_{n+1} - x \rangle + \frac{1}{2\beta} \|x_{n+1} - x_n\|^2$$

$$= K(x, v_{n+1}) + \langle P_V(r_n + L^* v_{n+1}) \mid x_{n+1} - x \rangle + \frac{1}{2\beta} \|x_{n+1} - x_n\|^2 - \zeta_n(x), \quad (4.8)$$

where, for every $n \in \mathbb{N}$, $\zeta_n(x) = \langle P_V(r_n + L^* v_{n+1}) \mid x_{n+1} - x \rangle - \langle \nabla h(x_n) + L^* v_{n+1} \mid x_{n+1} - x \rangle$. Let us set

$$\begin{cases} \overline{p}_n = P_V(x_n - \gamma_n(L^* v_n + \nabla h(x_n))) \\ \overline{v}_{n+1} = \text{prox}_{(\tau_n/\gamma_n)g^*}^{U^{-1}} (v_n + \frac{\tau_n}{\gamma_n} UL\overline{p}_n) \\ \overline{x}_{n+1} = P_V(x_n - \gamma_n(L^* \overline{v}_{n+1} + \nabla h(x_n))). \end{cases} \quad (4.9)$$

Since $x_{n+1} - x \in V$, proceeding as in (3.17)-(3.18), we get

$$(\forall n \in \mathbb{N}) \quad -\zeta_n(x) = -\langle P_V(r_n - \nabla h(x_n)) \mid \overline{x}_{n+1} - x \rangle - \langle P_V(r_n - \nabla h(x_n)) \mid x_{n+1} - \overline{x}_{n+1} \rangle$$

$$\leq -\langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle + \sigma_n \|r_n - \nabla h(x_n)\|^2, \quad (4.10)$$

where $\sigma_n = \gamma_n((\tau_n \gamma_n \|U\|)^{1/2} \|L\|^2 + 1)$. Using (4.5), we get $P_V(r_n + L^* v_{n+1}) = \gamma_n^{-1}(x_n - x_{n+1})$, and hence

$$K(x_{n+1}, v_{n+1}) - K(x, v_{n+1}) \leq \gamma_n^{-1} \langle x_n - x_{n+1} \mid x_{n+1} - x \rangle + \frac{1}{2\beta} \|x_{n+1} - x_n\|^2 - \zeta_n(x)$$

$$\leq \frac{1}{2\gamma_n} \left( \|x_n - x\|^2 - \|x_{n+1} - x\|^2 \right) - \frac{1}{2} \left( \frac{1}{\gamma_n} - \frac{1}{\beta} \right) \|x_{n+1} - x_n\|^2$$

$$- \langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle + \sigma_n \|r_n - \nabla h(x_n)\|^2$$

$$\leq \frac{1}{2\gamma_n} \left( \|x_n - x\|^2 - \|x_{n+1} - x\|^2 \right) - \langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle + \sigma_n \|r_n - \nabla h(x_n)\|^2. \quad (4.11)$$

Therefore,

$$\gamma_n \left( K(x_{n+1}, v_{n+1}) - K(x, v_{n+1}) \right) \leq \frac{1}{2} \left( \|x_n - x\|^2 - \|x_{n+1} - x\|^2 \right)$$

$$- \gamma_n \langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle + \sigma_n \gamma_n \|r_n - \nabla h(x_n)\|^2 \quad (4.12)$$

Since

$$v_{n+1} = \text{prox}_{(\tau_n/\gamma_n)g^*}^{U^{-1}} \left( v_n + \frac{\tau_n}{\gamma_n} ULp_n \right) = \text{prox}_{-(\tau_n/\gamma_n)K(p_n, \cdot)}^{U^{-1}}(v_n), \quad (4.13)$$

inequality (2.4) yields

$$K(p_n, v) - K(p_n, v_{n+1}) \leq \frac{\gamma_n}{\tau_n} \left\langle v - v_{n+1} \mid U^{-1}(v_{n+1} - v_n) \right\rangle. \quad (4.14)$$

Now, simple calculation shows that

$$K(x_{n+1}, v) - K(x_{n+1}, v_{n+1}) + K(p_n, v_{n+1}) - K(p_n, v) = \langle x_{n+1} - p_n \mid L^*(v - v_{n+1}) \rangle. \quad (4.15)$$

12

Therefore, using (4.14) and setting, for every $n \in \mathbb{N}$, $R_n = (\tau_n U)^{-1} - L P_V L^*$ we obtain

$$
\begin{aligned}
\gamma_n K(x_{n+1}, v) - \gamma_n K(x_{n+1}, v_{n+1}) &= \gamma_n \langle x_{n+1} - p_n \mid L^*(v - v_{n+1}) \rangle + \gamma_n(K(p_n, v) - K(p_n, v_{n+1})) \\
&= \gamma_n^2 \langle v_n - v_{n+1} \mid L P_V L^*(v - v_{n+1}) \rangle + \gamma_n(K(p_n, v) - K(p_n, v_{n+1})) \\
&\leq \gamma_n^2 \langle v_n - v_{n+1} \mid L P_V L^*(v - v_{n+1}) \rangle + \frac{\gamma_n^2}{\tau_n} \langle v - v_{n+1} \mid U^{-1}(v_{n+1} - v_n) \rangle \\
&= \gamma_n^2 \langle v_{n+1} - v_n \mid ((\tau_n U)^{-1} - L P_V L^*)(v - v_{n+1}) \rangle \\
&= \gamma_n^2 \Big( \frac{1}{2} \|v_n - v\|_{R_n}^2 - \frac{1}{2} \|v_{n+1} - v\|_{R_n}^2 - \frac{1}{2} \|v_{n+1} - v_n\|_{R_n}^2 \Big)
\end{aligned}
\tag{4.16}
$$

Now, by adding (4.16) and (4.11), and taking into account that $(\gamma_n)_{n \in \mathbb{N}}$ is decreasing and $(\tau_n)_{n \in \mathbb{N}}$ is increasing, we get

$$
\begin{aligned}
\gamma_n \big( K(x_{n+1}, v) - K(x, v_{n+1}) \big) \leq{}& \frac{1}{2} \big( \|x_n - x\|^2 - \|x_{n+1} - x\|^2 \big) + \frac{\gamma_n^2}{2} \|v_n - v\|_{R_n}^2 - \frac{\gamma_{n+1}^2}{2} \|v_{n+1} - v\|_{R_{n+1}}^2 \\
&- \gamma_n \langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle + \gamma_n \sigma_n \|r_n - \nabla h(x_n)\|^2.
\end{aligned}
\tag{4.17}
$$

Since $K(\cdot, \cdot)$ is convex-concave, we have

$$
\begin{aligned}
K(\tilde{x}_N, v) - K(x, \tilde{v}_N) \leq{}& \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1} \sum_{n=0}^{N} \gamma_n(K(x_{n+1}, v) - K(x, v_{n+1})) \\
\leq{}& \frac{1}{2} \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1} \Big( \|x_0 - x\|^2 + \gamma_0^2 \|v_0 - v\|_{G_0}^2 \\
&+ \sum_{n=0}^{N} \big( \gamma_n \sigma_n \|r_n - \nabla h(x_n)\|^2 - \gamma_n \langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle \big) \Big).
\end{aligned}
\tag{4.18}
$$

Since for every $n \in \mathbb{N}$, $\overline{x}_{n+1}$ is $\mathcal{F}_n$-measurable, we have

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \mathsf{E}[\langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle] &= \mathsf{E}[\mathsf{E}[\langle r_n - \nabla h(x_n) \mid \overline{x}_{n+1} - x \rangle \mid \mathcal{F}_n]] \\
&= \mathsf{E}[\langle \mathsf{E}[r_n - \nabla h(x_n) \mid \mathcal{F}_n] \mid \overline{x}_{n+1} - x \rangle] \\
&= 0.
\end{aligned}
\tag{4.19}
$$

Therefore, by taking the expectation of both sides of (4.18), we obtain

$$
\begin{aligned}
&\mathsf{E}[K(\tilde{x}_N, v) - K(x, \tilde{v}_N)] \\
&\qquad \leq \frac{1}{2} \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1} \Big( \mathsf{E}[\|x_0 - x\|^2 + \gamma_0^2 \|v_0 - v\|_{G_0}^2] + \sum_{n=0}^{N} \big( \gamma_n \sigma_n \mathsf{E}[\|r_n - \nabla h(x_n)\|^2] \big) \Big) \\
&\qquad \leq \frac{1}{2} \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1} \Big( \mathsf{E}[\|x_0 - x\|^2 + \gamma_0^2 \|v_0 - v\|_{G_0}^2] + 2((\gamma_0 \tau \|U\|)^{1/2} \|L\|^2 + 1) c_0 \Big),
\end{aligned}
\tag{4.20}
$$

which proves (4.6). $\square$

**Corollary 4.4** *Under the assumptions of Theorem 4.3, suppose that $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$, and that $\mathcal{P}_V$ and $\mathcal{D}_V$ are bounded. Let $\alpha = \sup_{(\overline{x},\overline{v})\in\mathcal{P}_V\times\mathcal{D}_V} \mathsf{E}[\|x_0-\overline{x}\|^2 + \gamma_0^2\|v_0-\overline{v}\|_{G_0}^2] + 2((\gamma_0\tau\|U\|)^{1/2}\|L\|^2+1)c_0$. Let $\varepsilon > 0$, and let $N_\varepsilon \in \mathbb{N}$ be such that $\sum_{n=1}^{N_\varepsilon} \gamma_n \leq 2/(\alpha\varepsilon)$. Then $(\tilde{x}_{N_\varepsilon}, \tilde{v}_{N_\varepsilon})$ is an $\varepsilon$-saddle point in expectation.*

*Proof.* The conclusion directly follows from (4.6), noting that $\sup_{(\overline{x},\overline{v})\in\mathcal{P}_V\times\mathcal{D}_V} c(\overline{x},\overline{v}) \leq \alpha$. $\square$

# 5 Application to sum of composite inclusions

Based on the standard product space reformulation technique, one can apply the proposed framework to more general composite inclusions. As an illustration, we present below an application to the sum of composite operators and a cocoercive operator [18, 55, 56].

**Problem 5.1** Let $\mathcal{H}$ be a real Hilbert space, let $m$ be a strictly positive integer and let $(\omega_i)_{1\leq i\leq m} \in [0,1]^m$ be such that $\sum_{i=1}^m \omega_i = 1$, let $C\colon \mathcal{H} \to \mathcal{H}$ be $\mu$-cocoercive for some $\mu \in ]0,+\infty[$. For every $i \in \{1,\ldots,m\}$, let $\mathcal{G}_i$ be a real Hilbert space, let $A_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be maximally monotone, and suppose that $L_i\colon \mathcal{H} \to \mathcal{G}_i$ is a nonzero bounded linear operator. The problem is to solve the primal inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in \sum_{i=1}^m \omega_i L_i^* A_i(L_i\overline{x}) + C\overline{x}, \tag{5.1}$$

together with the dual inclusion

find $\overline{v}_1 \in \mathcal{G}_1, \ldots, \overline{v}_m \in \mathcal{G}_m$ such that

$$(\exists x \in \mathcal{H}) \quad -Cx = \sum_{i=1}^m \omega_i L_i^* \overline{v}_i \quad \text{and} \quad (\forall i \in \{1,\ldots,m\}) \quad \overline{v}_i \in A_i(L_i x). \tag{5.2}$$

We denote by $\mathcal{P}_1$ and $\mathcal{D}_1$ the sets of solutions to (5.1) and (5.2), respectively.

Let us recall the following facts that show that Problem 5.1 is a special case of Problem 1.1 (see also [9, Theorem 3.8] for the case where $C$ is zero).

**Lemma 5.2** *Define $\boldsymbol{\mathcal{H}} = \mathcal{H}^m$ and $\boldsymbol{\mathcal{G}} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m$, endowed with the scalar product*

$$(\forall(\boldsymbol{v},\boldsymbol{w}) \in \boldsymbol{\mathcal{G}}^2) \quad \langle \boldsymbol{v} \mid \boldsymbol{w} \rangle_{\boldsymbol{\mathcal{G}}} = \sum_{i=1}^m \omega_i \langle v_i \mid w_i \rangle \tag{5.3}$$

*and*

$$\begin{cases} \boldsymbol{V} = \{\boldsymbol{x} \in \boldsymbol{\mathcal{H}} \mid x_1 = \ldots = x_m\} \\ \boldsymbol{A} = \bigtimes_{k=1}^m A_i \\ \boldsymbol{C}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{x} \mapsto (Cx_1,\ldots,Cx_m) \\ \boldsymbol{L}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}}\colon \boldsymbol{x} \mapsto (L_1 x_1,\ldots,L_m x_m). \end{cases}$$

*Then the following hold.*

14

(i) $\boldsymbol{C}$ *is $\mu$-cocoercive.*

(ii) *$x$ solves (5.1) if and only if $(x,\dots,x) \in \mathrm{zer}(\boldsymbol{C} + \boldsymbol{L}^*\boldsymbol{AL} + N_{\boldsymbol{V}})$.*

*Proof.* (i): Since $C$ is $\mu$-cocoercive, we have

$$(\forall \boldsymbol{x} \in \mathcal{H}^m)(\forall \boldsymbol{y} \in \mathcal{H}^m) \quad \langle \boldsymbol{Cx} - \boldsymbol{Cy} \mid \boldsymbol{x} - \boldsymbol{y} \rangle = \sum_{i=1}^m \omega_i \langle Cx_i - Cy_i \mid x_i - y_i \rangle$$
$$\geq \mu \sum_{i=1}^m \omega_i \|Cx_i - Cy_i\|^2$$
$$= \mu \|\boldsymbol{Cx} - \boldsymbol{Cy}\|^2, \tag{5.4}$$

which shows that $\boldsymbol{C}$ is $\mu$-cocoercive on $\boldsymbol{\mathcal{H}}$.

(ii): We have $\boldsymbol{L}^*\colon (v_1,\dots,v_m) \mapsto (L_1^* v_1, \dots, L_m^* v_m)$. Moreover, in view of (5.3), $\boldsymbol{V}^\perp = \left\{ \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \mid \sum_{i=1}^m \omega_i x_i = 0 \right\}$. We have

$$x \text{ solves } (5.1) \Leftrightarrow 0 \in \sum_{i=1}^m \omega_i L_i^* A_i(L_i x) + Cx$$
$$\Leftrightarrow (\exists (y_i)_{1 \leq i \leq m} \in \times_{i=1}^m L_i^* A_i L_i x) \quad 0 = \sum_{i=1}^m \omega_i(y_i) + Cx$$
$$\Leftrightarrow (\exists \boldsymbol{y} \in \times_{i=1}^m L_i^* A_i L_i x) \quad -\boldsymbol{y} - (Cx,\dots,Cx) \in \boldsymbol{V}^\perp = N_{\boldsymbol{V}}(x,\dots,x)$$
$$\Leftrightarrow (\exists \boldsymbol{y} \in \times_{i=1}^m L_i^* A_i L_i x) \quad 0 \in \boldsymbol{y} + (Cx,\dots,Cx) + N_{\boldsymbol{V}}(x,\dots,x)$$
$$\Leftrightarrow 0 \in \boldsymbol{L}^*\boldsymbol{AL}(x,\dots,x) + \boldsymbol{C}(x,\dots,x) + N_{\boldsymbol{V}}(x,\dots,x) \tag{5.5}$$

which proves (ii). $\square$

**Corollary 5.3** *Let $(\gamma_n)_{n\in\mathbb{N}}$ be a decreasing sequence of strictly positive real numbers and let $(\tau_n)_{n\in\mathbb{N}}$ be an increasing sequence of strictly positive real numbers such that $\sup_{n\in\mathbb{N}} \tau_n = \tau < +\infty$ For every $i \in \{1,\dots,m\}$ let $U_i$ be a self adjoint positive definite linear operator on $\mathcal{G}_i$, let $(r_n)_{n\in\mathbb{N}}$ be a $\mathcal{H}$-valued, squared integrable random process, let $x_0$ be a $\mathcal{H}$-valued, squared integrable random variable and let $v_{i,0}$ be a $\mathcal{G}_i$-valued, squared integrable random variable. Iterate*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_n = x_n - \gamma_n \sum_{i=1}^m \omega_i(L_i^* v_{i,n} + r_n) \\ \text{For } i = 1, \dots, m \\ \quad v_{i,n+1} = J_{(\tau_n/\gamma_n)U_i A_i^{-1}}\left(v_{i,n} + \frac{\tau_n}{\gamma_n} U_i L_i p_n\right) \\ x_{n+1} = x_n - \gamma_n \sum_{i=1}^m \omega_i(L_i^* v_{i,n+1} + r_n) \end{array} \right. \tag{5.6}$$

*Suppose that $\mathcal{P}_1$ is non empty and that the following conditions are satisfied for $\mathcal{F}_n = \sigma((x_k, (v_{i,k})_{1\leq i \leq m})_{0 \leq k \leq n})$*

(i) $(\forall i \in \{1,\dots,m\}) \quad (\tau U_i)^{-1} - L_i L_i^*$ *is positive definite.*

15

(ii) $\gamma \in {]}0, \mu[$.

(iii) $\mathsf{E}[r_n | \mathscr{F}_n] = Cx_n$.

(iv) $\sum_{n \in \mathbb{N}} \mathsf{E}[\|r_n - Cx_n\|^2 | \mathscr{F}_n] < +\infty \quad \mathsf{P}\text{-}a.s.$

*Then the following hold for some random vectors $\overline{x} \in \mathcal{P}_1$ a.s. and $(\overline{v}_1, \ldots, \overline{v}_m) \in \mathcal{D}_1$ a.s.*

(i) *$(x_n)_{n \in \mathbb{N}}$ converges weakly to $\overline{x}$ almost surely.*

(ii) *For every $i \in \{1, \ldots, m\}$, $(v_{i,n})_{n \in \mathbb{N}}$ converges weakly to $\overline{v}_i$ a.s.*

(iii) *$\sum_{n \in \mathbb{N}} \|Cx_n - Cx\|^2 < +\infty$ a.s.*

*Proof.* Let us define
$$\boldsymbol{U} \colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{G}} \colon (v_1, \ldots, v_m) \mapsto (U_1 v_1, \ldots, U_m v_m). \tag{5.7}$$
Then $\boldsymbol{U}$ is self adjoint, positive definite on $\boldsymbol{\mathcal{G}}$. Moreover, for every $n \in \mathbb{N}$,
$$(\forall \boldsymbol{v} = (v_1, \ldots, v_m) \in \boldsymbol{\mathcal{G}}) \quad J_{(\tau_n/\gamma_n)\boldsymbol{U}\boldsymbol{A}^{-1}} \boldsymbol{v} = (J_{(\tau_n/\gamma_n)U_i A_i^{-1}} v_i)_{1 \le i \le m}. \tag{5.8}$$
We recall that
$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad P_{\boldsymbol{V}} \boldsymbol{x} = \Big( \sum_{i=1}^{m} \omega_i x_i \Big)_{1 \le k \le m}. \tag{5.9}$$
Therefore, upon setting
$$\begin{cases} \boldsymbol{r}_n = (r_n, \ldots, r_n) \in \boldsymbol{V} \\ \boldsymbol{x}_n = (x_n, \ldots, x_n) \in \boldsymbol{V} \\ \boldsymbol{v}_n = (v_{1,n}, \ldots, v_{m,n}) \\ \boldsymbol{p}_n = (p_n, \ldots, p_n) \in \boldsymbol{V}, \end{cases} \tag{5.10}$$
we can rewrite (5.6) as
$$(\forall n \in \mathbb{N}) \quad \left| \begin{array}{l} \boldsymbol{p}_n = P_{\boldsymbol{V}}(\boldsymbol{x}_n - \gamma_n(\boldsymbol{L}^* \boldsymbol{v}_n + \boldsymbol{r}_n)) \\ \boldsymbol{v}_{n+1} = J_{(\tau_n/\gamma_n)\boldsymbol{U}\boldsymbol{A}^{-1}}(\boldsymbol{v}_n + \frac{\tau_n}{\gamma_n}\boldsymbol{U}\boldsymbol{L}\boldsymbol{p}_n) \\ \boldsymbol{x}_{n+1} = P_{\boldsymbol{V}}(\boldsymbol{x}_n - \gamma_n(\boldsymbol{L}^* \boldsymbol{v}_{n+1} + \boldsymbol{r}_n)), \end{array} \right. \tag{5.11}$$
which is a special case of (3.1). By Lemma 5.2, Problem 5.1 is a special case of Problem 1.1 Moreover, every specific condition in Theorem 3.2 is satisfied. Therefore, the first and the third conclusions follow from Theorem 3.2. We prove the second one. By Theorem 3.2, $(\boldsymbol{v}_n)_{n \in \mathbb{N}}$ converge weakly to $\boldsymbol{v}$ a.s., such that
$$0 \in -\boldsymbol{L}(N_{\boldsymbol{V}} + \boldsymbol{C})^{-1}(-\boldsymbol{L}^* \boldsymbol{v}) + \boldsymbol{A}^{-1} \boldsymbol{v}. \tag{5.12}$$
We now prove that $\boldsymbol{v} \in \mathcal{D}_1$. We have
$$\begin{aligned} (5.12) &\Rightarrow (\exists \boldsymbol{x} \in \boldsymbol{V}) \quad -\boldsymbol{L}^* \boldsymbol{v} \in N_{\boldsymbol{V}} \boldsymbol{x} + \boldsymbol{C}\boldsymbol{x} \quad \text{and} \quad \boldsymbol{v} \in \boldsymbol{A}\boldsymbol{L}\boldsymbol{x}. \\ &\Leftrightarrow (\exists \boldsymbol{x} \in \boldsymbol{V}) \quad -\boldsymbol{L}^* \boldsymbol{v} - \boldsymbol{C}\boldsymbol{x} \in \boldsymbol{V}^\perp \quad \text{and} \quad (\forall i \in \{1, \ldots, m\}) \ v_i \in A_i L_i x. \\ &\Leftrightarrow (\exists x \in \mathcal{H}) \quad -\sum_{i=1}^{m} \omega_i L_i v_i = Cx \quad \text{and} \quad (\forall i \in \{1, \ldots, m\}) \ v_i \in A_i L_i x. \\ &\Leftrightarrow \boldsymbol{v} = (v_i)_{1 \le i \le m} \in \mathcal{D}_1. \end{aligned} \tag{5.13}$$

$\square$

## 5.1 Structured minimization problems

We next show how the previous result can be specialized to the case of minimization problems involving the sum of composite functions [15, 18].

**Problem 5.4** Let $\mathcal{H}$ be a real Hilbert space, let $m$ be a strictly positive integer and let $(\omega_i)_{1 \leq i \leq m} \in [0,1]^m$ be such that $\sum_{i=1}^m \omega_i = 1$, let $h \in \Gamma_0(\mathcal{H})$ be differentiable, with $1/\mu$-Lipschitz continuous gradient for some $\mu \in \,]0, +\infty[$. For every $i \in \{1, \ldots, m\}$, let $\mathcal{G}_i$ be a real Hilbert space, let $g_i \in \Gamma_0(\mathcal{G}_i)$ be maximally monotone, and suppose that $L_i \colon \mathcal{H} \to \mathcal{G}_i$ is a nonzero bounded linear operator. The problem is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \ \sum_{i=1}^m \omega_i g_i(L_i x) + h(x) \tag{5.14}$$

together with the dual

$$\underset{v_1 \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m}{\text{minimize}} \ \sum_{i=1}^m \omega_i g_i^*(v_i) + h^* \Big( - \sum_{i=1}^m \omega_i L_i^* v_i \Big). \tag{5.15}$$

under the assumption that problem (5.14) has at least a solution.

As in Section 4 we will also consider the saddle point formulation of Problem 5.4.

$$H(x, v_1, \ldots, v_m) = h(x) + \sum_{i=1}^m \omega_i \langle L_i^* v_i \mid x \rangle - \sum_{i=1}^m \omega_i g_i^*(v_i) \tag{5.16}$$

The following algorithm is a special case of Algorithm 3.1.

**Algorithm 5.5** Let $(\gamma_n)_{n \in \mathbb{N}}$ be a decreasing sequence of strictly positive real numbers, let $(\tau_n)_{n \in \mathbb{N}}$ be an increasing sequence of strictly positive real numbers, let $U$ be a self adjoint positive definite linear operator on $\mathcal{G}$, let $(r_n)_{n \in \mathbb{N}}$ be a $\mathcal{H}$-valued, square integrable random process, let $x_0$ be a $\mathcal{H}$-valued, squared integrable random variable and let $v_0$ be a $\mathcal{G}$-valued, squared integrable random variable. Iterate

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} p_n = x_n - \gamma_n \sum_{i=1}^m \omega_i (L_i^* v_{i,n} + r_n) \\ \text{For } i = 1, \ldots, m \\ \quad \left\lfloor v_{i,n+1} = \text{prox}_{g_i^*}^{(\tau_n/\gamma_n)U_i^{-1}} \left( v_{i,n} + \frac{\tau_n}{\gamma_n} U_i L_i p_n \right) \right. \\ x_{n+1} = x_n - \gamma_n \sum_{i=1}^m \omega_i (L_i^* v_{i,n+1} + r_n) \end{array} \right. \tag{5.17}$$

The following result is a direct consequence of Corollary 5.3.

**Corollary 5.6** *In the setting of Problem 5.4, suppose that*

$$0 \in \text{zer} \Big( \nabla h + \sum_{i=1}^m \omega_i L_i^* \partial g_i(L_i \cdot) \Big) \tag{5.18}$$

*Let $(x_n, v_n)_{n \in \mathbb{N}}$ be the sequence generated by Algorithm 5.5. Suppose that the following conditions are satisfied for $\mathcal{F}_n = \sigma((x_k, (v_{i,k})_{1 \leq i \leq m})_{0 \leq k \leq n})$*

17

(i) *there exists $\tau \in ]0, +\infty[$ such that $\sup_{n \in \mathbb{N}} \tau_n \leq \tau$ and $(\forall i \in \{1, \ldots, m\})$ $(\tau U_i)^{-1} - L_i L_i^*$ is positive definite.*

(ii) $\gamma_0 \in ]0, \mu[$ *and* $\inf \gamma_n > 0$.

(iii) $\mathsf{E}[r_n | \mathcal{F}_n] = \nabla h(x_n)$.

(iv) $\sum_{n \in \mathbb{N}} \mathsf{E}[\|r_n - \nabla h(x_n)\|^2 | \mathcal{F}_n] < +\infty$ P-*a.s.*

*Then the following hold for some random vector $\overline{x}$ solving (5.14) and some random vector $(\overline{v}_1, \ldots, \overline{v}_m)$ solving (5.15) a.s.*

(i) $(x_n)_{n \in \mathbb{N}}$ *converges weakly to $\overline{x}$ almost surely.*

(ii) *For every $i \in \{1, \ldots, m\}$ $(v_{i,n})_{n \in \mathbb{N}}$ converges weakly to $\overline{v}_i$ a.s.*

(iii) $\sum_{n \in \mathbb{N}} \|\nabla h(x_n) - \nabla h(x)\|^2 < +\infty$ *a.s.*

**Corollary 5.7** *In the setting of Problem 5.4, let $(x_n, v_n)_{n \in \mathbb{N}}$ be the sequence generated by Algorithm 5.5. For every $N \in \mathbb{N}$, define*

$$\tilde{x}_N = \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1} \sum_{n=0}^{N} \gamma_n x_{n+1}$$

$$(\forall i \in \{1, \ldots, m\}) \quad \tilde{v}_{i,N} = \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1} \sum_{n=0}^{N} \gamma_n v_{n+1}.$$

*Suppose that the following conditions are satisfied for $\mathcal{F}_n = \sigma((x_k, v_k)_{0 \leq k \leq n})$*

(i) $\gamma_0 \in ]0, \mu[$.

(ii) *there exists $\tau \in ]0, +\infty[$ such that $\sup_{n \in \mathbb{N}} \tau_n \leq \tau$ and, for every $i \in \{1, \ldots, m\}$, the operator $(\tau U_i)^{-1} - L P_V L^*$ is positive semidefinite.*

(iii) $(\forall n \in \mathbb{N})$ $\mathsf{E}[r_n | \mathcal{F}_n] = \nabla h(x_n)$.

(iv) $c_0 = \sum_{n \in \mathbb{N}} \gamma_n^2 \mathsf{E}[\|r_n - \nabla h(x_n)\|^2] < \infty$.

*Let $x \in \mathcal{H}$ and let $\boldsymbol{v} = (v_i)_{1 \leq i \leq m} \in \times_{i=1}^m \operatorname{dom} g_i^*$. Set $c(x, \boldsymbol{v}) = \mathsf{E}[\|x_0 - x\|^2 + \gamma_0^2 \sum_{i=1}^m \omega_i \|v_{i,0} - v_i\|^2_{(\tau_0 U_i^{-1} - L P_V L^*)}] + 2((\gamma_0 \tau \|U\|)^{1/2} \|L\|^2 + 1) c_0$. Then*

$$\mathsf{E}[H(\tilde{x}_N, (v_1, \ldots, v_m)) - H(x, (\tilde{v}_{1,N}, \ldots, \tilde{v}_{m,N}))] \leq \frac{c(x, \boldsymbol{v})}{2} \Big( \sum_{n=0}^{N} \gamma_n \Big)^{-1}. \tag{5.19}$$

# References

[1] F. Alvarez and H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping, *Set-Valued Anal.*, vol. 9, pp. 3–11, 2001.

[2] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, A parallel splitting method for coupled monotone inclusions, *SIAM J. Control Optim.*, vol. 48, pp. 3246–3270, 2010.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[4] K. Barty , J.-S. Roy and C. Strugarek, Hilbert-valued perturbed subgradient algorithms, *Math. Oper. Res.*, vol. 32, pp. 551–562, 2007.

[5] A. Bennar and J.-M. Monnez, Almost sure convergence of a stochastic approximation process in a convex set, *Int. J. Appl. Math.*, vol. 20, pp. 713-722, 2007.

[6] P. Bianchi, W. Hachem and F. Iutzeler, A Coordinate Descent Primal-Dual Algorithm and Application to Distributed Asynchronous Optimization, *IEEE Trans. Automat. Contr.*, to appear.

[7] P. Bianchi and W. Hachem, Dynamical behavior of a stochastic forward-backward algorithm using random monotone operators, http://arXiv.org/abs/1508.02845.

[8] R. I. Bot, E. R. Csetnek, A. Heinrich, and C. Hendrich On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems, *Math. Program.*, vol. 150, pp. 251-279, 2015.

[9] L. M. Briceño-Arias and P. L. Combettes, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.*, vol. 21, pp. 1230–1250, 2011.

[10] L. M. Briceño-Arias and P. L. Combettes, Monotone operator methods for Nash equilibria in non-potential games, in *Computational and Analytical Mathematics,* (D. Bailey, H. H. Bauschke, P. Borwein, F. Garvan, M. Théra, J. Vanderwerff, and H. Wolkowicz, eds.). Springer, New York, 2013.

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

[12] A. Chambolle and T. Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, *Math. Program.*, online first, 2015.

[13] P. Chen, J. Huang, and X. Zhang, A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration, *Inverse Problems*, vol. 29, pp.025011, 2013.

[14] P. L. Combettes, Systems of structured monotone inclusions: duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420-2447, 2013.

[15] P. L. Combettes and J.-C. Pesquet, Primal-Dual Splitting Algorithm for Solving Inclusions with Mixtures of Composite, Lipschitzian, and Parallel-Sum Type Monotone Operators, *Set-Valued Var. Anal.*, vol. 20, pp. 307-330, 2012.

[16] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, vol. 53, pp. 475–504, 2004.

[17] P. L Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, A forward-backward view of some primal-dual optimization methods in image recovery, *In Proc. Int. Conf. Image Process.*, Paris, France, 27-30 Oct. 2014

[18] P. L. Combettes and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization*, vol. 63, pp. 1289-1318, 2013.

[19] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, *SIAM J. Optim.*, vol. 25, pp. 1221-1248, 2015.

[20] P. L. Combettes and J.-C. Pesquet, Stochastic approximations and perturbations in forward-backward splitting for monotone operators, *Pure Appl. Funct. Anal.*, to appear.

[21] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.* vol. 4, pp. 1168–1200, 2005.

[22] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *J. Optim. Theory Appl.*, vol. 158, pp.460–479, 2013.

[23] I. Daubechies, M. Defrise, and C. De Mol, An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint, *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[24] D. Davis, Convergence rate analysis of primal-dual splitting schemes, preprint, *SIAM J. Optim.*, vol. 25, pp. 1912–1943, 2015.

[25] E. De Vito, V. Umanità, and S. Villa, A consistent algorithm to solve Lasso, elastic-net and Tikhonov regularization, *J. Complexity,* vol. 27, pp. 188–200, 2011.

[26] Y. Drori, S. Sabach, and M. Teboulle, A simple algorithm for a class of nonsmooth convex-concave saddle-point problems, *Oper. Res. Lett.*, vol. 43, pp. 209–214, 2015.

[27] J. Duchi and Y. Singer, Efficient online and batch learning using forward backward splitting, *J. Mach. Learn. Res.,* vol. 10, pp. 2899–2934, 2009.

[28] R. Durrett, *Probability: theory and example.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[29] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer-Verlag, New York, 2003.

[30] R. Fortet, *Vecteurs, fonctions et distributions aléatoires dans les espaces de Hilbert*, Editions Hermès, Paris, 1995.

[31] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics.* SIAM, Philadelphia, 1989.

[32] H. Jiang and H. Xu, Stochastic approximation approaches to the stochastic variational inequality problem, *IEEE Trans. Automat. Control*, vol. 53, pp. 1462–1475, 2008.

[33] G. M. Korpelevich, The extragradient method for finding saddle points and other problems *Matecon*, vol.12, pp. 747–756, 1976.

[34] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes.* Springer, New York, 1991.

[35] I. Loris and C. Verhoeven, On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty *Inverse Problems*, vol. 27, pp. 125007, 2011.

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

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

[38] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Sér. A*, vol. 255, pp. 2897–2899, 1962.

[39] S. Mosci, L. Rosasco, M. Santoro, A. Verri and S. Villa, Solving Structured Sparsity Regularization with Proximal Methods, in *Machine Learning and Knowledge discovery in Databases European Conference*, pp. 418-433, 2010.

[40] A. Moudafi and M. Oliny, Convergence of a splitting inertial proximal method for monotone operators, *J. Comput. Appl. Math.*, vol. 155, pp. 447-454, 2003.

[41] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM J. Optim.*, vol. 15, pp. 229–251, 2004.

[42] A. Nemirovski and D. Yudin, *Problem complexity and method efficiency in Optimization* Wiley, New York, 1983.

[43] J.-C. Pesquet and N. Komodakis, Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems *IEEE Signal Processing Magazine*, to appear.

[44] J. C. Pesquet and N. Pustelnik, A parallel inertial proximal optimization method, *Pacific J. Optim.* vol. 8, pp. 273-306, 2012.

[45] J.-C. Pesquet and A. Repetti, A class of randomized primal-dual algorithms for distributed optimization, *J. Nonlinear Convex Anal.*, vol. 16, pp. 2453–2490, 2015.

[46] B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *U.S.S.R. Comput. Math. Math. Phys.*, vol. 4, pp. 1–17, 1964.

[47] D. A. Lorenz and T. Pock, An inertial forward-backward method for monotone inclusions *J. Math. Imaging Vis.*, vol. 51, pp.311–325, 2015.

[48] I. Loris and C. Verhoeven, On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty, *Inverse Problems*, vol. 27, no. 12, p. 125007, 2011.

[49] H. Raguet, J. Fadili, and G. Peyré, Generalized forward-backward splitting, *SIAM J. Imaging Sci.*, vol. 6, pp. 1199–1226, 2013.

[50] L. Rosasco, S. Villa, S. Mosci, M. Santoro, M. and A. Verri, Nonparametric sparsity and regularization, *J. Mach. Learn. Res.*, vol. 14, pp. 1665–1714, 2013.

[51] L. Rosasco, S. Villa, and B. C. Vũ, A Stochastic forward-backward splitting method for solving monotone inclusions in Hilbert spaces, J. Optim Theory Appl., to appear.

[52] L. Rosasco, S. Villa, and B. C. Vũ, A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions, *Optimization*, to appear.

[53] L. Rosasco, S. Villa, and B. C. Vũ, Convergence of stochastic proximal gradient, 2014, arXiv:1403.5074.

[54] H. Robbins and D. Siegmund, A Convergence theorem for non negative almost supermartingales and some applications, in: *Optimizing Methods in Statistic*, (J. S. Rustagi, Ed. ), pp. 233–257. Academic Press, New York, 1971.

[55] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[56] B. C. Vũ, A splitting algorithm for coupled system of primal-dual monotone inclusions, *J. Optim. Theory Appl.*, published online February 2014.

[57] P. Tseng, Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming, *Math. Programming*, vol. 48, pp. 249–263, 1990.

[58] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.*, vol. 29, pp. 119–138, 1991.

[59] S. Villa, L. Rosasco, S. Mosci, and A. Verri, Proximal methods for the latent group lasso penalty, *Comput. Anal. Optim.*, vol. 58, pp. 381–407, 2014.

[60] S. Villa, S. Salzo, L. Baldassarre, and A. Verri, Accelerated and inexact forward-backward algorithms, *SIAM J. Optim.*, vol. 23, pp. 1607–1633, 2013.

[61] B. C. Vũ, Almost sure convergence of the forward-backward-forward splitting algorithm *Optim. Letters*, online first, pp. 1–23, 2015.

[62] D. L. Zhu and P. Marcotte, Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities, *SIAM J. Optim.,* vol. 6, pp. 714–726, 1996.