A constant step Forward-Backward algorithm involving random maximal monotone operators

Pascal Bianchi* Walid Hachem† Adil Salim*

January 15, 2018

Abstract

A stochastic Forward-Backward algorithm with a constant step is studied. At each time step, this algorithm involves an independent copy of a couple of random maximal monotone operators. Defining a mean operator as a selection integral, the differential inclusion built from the sum of the two mean operators is considered. As a first result, it is shown that the interpolated process obtained from the iterates converges narrowly in the small step regime to the solution of this differential inclusion. In order to control the long term behavior of the iterates, a stability result is needed in addition. To this end, the sequence of the iterates is seen as a homogeneous Feller Markov chain whose transition kernel is parameterized by the algorithm step size. The cluster points of the Markov chains invariant measures in the small step regime are invariant for the semiflow induced by the differential inclusion. Conclusions regarding the long run behavior of the iterates for small steps are drawn. It is shown that when the sum of the mean operators is demipositive, the probabilities that the iterates are away from the set of zeros of this sum are small in Cesàro mean. The ergodic behavior of these iterates is studied as well. Applications of the proposed algorithm are considered. In particular, a detailed analysis of the random proximal gradient algorithm with constant step is performed.

Keywords: Dynamical systems, Narrow convergence of stochastic processes, Random maximal monotone operators, Stochastic approximation with constant step, Stochastic Forward - Backward algorithm, Stochastic proximal point algorithm.
47H05, 47N10, 62L20, 34A60.

1 Introduction

Given two maximal monotone operators A and B on the space $E = \mathbb{R}^N$, where B is single valued, the Forward-Backward splitting algorithm is an iterative algorithm for finding a zero of the sum operator $\mathsf{A} + \mathsf{B}$. It reads

$$x_{n+1} = (I + \gamma \mathsf{A})^{-1}(x_n - \gamma \mathsf{B}(x_n)), \tag{1}$$

where γ is a positive step. This algorithm consists in a forward step $(I - \gamma \mathsf{B})(x_n)$ followed by a backward step, where the resolvent $(I + \gamma \mathsf{A})^{-1}$ of A, known to be single valued as A is maximal monotone, is applied to the output of the former. When B satisfies a so called cocoercivity condition, and when the step γ is small enough, the convergence of the algorithm towards a zero of $\mathsf{A} + \mathsf{B}$ (provided it exists) is a well established fact [6, Ch. 25]. In the field of convex optimization, this algorithm can be used to find a minimizer of the sum of two real functions $F + G$ on E, where F is a convex function which is defined on the whole E and which has a

*LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France.
46, rue Barrault, 75634 Paris Cedex 13, France.
† CNRS / LIGM (UMR 8049), Université Paris-Est Marne-la-Vallée.
5, boulevard Descartes, Champs-sur-Marne, 77454, Marne-la-Vallée Cedex 2, France.
pascal.bianchi@telecom-paristech.fr, walid.hachem@u-pem.fr, adil.salim@telecom-paristech.fr

Lipschitz gradient, and where G is a convex, proper, and lower semi continuous (lsc) function. In this case, the Forward-Backward algorithm is known as the proximal gradient algorithm, and is written as $x_{n+1} = \text{prox}_{\gamma G}(x_n - \gamma \nabla F(x_n))$, where $\text{prox}_{\gamma G} := (I + \gamma \partial G)^{-1}$ is Moreau's proximity operator of γG.

In this paper, we are interested in the situation where the operators A and B are replaced with random maximal monotone operators. Denote as $\mathscr{M}$ the set of maximal monotone operators on E, let $A, B : \Xi \to \mathscr{M}$ be two functions from a measurable space $(\Xi, \mathscr{G})$ to $\mathscr{M}$, and let (ξ_n) be a sequence of independent and identically distributed (iid) random variables from some probability space to $(\Xi, \mathscr{G})$ with the probability distribution μ. Assuming that $B(s)$ is single-valued operator defined on the whole E, we examine the stochastic version of the Forward-Backward algorithm

$$x_{n+1} = (I + \gamma A(\xi_{n+1}))^{-1}(I - \gamma B(\xi_{n+1}))x_n\,, \quad \gamma > 0\,. \tag{2}$$

Our aim is to study the dynamical behavior of this algorithm in the limit of the small steps γ, where the effect of the noise due to the ξ_n will be smoothened.

To give an application example for this algorithm, let us consider again the minimization problem of the sum $F + G$, and let us assume that these functions are unknown to the observer (or difficult to compute), and are written as $F(x) = \mathbb{E}_{\xi_1} f(\xi_1, x)$ and $G(x) = \mathbb{E}_{\xi_1} g(\xi_1, x)$. When the functions f and g are known with $f(\xi_1, \cdot)$ being convex differentiable, and $g(\xi_1, \cdot)$ being convex, proper, and lsc, and when an iid sequence (ξ_n) is available, we can approximatively solve the minimization problem of $F + G$ by resorting to the stochastic proximal gradient algorithm $x_{n+1} = \text{prox}_{\gamma g(\xi_{n+1}, \cdot)}(x_n - \gamma \nabla_x f(\xi_{n+1}, x_n))$. Similar algorithms has been studied in [12, 38] with the additional assumption that the step size γ vanishes as n tends to infinity. The main asset of such vanishing step size algorithms is that the iterates (with or without averaging) converge almost surely as the iteration index goes to infinity. This paper focuses on the case where the step size γ is fixed w.r.t. n. As we shall see below, convergence hold in a weaker sense in this case. Loosely speaking, the iterates fluctuate in a small neighborhood of the set of sought solutions, but do not converge in an almost sure sense as $n \to \infty$. Yet, constant step size algorithms have raised a great deal of attention in the signal processing and machine learning literature ([21]). First, they are known to reach a neighborhood of the solution in a fewer number of iterations than the decreasing step algorithms. Second, they are in practice able to adapt to non stationary or slowly changing environments, and thus track a possible changing set of solutions. This is particularly helpful in adaptive signal processing for instance.

In order to study the dynamical behavior of (2), we introduce the operators

$$\mathcal{A} = \int A(s)\, \mu(ds) \quad \text{and} \quad \mathcal{B} = \int B(s)\, \mu(ds)\,,$$

where the first integral is a set-valued integral, which is to be recognized as a *selection integral* [30]. Assuming that the monotone operator $\mathcal{A} + \mathcal{B}$ is maximal, it is a standard fact of the monotone operator theory that for any x_0 in the domain of $\mathcal{A} + \mathcal{B}$, the Differential Inclusion (DI)

$$\begin{cases} \dot{\mathsf{x}}(t) & \in & -(\mathcal{A} + \mathcal{B})(\mathsf{x}(t)) \\ \mathsf{x}(0) & = & x_0 \end{cases} \tag{3}$$

admits a unique absolutely continuous solution on $\mathbb{R}_+ := [0, \infty)$ [15, 3]. Let $\mathsf{x}_\gamma(t)$ be the continuous random process obtained by assuming that the iterates x_n are distant apart by the time step γ, and by interpolating linearly these iterates. Then, the first step of the approach undertaken in this paper is to show that x_γ shadows the solution of the DI for small γ, in the sense that it converges narrowly to this solution as $\gamma \to 0$ in the topology of convergence on the compact sets of $\mathbb{R}_+$. The same idea is behind the so-called ODE method which is frequently used in the stochastic approximation literature [7, 28].

The compact convergence alone is not enough to control the long term behavior of the iterates. A stability result is needed. To that end, the second step of the approach is to view the sequence

(x_n) as a homogeneous Feller Markov chain whose transition kernel is parameterized by γ. In this context, the aim is to show that the set of invariant measures for this kernel is non empty, and that the family of invariant measures obtained for all γ belonging to some interval $(0, \gamma_0]$ is tight. We shall obtain a general tightness criterion which will be made more explicit in a number of situations of interest involving random maximal monotone operators.

The narrow convergence of x_γ, together with the tightness of the Markov chain invariant measures, lead to the invariance of the small γ cluster points of these invariant measures with respect to the semiflow induced by the DI (3) (see [24, 23, 8] for similar contexts). Using these results, it becomes possible to characterize the long run behavior of the iterates (x_n). In particular, the proximity of these iterates to the set of zeros $Z(\mathcal{A}+\mathcal{B})$ of $\mathcal{A}+\mathcal{B}$ is of obvious interest. First, we show that when the operator $\mathcal{A}+\mathcal{B}$ is *demipositive* [16], the probabilities that the iterates are away from $Z(\mathcal{A}+\mathcal{B})$ are small in Cesàro mean. Whether $\mathcal{A}+\mathcal{B}$ is demipositive or not, we can also characterize the ergodic behavior of the algorithm, showing that when γ is small, the partial sums $n^{-1}\sum_1^n x_k$ stay close to $Z(\mathcal{A}+\mathcal{B})$ with a high probability.

Stochastic approximations with differential inclusions were considered in [9] and in [22] from the dynamical systems viewpoint. The case where the DI is defined by a maximal monotone operator was studied in [11], [12], and [38]. Instances of the random proximal gradient algorithm were treated in *e.g.*, [1] or [37]. All these references dealt with the decreasing step case, which requires quite different tools from the constant step case. This case is considered in [19] (see also [18]), which relies on a Robbins-Siegmund like approach requiring summability assumptions on the random errors. The constant step case is also dealt with in [39] and in [13] for generic differential inclusions. In the present work, we follow the line of reasoning of our paper [13], noting that the case where the DI is defined by a maximal monotone operator has many specificities. For instance, a maximal monotone operator is not upper semi continuous in general, as it was assumed for the differential inclusions studied in [39] and [13]. Another difference lies in the fact that we consider here the case where the domains of the operators $A(s)$ can be different. Finally, the tightness criterion for the Markov chain invariant measures requires a quite specific treatment in the context of the maximal monotone operators.

We close this paragraph by mentioning [10], where one of the studied stochastic proximal gradient algorithms can be cast in the general framework of (2).

Paper organization. Section 2 introduces the main algorithm and recalls some known facts about random monotone operators and their selection integrals. Section 3 provides our assumptions and states our main result about the long run behavior of the iterates. A brief sketch of the proof is also provided for convenience, the detailed arguments being postponed to the end of the paper. Section 4 provides some illustrations of our results in particular cases. The monotone operators involved are assumed to be subdifferentials, hence covering the context of numerical optimization. Our assumptions are discussed at length in this scenario. The case when the monotone operators are linear maps is addressed as well. Section 5 analyzes the dynamical behavior of the iterates. It is shown that the piecewise linear interpolation of the iterates converges narrowly, uniformly on compact sets, to a solution to the DI. The result, which has its own interest, is the first key argument to establish the main Theorem of Section 3. The second argument is provided in Section 6, where we characterize the cluster points of the invariant mesures (indexed by the step size) of the Markov chain formed by the iterates. The appendices A and B are devoted to the proofs relative to Sections 4 and 5 respectively.

2 Background and problem statement

2.1 Basic facts on maximal monotone operators

We start by recalling some basic facts related with the maximal monotone operators on E and with their associated differential inclusions. These facts will be used in the proofs without mention. For more details, the reader is referred to the treatises [15], [3], or [6], or to the tutorial paper [32].

Consider a set valued mapping $\mathsf{A} : E \rightrightarrows E$, *i.e.*, for each $x \in E$, $\mathsf{A}(x)$ is a subset of E. The domain and the graph of A are the respective subsets of E and $E \times E$ defined as $\mathrm{dom}(\mathsf{A}) := \{x \in E : \mathsf{A}(x) \neq \emptyset\}$, and $\mathrm{gr}(\mathsf{A}) := \{(x, y) \in E \times E : y \in \mathsf{A}(x)\}$. The operator A is proper if $\mathrm{dom}(\mathsf{A}) \neq \emptyset$. The operator A is said to be monotone if $\forall x, x' \in \mathrm{dom}(\mathsf{A})$, $\forall y \in \mathsf{A}(x), \forall y' \in \mathsf{A}(x')$, it holds that $\langle y - y', x - x' \rangle \geq 0$. A proper monotone operator A is said maximal if its graph $\mathrm{gr}(\mathsf{A})$ is a maximal element in the inclusion ordering among graphs of monotone operators.

Denote by I the identity operator, and by A^{-1} the inverse of the operator A, defined by the fact that $(x, y) \in \mathrm{gr}(\mathsf{A}^{-1}) \Leftrightarrow (y, x) \in \mathrm{gr}(\mathsf{A})$. It is well known that A belongs to the set $\mathscr{M}$ of the maximal monotone operators on E if and only if, for all $\gamma > 0$, the so called resolvent operator $\mathsf{J}_\gamma := (I + \gamma \mathsf{A})^{-1}$ is a contraction defined on the whole space E (in particular, J_γ is single valued). We also know that when $\mathsf{A} \in \mathscr{M}$, the closure $\mathrm{cl}(\mathrm{dom}(\mathsf{A}))$ of $\mathrm{dom}(\mathsf{A})$ is convex, and $\lim_{\gamma \to 0} \mathsf{J}_\gamma(x) = \Pi_{\mathrm{cl}(\mathrm{dom}(\mathsf{A}))}(x)$, where Π_S is the projector on the closed convex set S. It holds that $\mathsf{A}(x)$ is closed and convex for all $x \in \mathrm{dom}(\mathsf{A})$. We can therefore put $\mathsf{A}_0(x) = \Pi_{\mathsf{A}(x)}(0)$, in other words, $\mathsf{A}_0(x)$ is the minimum norm element of $\mathsf{A}(x)$. Of importance is the so called Yosida regularization of A for $\gamma > 0$, defined as the single-valued operator $\mathsf{A}_\gamma = (I - \mathsf{J}_\gamma)/\gamma$. This is a $1/\gamma$-Lipschitz operator on E that satisfies $\mathsf{A}_\gamma(x) \to \mathsf{A}_0(x)$ and $\|\mathsf{A}_\gamma(x)\| \uparrow \|\mathsf{A}_0(x)\|$ for all $x \in \mathrm{dom}(\mathsf{A})$. One can also check that $\mathsf{A}_\gamma(x) \in \mathsf{A}(\mathsf{J}_\gamma(x))$ for all $x \in E$.

A typical maximal monotone operator is the subdifferential ∂f of a function $f \in \Gamma_0$, the set of proper, convex, and lsc functions on E. In this case, the resolvent $(I + \gamma \partial f)^{-1}$ for $\gamma > 0$ is the well known proximity operator of γf, and is denoted as $\mathrm{prox}_{\gamma f}$. The Yosida regularization of ∂f for $\gamma > 0$ coincides with the gradient of the so-called Moreau's envelope $f_\gamma(x) := \min_w (f(w) + \|w - x\|^2/(2\gamma))$ of f.

2.2 Set valued integrals and random maximal monotone operators

Let $(\Xi, \mathscr{G}, \mu)$ be a probability space where the σ-field $\mathscr{G}$ is μ-complete. For any Euclidean space E, denote as $\mathscr{B}(E)$ the Borel field of E, and let $F : \Xi \rightrightarrows E$ be a set valued function such that $F(s)$ is a closed set for each $s \in \Xi$. The function F is said *measurable* if $\{s : F(s) \cap H \neq \emptyset\} \in \mathscr{G}$ for any set $H \in \mathscr{B}(E)$. An equivalent definition for the mesurability of F requires that the domain $\mathrm{dom}(F) := \{s \in \Xi : F(s) \neq \emptyset\}$ of F belongs to $\mathscr{G}$, and that there exists a sequence of measurable functions $\varphi_n : \mathrm{dom}(F) \to E$ such that $F(s) = \mathrm{cl}\,\{\varphi_n(s)\}_n$ for all $s \in \mathrm{dom}(F)$ [17, Chap. 3] [25].

Assume now that F is measurable and that $\mu(\mathrm{dom}(F)) = 1$. Given $1 \leq p < \infty$, let $\mathcal{L}^p(\Xi, \mathscr{G}, \mu; E)$ be the Banach space of the $\mathscr{G}$-measurable functions $\varphi : \Xi \to E$ such that $\int \|\varphi\|^p d\mu < \infty$, and let

$$\mathfrak{S}^p_F := \{\varphi \in \mathcal{L}^p(\Xi, \mathscr{G}, \mu; E) \,:\, \varphi(s) \in F(s) \ \mu - \text{a.e.}\} .$$

If $\mathfrak{S}^1_F \neq \emptyset$, the function F is said integrable. The *selection integral* [30] of F is the set

$$\int F d\mu := \mathrm{cl} \left\{ \int_\Xi \varphi d\mu \ : \ \varphi \in \mathfrak{S}^1_F \right\}.$$

Now, consider a function $A : \Xi \to \mathscr{M}$. By the maximality of $A(s)$, the graph $\mathrm{gr}(A(s))$ of $A(s)$ is a closed subset of $E \times E$ [15]. For any $\gamma > 0$, denote by $J_\gamma(s, \cdot) := (I + \gamma A(s))^{-1}(\cdot)$ the resolvent of $A(s)$. Assume that the function $s \mapsto \mathrm{gr}(A(s))$ is measurable as a closed set-valued $\Xi \rightrightarrows E \times E$ function. As shown in [2, Ch. 2], this is equivalent to saying that the function $s \mapsto J_\gamma(s, x)$ is measurable from Ξ to E for any $\gamma > 0$ and any $x \in E$. Observe that since $J_\gamma(s, x)$ is measurable in s and continuous in x (being non expansive), $J_\gamma : \Xi \times E \to E$ is $\mathscr{G} \otimes \mathscr{B}(E)/\mathscr{B}(E)$ measurable by Carathéodory's theorem. Denoting by $D(s)$ the domain of $A(s)$, the measurability of $s \mapsto \mathrm{gr}(A(s))$ implies that the set-valued function $s \mapsto \mathrm{cl}(D(s))$ is measurable, which implies that the function $s \mapsto d(x, D(s))$ is measurable for each $x \in E$, where $d(x, S)$ is the distance between the point x and the set S. Denoting as $A(s, x)$ the image of x by the operator $A(s)$, the measurability of the set valued function $s \mapsto A(s, x)$ for each $x \in E$ is another consequence of the measurability of $s \mapsto \mathrm{gr}(A(s))$. In particular, the function $s \mapsto A_0(s, x)$ is measurable for each $x \in E$, where $A_0(s, x) := \Pi_{A(s,x)}(0)$.

The essential intersection $\mathcal{D}$ of the domains $D(s)$ is defined as [27]

$$\mathcal{D} := \bigcup_{G\in\mathscr{G}:\mu(G)=0} \ \bigcap_{s\in\Xi\setminus G} D(s)\,,$$

in other words, $x \in \mathcal{D} \;\Leftrightarrow\; \mu(\{s : x \in D(s)\}) = 1$. Let us assume that $\mathcal{D} \neq \emptyset$, and that the set-valued mapping $A(\cdot, x)$ is integrable for each $x \in \mathcal{D}$. For all $x \in \mathcal{D}$, we can define

$$\mathcal{A}(x) := \int_\Xi A(s,x)\,\mu(ds)\,.$$

One can immediately see that the operator $\mathcal{A} : \mathcal{D} \rightrightarrows E$ so defined is a monotone operator.

2.3 Differential inclusion involving maximal monotone operators

We now turn to the differential inclusions induced by maximal monotone operators. Given $\mathsf{A} \in \mathscr{M}$ and $x_0 \in \operatorname{dom}(\mathsf{A})$, the DI $\dot{\mathsf{x}}(t) \in -\mathsf{A}(\mathsf{x}(t))$ on $\mathbb{R}_+$ with $\mathsf{x}(0) = x_0$ has a unique solution, *i.e.*, a unique absolutely continuous mapping $\mathsf{x} : \mathbb{R}_+ \to E$ such that $\mathsf{x}(0) = x_0$, and $\dot{\mathsf{x}}(t) \in -\mathsf{A}(\mathsf{x}(t))$ for almost all $t > 0$.

Consider the map $\Phi : \operatorname{dom}(\mathsf{A}) \times \mathbb{R}_+ \to \operatorname{dom}(\mathsf{A})$, $(x_0, t) \mapsto \mathsf{x}(t)$ where $\mathsf{x}(t)$ is the DI solution with initial value x_0. Then, Φ satisfies $\|\Phi(x,t) - \Phi(y,t)\| \le \|x - y\|$ for all $t \ge 0$ and all $x, y \in \operatorname{dom}(\mathsf{A})$. Since E is complete, Φ can be extended to a map from $\operatorname{cl}(\operatorname{dom}(\mathsf{A})) \times \mathbb{R}_+$ to $\operatorname{cl}(\operatorname{dom}(\mathsf{A}))$. This extension that we still denote as Φ is a semiflow on $\operatorname{cl}(\operatorname{dom}(\mathsf{A})) \times \mathbb{R}_+$, being a continuous $\operatorname{cl}(\operatorname{dom}(\mathsf{A})) \times \mathbb{R}_+ \to \operatorname{cl}(\operatorname{dom}(\mathsf{A}))$ function satisfying $\Phi(\cdot, 0) = I$, and $\Phi(x, t+s) = \Phi(\Phi(x,s), t)$ for each $x \in \operatorname{cl}(\operatorname{dom}(\mathsf{A}))$, and $t, s \ge 0$.

The set of zeros $Z(\mathsf{A}) := \{x \in \operatorname{dom}(\mathsf{A}) : 0 \in \mathsf{A}(x)\}$ of A is a closed convex set which coincides with the set of equilibrium points $\{x \in \operatorname{cl}(\operatorname{dom}(\mathsf{A})) : \forall t \ge 0, \Phi(x,t) = x\}$ of Φ. The trajectories $\Phi(x, \cdot)$ of the semiflow do not necessarily converge to $Z(\mathsf{A})$ (see [32] for a counterexample). However, the ergodic theorem for the semiflows generated by the elements of $\mathscr{M}$ states that if $Z(\mathsf{A}) \neq \emptyset$, then for each $x \in \operatorname{cl}(\operatorname{dom}(\mathsf{A}))$, the averaged function

$$\begin{array}{rccc} \overline{\Phi} & : & \operatorname{cl}(\operatorname{dom}(\mathsf{A})) \times \mathbb{R}_+ & \longrightarrow & \operatorname{cl}(\operatorname{dom}(\mathsf{A})) \\ & & (x,t) & \longmapsto & \dfrac{1}{t}\displaystyle\int_0^t \Phi(x,s)\,ds \end{array}$$

(with $\overline{\Phi(\cdot,0)} = \Phi(\cdot,0)$), converges to an element of $Z(\mathsf{A})$ as $t \to \infty$. The convergence of the trajectories of the semiflow itself to an element of $Z(\mathsf{A})$ is ensured when A is demipositive [16]. An operator $\mathsf{A} \in \mathscr{M}$ is said demipositive if there exists $w \in Z(\mathsf{A})$ such that for every sequence $((u_n, v_n) \in \operatorname{gr}(\mathsf{A}))$ such that (u_n) converges to u, and such that (v_n) is bounded,

$$\langle u_n - w, v_n \rangle \xrightarrow[n\to\infty]{} 0 \quad \Rightarrow \quad u \in Z(\mathsf{A}).$$

Under this condition and if $Z(\mathsf{A}) \neq \emptyset$, then for all $x \in \operatorname{cl}(\operatorname{dom}(\mathsf{A}))$, $\Phi(x,t)$ converges as $t \to \infty$ to an element of $Z(\mathsf{A})$.

We recall some of the most important notions related with the dynamical behavior of the semiflow Φ. Denote as $\mathcal{M}(E)$ the space of probability measures on E equipped with its Borel σ-field $\mathscr{B}(E)$. An element $\pi \in \mathcal{M}(E)$ is called an invariant measure for Φ if $\pi = \pi\Phi(\cdot,t)^{-1}$ for every $t > 0$. The set of invariant measures for Φ will be denoted $\mathcal{I}(\Phi)$. The limit set of the trajectory $\Phi(x, \cdot)$ of the semiflow Φ starting at x is the set

$$L_{\Phi(x,\cdot)} := \bigcap_{t\ge 0} \operatorname{cl}\left(\Phi(x, [t, \infty))\right)$$

of the limits of the convergent subsequences $(\Phi(x, t_k))_k$ as $t_k \to \infty$. A point $x \in \operatorname{cl}(\operatorname{dom}\mathsf{A})$ is said recurrent if $x \in L_{\Phi(x,\cdot)}$. The Birkhoff center BC_Φ of Φ is

$$\mathrm{BC}_\Phi := \operatorname{cl}\left\{x \in \operatorname{cl}(\operatorname{dom}\mathsf{A}) : x \in L_{\Phi(x,\cdot)}\right\},$$

i.e., the closure of the set of recurrent points of Φ. The celebrated Poincaré's recurrence theorem [20, Th. II.6.4 and Cor. II.6.5] says that the support of any $\pi \in \mathcal{I}(\Phi)$ is a subset of BC_Φ.

Proposition 2.1. Assume that $Z(\mathsf{A}) \neq \emptyset$, and let $\pi \in \mathcal{I}(\Phi)$. If A is demipositive, then $\mathrm{supp}(\pi) \subset Z(\mathsf{A})$. If π has a first moment, then, whether A is demipositive or not,

$$\int x\, \pi(dx) \in Z(\mathsf{A}) \,.$$

Proof. When A is demipositive, $Z(\mathsf{A})$ coincides straightforwardly with BC_Φ, and the first inclusion follows from Poincaré's recurrence theorem.

To show the second result, we start by proving that $\{\overline{\Phi}(\cdot, t) : t > 0\}$ is uniformly integrable as a family of random variables in $(E, \mathscr{B}(E), \pi)$. Let $\varepsilon > 0$. Since the family $\{\Phi(\cdot, t) : t \geq 0\}$ is identically distributed, it is uniformly integrable, thus, there exists $\eta_\varepsilon > 0$ such that $\sup_t \int_S \|\Phi(x, t)\|\, \pi(dx) \leq \varepsilon$ for all $S \in \mathscr{B}(E)$ satisfying $\pi(S) \leq \eta_\varepsilon$. By Tonelli's theorem,

$$\sup_{t>0} \int_S \|\overline{\Phi}(x, t)\|\, \pi(dx) \leq \sup_{t>0} \frac{1}{t} \int_0^t \int_S \|\Phi(x, s)\|\, \pi(dx) ds \leq \varepsilon \,,$$

which shows that, indeed, $\{\overline{\Phi}(\cdot, t) : t > 0\}$ is uniformly integrable [31, Prop. II-5-2]. By the ergodic theorem for semiflows generated by elements of $\mathscr{M}$, there exists a function $f : \mathrm{cl}(\mathrm{dom}\, \mathsf{A}) \to Z(\mathsf{A})$ such that $\overline{\Phi}(\cdot, t) \to f$ as $t \to \infty$. Since

$$\int x\, \pi(dx) = \int \overline{\Phi}(x, t)\, \pi(dx) \quad \text{for all } t \geq 0 \,,$$

we can make $t \to \infty$ and use the uniform integrability of $\{\overline{\Phi}(\cdot, t) : t > 0\}$ to obtain that $\int \|f\|\, d\pi < \infty$, and $\int x\, \pi(dx) = \int f(x)\, \pi(dx)$. The result follows from the closed convexity of $Z(\mathsf{A})$. □

2.4 Presentation of the stochastic Forward-Backward algorithm

Let $B : \Xi \times E \to E$ be a mapping such that $B(\cdot, x)$ is $\mathscr{G}$-measurable for all $x \in E$, and $B(s, \cdot)$ is continuous and monotone (seen as a single-valued operator) on E. By Carathéodory's theorem, B is $\mathscr{G} \otimes \mathscr{B}(E)$-measurable. Furthermore, since $B(s, \cdot)$ is continuous on E, this monotone operator is maximal [15, Prop. 2.4]. We also assume that the mapping $B(\cdot, x) : \Xi \to E$ is integrable for all $x \in E$, and we set $\mathcal{B}(x) := \int B(s, x) \mu(ds)$. Note that $\mathrm{dom}\, \mathcal{B} = E$.

Let (ξ_n) be an i.i.d. sequence of random variables from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\Xi, \mathscr{G})$ with the distribution μ. Let x_0 be a E-valued random variable with probability law ν, and assume that x_0 and (ξ_n) are independent. Starting from x_0, our purpose is to study the behavior of the iterates

$$x_{n+1} = J_\gamma(\xi_{n+1}, x_n - \gamma B(\xi_{n+1}, x_n)), \quad n \in \mathbb{N} \,, \tag{4}$$

for a given $\gamma > 0$, where we recall the notation $J_\gamma(s, \cdot) := (I + \gamma A(s))^{-1}(\cdot)$ for every $s \in \Xi$.

In the deterministic case where the functions $A(s, \cdot)$ and $B(s, \cdot)$ are replaced with deterministic maximal monotone operators $\mathsf{A}(\cdot)$ and $\mathsf{B}(\cdot)$, with B still being assumed single-valued with $\mathrm{dom}(\mathsf{B}) = E$, the algorithm coincides with the well-known Forward-Backward algorithm (1). Assuming that B is so-called cocoercive and that γ is not too large, the iterates given by (1) are known to converge to an element of $Z(\mathsf{A} + \mathsf{B})$, provided this set is not empty [6, Th. 25.8]. In the stochastic case who is of interest here, this convergence does not hold in general. Nonetheless, we shall show below that in the long run, the probability that the iterates or their empirical means stay away of $Z(\mathcal{A} + \mathcal{B})$ is small when γ is close to zero.

3 Assumptions and main results

We first observe that the process (x_n) described by Eq. (4) is a homogeneous Markov chain whose transition kernel P_γ is defined by the identity

$$P_\gamma(x, f) = \int f(J_\gamma(s, x - \gamma B(s, x)))\, \mu(ds) \,, \tag{5}$$

valid for each measurable and positive function f. The kernel P_γ and the initial measure ν determine completely the probability distribution of the process (x_n), seen as a $(\Omega, \mathscr{F}) \to (E^{\mathbb{N}}, \mathscr{B}(E)^{\otimes\mathbb{N}})$ random variable. We shall denote this probability distribution on $(E^{\mathbb{N}}, \mathscr{B}(E)^{\otimes\mathbb{N}})$ as $\mathbb{P}^{\nu,\gamma}$. We denote by $\mathbb{E}^{\nu,\gamma}$ the corresponding expectation. When $\nu = \delta_a$ for some $a \in E$, we shall prefer the notations $\mathbb{P}^{a,\gamma}$ and $\mathbb{E}^{a,\gamma}$ to $\mathbb{P}^{\delta_a,\gamma}$ and $\mathbb{E}^{\delta_a,\gamma}$. From nom on, (x_n) will denote the canonical process on the canonical space $(E^{\mathbb{N}}, \mathscr{B}(E)^{\otimes\mathbb{N}})$.

We denote as $\mathscr{F}_n$ the sub-σ-field of $\mathscr{F}$ generated by the family $\{x_0, \{\xi_k^\gamma : 1 \le k \le n\}\}$, and we write $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \,|\, \mathscr{F}_n]$ for $n \in \mathbb{N}$.

In the remainder of the paper, C will always denote a positive constant that does not depend on the time n nor on γ. This constant may change from a line of calculation to another. In all our derivations, γ will lie in the interval $(0, \gamma_0]$ where γ_0 is a fixed constant which is chosen as small as needed.

3.1 Assumptions

Assumption 3.1. For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

$$\sup_{x \in \mathcal{K} \cap \mathcal{D}} \int \|A_0(s, x)\|^{1+\varepsilon} \, \mu(ds) < \infty.$$

Assumption 3.2. The monotone operator $\mathcal{A}$ is maximal.

Assumption 3.3. For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

$$\sup_{x \in \mathcal{K}} \int \|B(s, x)\|^{1+\varepsilon} \, \mu(ds) < \infty \,.$$

The next assumption will mainly lead to the tightness of the invariant measures mentioned in the introduction.

We know that a point $x_\star$ is an element of $Z(\mathcal{A} + \mathcal{B})$ if there exists $\varphi \in \mathfrak{S}^1_{A(\cdot, x_\star)}$ such that $\int \varphi(s) \, \mu(ds) + \int B(s, x_\star) \, \mu(ds) = 0$. When $B(\cdot, x_\star) \in \mathcal{L}^2(\Xi, \mathscr{G}, \mu; E)$, and when the above function φ can be chosen in $\mathcal{L}^2(\Xi, \mathscr{G}, \mu; E)$, we say that such a zero admits a $\mathcal{L}^2$ representation (φ, B). In this case, we define

$$\begin{aligned}
\psi_\gamma(x) := \int \Big\{ &\langle A_\gamma(s, x - \gamma B(s, x)) - \varphi(s), J_\gamma(s, x - \gamma B(s, x)) - x_\star \rangle \\
&+ \langle B(s, x) - B(s, x_\star), x - x_\star \rangle \Big\} \, \mu(ds) \\
+ \gamma \int &\|A_\gamma(s, x - \gamma B(s, x))\|^2 \mu(ds) - 6\gamma \int \|B(s, x) - B(s, x_\star)\|^2 \mu(ds) \,, \qquad (6)
\end{aligned}$$

where

$$A_\gamma(s, x) := \frac{x - J_\gamma(s, x)}{\gamma}$$

is the Yosida regularization of $A(s, x)$ for $\gamma > 0$.

Assumption 3.4. There exists $x_\star \in Z(\mathcal{A} + \mathcal{B})$ admitting a $\mathcal{L}^2$ representation (φ, B). The function $\Psi(x) := \inf_{\gamma \in (0, \gamma_0]} \psi_\gamma(x)$ satisfies one of the following properties:

(a) $\displaystyle \liminf_{\|x\| \to \infty} \frac{\Psi(x)}{\|x\|} > 0$.

(b) $\displaystyle \frac{\Psi(x)}{\|x\|} \xrightarrow[\|x\| \to \infty]{} \infty$.

(c) $\displaystyle \liminf_{\|x\| \to \infty} \frac{\Psi(x)}{\|x\|^2} > 0$.

Let us comment these assumptions.

Assumptions 3.1 and 3.3 are moment assumptions on $A_0(s,x)$ and $B(s,x)$ that are usually easy to check. Assumption 3.1 implies that for every $x \in \mathcal{D}$, $A_0(\,.\,,x)$ is integrable. Therefore, $A(\,.\,,x)$ is integrable. This implies that the domain of the selection integral $\mathcal{A}$ coincides with $\mathcal{D}$.

Conditions where Assumption 3.2 are satisfied can be found in [15, Chap. II.6] in the case where μ has a finite support, and in [12, Prop. 3.1] in other cases. When $A(s)$ is the subdifferential of a function $g(s,\cdot)$ belonging to Γ_0, the maximality of $\mathcal{A}$ is established if we can exchange the expectation of $g(\xi_1,x)$ w.r.t. ξ_1 with the subdifferentiation w.r.t. x, in which case $\mathcal{A}$ would be equal to ∂G, where $G(x) = \int g(s,x)\,\mu(ds)$. This problem is dealt with in [40] (see also Sec. 4.1 below).

The first role of Assumption 3.4 is to ensure the tightness of the invariant measures of the kernels P_γ, as mentioned in the introduction. Beyond the tightness, this assumption controls the asymptotic behavior of functionals of the iterates with a prescribed growth condition at infinity. Assumption 3.4 will be specified and commented at length in Section 4.

Regarding the domains of the operators $A(s)$, two cases will be considered, according to whether these domains vary with s or not. We shall name these two cases the "common domain" case and the "different domains" case respectively. In the common domain case, our assumption is therefore:

Assumption 3.5 (Common domain case)**.** The set-valued function $s \mapsto D(s)$ is μ-almost everywhere constant.

In the common domain case, Assumptions 3.1–3.4 will be sufficient to state our results, whereas in the different domains case, three supplementary assumptions will be needed:

Assumption 3.6 (Different domains case)**.** $\forall x \in E$, $\displaystyle\int d(x,D(s))^2\,\mu(ds) \geq C\boldsymbol{d}(x)^2$, where $\boldsymbol{d}(\cdot)$ is the distance function to $\mathcal{D}$.

Assumption 3.7 (Different domains case)**.** For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

$$\sup_{\gamma\in(0,\gamma_0],x\in\mathcal{K}} \frac{1}{\gamma^{1+\varepsilon}} \int \|J_\gamma(s,x) - \Pi_{\mathrm{cl}(D(s))}(x)\|^{1+\varepsilon}\,\mu(ds) < \infty\,.$$

Assumption 3.8 (Different domains case)**.** For all $\gamma \in (0,\gamma_0]$ and all $x \in E$,

$$\int \left(\frac{\|J_\gamma(s,x) - \Pi_{\mathrm{cl}(D(s))}(x)\|}{\gamma} + \|B(s,x)\| \right) \mu(ds) \leq C(1+\psi_\gamma(x))\,.$$

Assumption 3.6 is rather mild, and is easy to illustrate in the case where μ is a finite sum of Dirac measures. Following [5], we say that a finite collection of closed and convex subsets $\{\mathcal{C}_1,\ldots,\mathcal{C}_m\}$ over E is *linearly regular* if there exists $\kappa > 0$ such that for every x,

$$\max_{i=1\ldots m} d(x,\mathcal{C}_i) \geq \kappa d(x,\mathcal{C}), \quad \text{where } \mathcal{C} = \bigcap_{i=1}^m \mathcal{C}_i\,,$$

and where implicitly $\mathcal{C} \neq \emptyset$. Sufficient conditions for a collection of sets to satisfy the above condition can be found in [5] and the references therein.

We know that when $\gamma \to 0$, $J_\gamma(s,x)$ converges to $\Pi_{\mathrm{cl}(D(s))}(x)$ for each (s,x). Assumptions 3.7 and 3.8 add controls on the convergence rate. The instantiations of these assumptions in the case of the stochastic proximal gradient algorithm will be provided in Section 4.1 below.

3.2 Main result

Lemma 3.1. Let Assumptions 3.2 and 3.3 hold true. Then, the monotone operator $\mathcal{A} + \mathcal{B}$ is maximal.

Proof. Assumption 3.3 implies that the monotone operator $\mathcal{B}$ is continuous on E. Therefore, $\mathcal{B}$ is maximal [15, Prop. 2.4]. The maximality of $\mathcal{A}+\mathcal{B}$ follows, since $\mathcal{A}$ is maximal by Assumption 3.2, and $\mathcal{B}$ has a full domain [15, Cor. 2.7]. □

Note that $\mathrm{dom}(\mathcal{A}+\mathcal{B}) = \mathcal{D}$. In the remainder of the paper, we denote as $\Phi : \mathrm{cl}(\mathcal{D})\times\mathbb{R}_+ \to \mathrm{cl}(\mathcal{D})$ the semiflow produced by the DI $\dot{\mathsf{x}}(t) \in -(\mathcal{A}+\mathcal{B})(\mathsf{x}(t))$. Recall that $\mathcal{I}(\Phi)$ is the set of invariant measures for the semiflow Φ.

We also write

$$\bar{x}_n := \frac{1}{n+1}\sum_{k=0}^{n} x_k\,.$$

We now state our main theorem.

Theorem 3.2. Let Assumptions 3.1, 3.2, 3.3, and 3.4–(a) be satisfied. Moreover, assume that either Assumption 3.5 or Assumptions 3.6–3.8 are satisfied.

Then, $\mathcal{I}(\Phi) \neq \emptyset$. Let $\nu \in \mathcal{M}(E)$ be with a finite second moment, and let $\mathcal{U} := \bigcup_{\pi\in\mathcal{I}(\Phi)} \mathrm{supp}(\pi)$. Then, for all $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n+1}\sum_{k=0}^{n} \mathbb{P}^{\nu,\gamma}(d(x_k,\mathcal{U}) > \varepsilon) \xrightarrow[\gamma\to 0]{} 0\,. \tag{7}$$

In particular, if the operator $\mathcal{A}+\mathcal{B}$ is demipositive, then

$$\limsup_{n\to\infty} \frac{1}{n+1}\sum_{k=0}^{n} \mathbb{P}^{\nu,\gamma}\left(d(x_k, Z(\mathcal{A}+\mathcal{B})) > \varepsilon\right) \xrightarrow[\gamma\to 0]{} 0\,. \tag{8}$$

Moreover, the set $\{\pi \in \mathcal{I}(\Phi) : \pi(\Psi) < \infty\}$ is not empty. Let $N' \in \mathbb{N}^*$, and let $f : E \to \mathbb{R}^{N'}$ be continuous. Assume that there exists $M \geq 0$ and $\varphi : \mathbb{R}^{N'} \to \mathbb{R}_+$ such that $\lim_{\|a\|\to\infty} \varphi(a)/\|a\| = \infty$, and

$$\forall a \in E,\ \varphi(f(a)) \leq M(1+\Psi(a))\,.$$

Then, for all $n \in \mathbb{N}$, $\gamma \in (0,\gamma_0]$, the r.v.

$$F_n := \frac{1}{n+1}\sum_{k=0}^{n} f(x_k)$$

is $\mathbb{P}$-integrable, and satisfies for all $\varepsilon > 0$,

$$\limsup_{n\to\infty} \mathbb{P}^{\nu,\gamma}\left(d\left(F_n, \mathcal{S}_f\right) \geq \varepsilon\right) \xrightarrow[\gamma\to 0]{} 0\,, \tag{9}$$

$$\limsup_{n\to\infty}\ d\left(\mathbb{E}^{\nu,\gamma}(F_n), \mathcal{S}_f\right) \xrightarrow[\gamma\to 0]{} 0\,. \tag{10}$$

where $\mathcal{S}_f := \{\pi(f) : \pi \in \mathcal{I}(\Phi)\}$. In particular, if $f(x) = x$, and if Assumption 3.4–(b) is satisfied, then

$$\limsup_{n\to\infty}\ \mathbb{P}^{\nu,\gamma}\left(d\left(\bar{x}_n, Z(\mathcal{A}+\mathcal{B})\right) \geq \varepsilon\right) \xrightarrow[\gamma\to 0]{} 0\,, \tag{11}$$

$$\limsup_{n\to\infty}\ d\left(\mathbb{E}^{\nu,\gamma}(\bar{x}_n), Z(\mathcal{A}+\mathcal{B})\right) \xrightarrow[\gamma\to 0]{} 0\,. \tag{12}$$

By Lem. 3.1 and Prop. 2.1, the convergences (8), (11), and (12) are the consequences of (7), (9), and (10) respectively. We need to prove the latter.

3.3 Proof technique

We first observe that the Markov kernels P_γ are Feller, *i.e.*, they take the set $C_b(E)$ of the real, continuous, and bounded functions on E to $C_b(E)$. Indeed, for each $f \in C_b(E)$, Eq. (5) shows that $P_\gamma(\cdot, f) \in C_b(E)$ by the continuity of $J_\gamma(s, \cdot)$ and $B(s, \cdot)$, and by dominated convergence.

For each $\gamma > 0$, we denote as

$$\mathcal{I}(P_\gamma) := \{\pi \in \mathcal{M}(E) \,:\, \pi = \pi P_\gamma\}$$

the set of invariant probability measures of P_γ. Define the family of kernels $\mathcal{P} := \{P_\gamma\}_{\gamma \in (0,\gamma_0]}$, and let

$$\mathcal{I}(\mathcal{P}) := \bigcup_{\gamma \in (0,\gamma_0]} \mathcal{I}(P_\gamma)$$

be the set of distributions π such that $\pi = \pi P_\gamma$ for at least one P_γ with $\gamma \in (0, \gamma_0]$.

The following proposition, which is valid for Feller Markov kernels, has been proven in [13] in the more general context of set-valued differential inclusions.

Proposition 3.3. Let $V : E \to [0, +\infty)$ and $Q : E \to [0, +\infty)$ be measurable. Assume that $Q(x) \to \infty$ as $\|x\| \to \infty$. Assume that for each $\gamma \in (0, \gamma_0]$,

$$P_\gamma(x, V) \le V(x) - \alpha(\gamma)Q(x) + \beta(\gamma)\,, \tag{13}$$

where $\alpha : (0, \gamma_0] \to (0, +\infty)$ and $\beta : (0, \gamma_0] \to \mathbb{R}$ satisfy $\sup_{\gamma \in (0,\gamma_0]} \frac{\beta(\gamma)}{\alpha(\gamma)} < \infty$. Then, the family $\mathcal{I}(\mathcal{P})$ is tight. Moreover, $\sup_{\pi \in \mathcal{I}(\mathcal{P})} \pi(Q) < \infty$.

Assume moreover that, as $\gamma \to 0$, any cluster point of $\mathcal{I}(\mathcal{P})$ is an element of $\mathcal{I}(\Phi)$. In particular, $\{\pi \in \mathcal{I}(\Phi) \,:\, \pi(Q) < \infty\}$ is not empty. Let $\nu \in \mathcal{M}(E)$ s.t. $\nu(V) < \infty$. Let $\mathcal{U} := \bigcup_{\pi \in \mathcal{I}(\Phi)} \operatorname{supp}(\pi)$. Then, for all $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{P}^{\nu,\gamma}(d(x_k, \mathcal{U}) > \varepsilon) \xrightarrow[\gamma\to 0]{} 0\,.$$

Let $N' \in \mathbb{N}^*$ and $f : E \to \mathbb{R}^{N'}$ be continuous. Assume that there exists $M \ge 0$ and $\varphi : \mathbb{R}^{N'} \to \mathbb{R}_+$ such that $\lim_{\|a\|\to\infty} \varphi(a)/\|a\| = \infty$ and

$$\forall a \in E,\ \varphi(f(a)) \le M(1 + Q(a))\,.$$

Then, for all $n \in \mathbb{N}$, $\gamma \in (0, \gamma_0]$, the r.v.

$$F_n := \frac{1}{n+1} \sum_{k=0}^{n} f(x_k)$$

is $\mathbb{P}^{\nu,\gamma}$-integrable, and satisfies for all $\varepsilon > 0$,

$$\limsup_{n\to\infty} d\left(\mathbb{E}^{\nu,\gamma}(F_n)\,, \mathcal{S}_f\right) \xrightarrow[\gamma\to 0]{} 0\,, \quad \text{and} \quad \limsup_{n\to\infty} \mathbb{P}^{\nu,\gamma}\left(d\left(F_n\,, \mathcal{S}_f\right) \ge \varepsilon\right) \xrightarrow[\gamma\to 0]{} 0\,,$$

where $\mathcal{S}_f := \{\pi(f) \,:\, \pi \in \mathcal{I}(\Phi)\}$.

Proof. Assume that Eq. (13) holds. By [13, Prop. 6.7], $\mathcal{I}(\mathcal{P})$ is tight and $\sup_{\pi \in \mathcal{I}(\mathcal{P})} \pi(Q) < \infty$, which proves the first point. Assume moreover that, as $\gamma \to 0$, any cluster point of $\mathcal{I}(\mathcal{P})$ is an element of $\mathcal{I}(\Phi)$. By the tightness of $\mathcal{I}(\mathcal{P})$ and the Prokhorov theorem, such a cluster point π exists, and satisfies $\pi(Q) < \infty$ by the first point just shown. The rest of the proof follows [13, Section 6.4] word-for-word. □

In order to prove Th. 3.2, it is enough to show that the assumptions of Prop. 3.3 are satisfied. Namely, we need to establish (13) and to show that the cluster points of $\mathcal{I}(\mathcal{P})$ as $\gamma \to 0$ are elements of $\mathcal{I}(\Phi)$.

In Sec. 5, we show that the linearly interpolated process constructed from the sequence (x_n) converges narrowly as $\gamma \to 0$ to a DI solution in the topology of uniform convergence on compact sets. The main result of this section is Th. 5.1, which has its own interest. To prove this theorem, we establish the tightness of the linearly interpolated process (Lem. 5.3), then we show that the limit points coincide with the DI solution (Lem. 5.4–5.8). In Sec. 6, we start by establishing the inequality (13), which is shown in Lem. 6.1 with $Q(x) = \Psi(x)$. Using the tightness of $\mathcal{I}(\mathcal{P})$ in conjunction with Th. 5.1, Lem 6.2 shows that the cluster points of $\mathcal{I}(\mathcal{P})$ are elements of $\mathcal{I}(\Phi)$. In the different domains case, this lemma requires that the invariant measures of P_γ put most of their weights in a thickening of the domain $\mathcal{D}$ of order γ. This fact is established by Lem. 6.3.

4 Case studies - Tightness of the invariant measures

Before proving the main results, we first address three important cases: the case of the random proximal gradient algorithm, the case where $A(s)$ is an affine monotone operator and $B(s) = 0$, and the case where $\mathcal{D}$ is bounded. The main problem is to ensure that one of the cases of Assumption 3.4 is verified. We close the section with a general condition ensuring that Assumption 3.4–(a) is verified. The proofs are postponed to Appendix A.

4.1 A random proximal gradient algorithm

Let $(\Sigma, \mathscr{A}, \zeta)$ be a probability space, where $\mathscr{A}$ is ζ-complete. Denoting as epi the epigraph of a function, a function $h : \Sigma \times E \to (-\infty, \infty]$ is called a convex normal integrand [34] if the set-valued mapping $s \mapsto \operatorname{epi} h(s, \cdot)$ is closed-valued and measurable, and if $h(s, \cdot)$ is convex. To simplify the presentation, we furthermore assume that h is finite everywhere, noting that the results can be extended to the case where h can take the value ∞. Observe that the set-valued function $s \mapsto \partial h(s, \cdot)$ is a measurable $\Sigma \to \mathscr{M}$ function in the sense of Section 2.2 [2] (in all what follows, the subdifferential or the gradient of a function in (s, x) will be meant to be taken w.r.t. x). Assume that $\int |h(s, x)| \zeta(ds) < \infty$ for all $x \in E$, and consider the convex function $H(x) := \int h(s, x)\, \zeta(ds)$ defined on E. By *e.g.*, [36, page 179], $\partial H(x) = \int \partial h(s, x)\, \zeta(ds)$.

Let $f : \Sigma \times E \to \mathbb{R}$ be such that $f(\cdot, x)$ is $\mathscr{A}$-measurable for all $x \in E$, and $f(s, \cdot)$ is convex and continuously differentiable for all $s \in \Sigma$. Moreover, assume that $\int |f(s, x)|\, \zeta(ds) < \infty$ for all $x \in E$, and define the function $F(x) := \int f(s, x)\, \zeta(ds)$ on E. This function is differentiable with $\nabla F(x) = \int \nabla f(s, x)\, \zeta(ds)$.

Finally, given $m \in \mathbb{N}^*$, let $\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ be a collection of closed and convex subsets of E. We assume that $\bigcap_{i=1}^m \operatorname{ri}(\mathcal{C}_i) \neq \emptyset$, where ri is the relative interior of a set.

Our purpose is to approximatively solve the optimization problem

$$\min_{x \in \mathcal{C}} F(x) + H(x), \quad \mathcal{C} := \bigcap_{i=1}^m \mathcal{C}_i \tag{14}$$

whether the minimum is attained. Let (u_n) be an iid sequence on Σ with the probability measure ζ. Let (I_n) be an iid sequence on $\{0, 1, \ldots, m\}$ with the probability measure α such that $\alpha(k) = \mathbb{P}(I_1 = k) > 0$ for each k. Assume that (I_n) and (u_n) are independent. In order to solve the problem (14), we consider the iterates

$$x_{n+1} = \begin{cases} \operatorname{prox}_{\alpha(0)^{-1}\gamma h(u_{n+1}, \cdot)}(x_n - \gamma \nabla f(u_{n+1}, x_n)) & \text{if } I_{n+1} = 0, \\ \Pi_{\mathcal{C}_{I_{n+1}}}(x_n - \gamma \nabla f(u_{n+1}, x_n)) & \text{otherwise,} \end{cases} \tag{15}$$

for $\gamma > 0$. This problem can be cast in the general framework of the stochastic proximal gradient algorithm presented in the introduction. On the space $\Xi := \Sigma \times \{0, \ldots, m\}$, define the iid random variables $\xi_n := (u_n, I_n)$ with the measure $\mu := \zeta \otimes \alpha$. Denoting as ι_S the indicator function of the set S, let $g : \Xi \times E \to (-\infty, \infty]$ be defined as

$$g(s, x) := \begin{cases} \alpha(0)^{-1} h(u, x) & \text{if } i = 0, \\ \iota_{\mathcal{C}_i}(x) & \text{otherwise,} \end{cases}$$

where $s = (u, i)$. Then, Problem (14) is equivalent to minimizing the sum $F(x) + G(x)$, where

$$G(x) := \int g(s, x)\, \mu(ds) = \sum_{k=1}^{m} \iota_{\mathcal{C}_k}(x) + H(x)\,.$$

It is furthermore clear that the algorithm (15) is the instance of the general algorithm (4) that corresponds to $A(s) = \partial g(s, \cdot)$ and $B(s) = \nabla f(u, \cdot)$ for $s = (u, i)$. With our assumptions, the qualification conditions hold, and the three sets $\arg\min(F + G)$, $Z(\partial G + \nabla F)$, and $Z(\mathcal{A} + \mathcal{B})$ coincide.

Before going further, we recall some well known facts regarding the coercive functions belonging to Γ_0. A function $q \in \Gamma_0$ is said coercive if $\lim_{\|x\|\to\infty} q(x) = \infty$. It is said supercoercive if $\lim_{\|x\|\to\infty} q(x)/\|x\| = \infty$. The three following conditions are equivalent: i) q is coercive, ii) there exists $a \in \mathbb{R}$ such that the level set $\text{lev}_{\le a}\, q$ is non empty and compact, iii) $\liminf_{\|x\|\to\infty} q(x)/\|x\| > 0$ (see *e.g.*, [6, Prop. 11.11 and 11.12] and [14, Prop. 1.1.5]).

The main result of this paragraph is the following:

Proposition 4.1. Let the following hypotheses hold true:

H1 There exists $x_\star \in Z(\partial G + \nabla F)$ admitting a $\mathcal{L}^2$ representation $(\varphi((u, i)), \nabla f(u, x_\star))$.

H2 There exists $c > 0$ s.t. for every $x \in E$,

$$\int \langle \nabla f(s, x) - \nabla f(s, x_\star), x - x_\star \rangle\, \zeta(ds) \ge c \int \|f(s, x) - f(s, x_\star)\|^2\, \zeta(ds).$$

H3 The function $F + G$ satisfies one of the following properties:

(a) $F + G$ is coercive.

(b) $F + G$ is supercoercive.

Then, Assumption 3.4–(a) (resp., Assumption 3.4–(b)) holds true if Hypothesis H3–(a) (resp., Hypothesis H3–(b)) holds true.

Let us comment these hypotheses. A light condition ensuring the truth of Hypothesis H1 is provided by the following lemma.

Lemma 4.2. Assume that there exists $x_\star \in Z(\partial G + \nabla F)$ satisfying the two following conditions: $\int \|\nabla f(u, x_\star)\|^2\, \zeta(du) < \infty$, and there exists an open neighborhood $\mathcal{N}$ of $x_\star$ such that $\int h(u, x)^2\, \zeta(du) < \infty$ for all $x \in \mathcal{N}$. Then, Hypothesis H1 is verified.

We now turn to Hypothesis H2. When studying the deterministic Forward-Backward algorithm (1), it is standard to assume that B is cocoercive, in other words, that there exists a constant $L > 0$ such that $\langle \mathsf{B}(x) - \mathsf{B}(y), x - y \rangle \ge L\|\mathsf{B}(x) - \mathsf{B}(y)\|^2$ [6, Th. 25.8]. A classical case where this is satisfied is the case where B is the gradient of a convex differentiable function having a $1/L$-Lipschitz continuous gradient, as is shown by the Baillon-Haddad theorem [6, Cor. 18.16]. In our case, if we assume that there exists a nonnegative measurable function $\beta(s)$ such that $\|\nabla f(s, x) - \nabla f(s, x')\| \le \beta(s)\|x - x'\|$, then by the Baillon-Haddad theorem,

$$\langle \nabla f(s, x) - \nabla f(s, x'), x - x' \rangle \ge \frac{1}{\beta(s)} \|\nabla f(s, x) - \nabla f(s, x')\|^2\,.$$

Thus, one obvious case where Hypothesis H2 is satisfied is the case where $\beta(s)$ is bounded.

Using proposition 4.1, we can now obtain the following corollary to Th. 3.2.

Corollary 4.3. Let Hypotheses H1–H3 hold true. Assume in addition the following hypotheses:

C1 For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

$$\sup_{x \in \mathcal{K} \cap \mathcal{C}} \int \|\partial h_0(u,x)\|^{1+\varepsilon} \, \zeta(du) < \infty,$$

where $\partial h_0(u,\cdot)$ is the least norm element of $\partial h(u,\cdot)$.

C2 For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

$$\sup_{x \in \mathcal{K}} \int \|\nabla f(u,x)\|^{1+\varepsilon} \, \zeta(du) < \infty \,.$$

C3 The sets $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are linearly regular.

C4 For all $\gamma \in (0, \gamma_0]$ and all $x \in E$,

$$\int \left(\|\nabla h_\gamma(u,x)\| + \|\nabla f(u,x)\|\right) \zeta(du) \le C(1 + |F(x) + H_\gamma(x)|) \,,$$

where $h_\gamma(u,\cdot)$ is the Moreau envelope of $h(u,\cdot)$.

Then, for each probability measure ν having a finite second moment,

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{P}^{\nu,\gamma}\left(d(x_k, \arg\min(F+G)) > \varepsilon\right) \xrightarrow[\gamma\to 0]{} 0 \,.$$

Moreover, if Hypothesis H3–(b) is satisfied, then

$$\limsup_{n\to\infty} \; \mathbb{P}^{\nu,\gamma}\left(d\left(\bar{x}_n, \arg\min(F+G)\right) \ge \varepsilon\right) \xrightarrow[\gamma\to 0]{} 0, \text{ and}$$
$$\limsup_{n\to\infty} \; d\left(\mathbb{E}^{\nu,\gamma}(\bar{x}_n), \arg\min(F+G)\right) \xrightarrow[\gamma\to 0]{} 0 \,.$$

Proof. With the hypotheses H1–H3 and C1–C4, one can check that the assumptions 3.1–3.8 are verified. Note that $\partial G + \nabla F$ is a demipositive operator, being the subdifferential of a Γ_0 function having a minimizer [16]. The results of the corollary follow from those of Th. 3.2. □

4.2 The case where $A(s)$ is affine

In all the remainder of this section, we shall focus on the validity of Assumption 3.4. We assume that $B = 0$, and that

$$A(s,x) = H(s)x + d(s),$$

where $H : \Xi \to \mathbb{R}^{N\times N}$ and $d : \Xi \to E$ are two $\mathscr{G}$-measurable functions. It is easily seen that the linear operator $A(s)$ is monotone if and only if $H(s) + H(s)^T \ge 0$ in the semidefinite ordering of matrices, a condition that we shall assume in this subsection. Moreover, assuming that

$$\int (\|H(s)\|^2 + \|d(s)\|^2) \, \mu(ds) < \infty,$$

the operator

$$\mathcal{A}(x) = \Big(\int H(s) \, \mu(ds)\Big) x + \int d(s) \, \mu(ds) := \boldsymbol{H} x + \boldsymbol{d}$$

exists and is a maximal monotone operator with the domain E. When $\boldsymbol{d}$ belongs to the image of $\boldsymbol{H}$, $Z(\mathcal{A}) \neq \emptyset$, and every $x_\star \in Z(\mathcal{A})$ has a unique $\mathcal{L}^2$ representation $(\varphi(s) = H(s)x_\star + d(s), 0)$. We have the following proposition:

Proposition 4.4. If $\boldsymbol{H} + \boldsymbol{H}^T > 0$, then $\boldsymbol{H}$ is invertible, $Z(\mathcal{A}) = \{x_\star\}$ with $x_\star = -\boldsymbol{H}^{-1}\boldsymbol{d}$, and and Assumption 3.4–(c) is verified.

4.3 The case where the domain $\mathcal{D}$ is bounded

Proposition 4.5. Let the following hypotheses hold true:

H1 The domain $\mathcal{D}$ is bounded.

H2 There exists a constant $C > 0$ such that

$$\forall x \in E, \int d(s,x)^2 \, \mu(ds) \geq C \boldsymbol{d}(x)^2.$$

H3 There exists $x_\star \in Z(\mathcal{A} + \mathcal{B})$ admitting a $\mathcal{L}^2$ representation.

H4 There exists $c > 0$ s.t. for every $x \in E$, For all γ small enough,

$$\int \langle B(s,x) - B(s,x_\star), x - x_\star \rangle \, \mu(ds) \geq c \int \|B(s,x) - B(s,x_\star)\|^2 \, \mu(ds) \, .$$

Then, Assumption 3.4–(c) is satisfied.

4.4 A case where Assumption 3.4–(a) is valid

We close this section by providing a general condition that guarantees the validity of Assumption 3.4–(a). For simplicity, we focus on the case where $B(s) = 0$, noting that the result can be easily extended to the case where $B(s) \neq 0$ when a cocoercivity hypothesis of the type of Prop. 4.5–H4 is satisfied.

We denote by $\boldsymbol{S}(\rho, d)$ the sphere of E with center ρ and radius d. We also denote by $\operatorname{int} S$ the interior of a set S.

Proposition 4.6. Assume that $B(s) = 0$, and that there exists $x_\star \in Z(\mathcal{A}) \cap \operatorname{int} \mathcal{D}$ admitting a $\mathcal{L}^2$ representation $\varphi \in \mathfrak{S}^2_{A(\cdot, x_\star)}$. Assume that there exists a set $\Sigma \in \mathscr{G}$ such that $\mathcal{D} \subset \cap_{s \in \Sigma} D(s)$, $\mu(\Sigma) > 0$, and such that for all $s \in \Sigma$, there exists $\delta(s) > 0$ satisfying $\boldsymbol{S}(\varphi(s), \delta(s)) \subset \operatorname{int} \mathcal{D}$, and

$$\forall x \in \boldsymbol{S}(\varphi(s), \delta(s)), \quad \inf_{y \in A(s,x)} \langle y - \varphi(s), x - x_\star \rangle > 0.$$

Then, Assumption 3.4–(a) is satisfied.

Note that the inf in the statement of this proposition is attained, as is revealed by the proof.

5 Narrow convergence towards the DI solutions

5.1 Main result

The set $C(\mathbb{R}_+, E)$ of continuous functions from $\mathbb{R}_+$ to E is equipped with the topology of uniform convergence on the compact intervals, who is known to be compatible with the distance d defined as

$$\mathsf{d}(\mathsf{x}, \mathsf{y}) := \sum_{n \in \mathbb{N}^*} 2^{-n} \left(1 \wedge \sup_{t \in [0,n]} \|\mathsf{x}(t) - \mathsf{y}(t)\| \right) .$$

For every $\gamma > 0$, we introduce the measurable map $\mathsf{X}_\gamma : (E^{\mathbb{N}}, \mathscr{B}(E)^{\otimes \mathbb{N}}) \to (C(\mathbb{R}_+, E), \mathscr{B}(C(\mathbb{R}_+, E)))$, defined for every $x = (x_n : n \in \mathbb{N})$ in $E^{\mathbb{N}}$ as

$$\mathsf{X}_\gamma(x) \, : t \mapsto x_{\lfloor \frac{t}{\gamma} \rfloor} + (t/\gamma - \lfloor t/\gamma \rfloor)(x_{\lfloor \frac{t}{\gamma} \rfloor + 1} - x_{\lfloor \frac{t}{\gamma} \rfloor}) \, .$$

This map will be referred to as the linearly interpolated process. When $x = (x_n)$ is the process with the probability measure $\mathbb{P}^{\nu,\gamma}$ defined above, the distribution of the r.v. X_γ is $\mathbb{P}^{\nu,\gamma} \mathsf{X}_\gamma^{-1}$. If S is a subset of E and $\varepsilon > 0$, we denote by $S_\varepsilon := \{a \in E : d(a, S) < \varepsilon\}$ the ε-neighborhood of S. The aim of the present section is to establish the following result:

Theorem 5.1. Let Assumptions 3.1–3.3 hold true. Let either Assumption 3.5 or Assumptions 3.6-3.7 hold true. Then, for every $\eta > 0$, for every compact set $\mathcal{K} \subset E$ s.t. $\mathcal{K} \cap \mathcal{D} \neq \emptyset$,

$$\forall M \geq 0, \quad \sup_{a \in \mathcal{K} \cap \mathcal{D}_{\gamma M}} \mathbb{P}^{a,\gamma} \left(\mathsf{d}(\mathsf{X}_\gamma, \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), \cdot)) > \eta \right) \xrightarrow[\gamma \to 0]{} 0. \tag{16}$$

Using the Yosida regularization $A_\gamma(s,x)$ of $A(s,x)$, the iterates (4) can be rewritten as $x_0 = a \in \mathcal{D}_{\gamma M}$ and

$$x_{n+1} = x_n - \gamma B(\xi_{n+1}, x_n) - \gamma A_\gamma(\xi_{n+1}, x_n - \gamma B(\xi_{n+1}, x_n)). \tag{17}$$

Setting $h_\gamma(s,x) := -B(s,x) - A_\gamma(s, x - \gamma B(s,x))$, the iterates (4) can be cast into the same form as the one studied in [13]. The following result, which we state here mainly for the ease of the reading, is a straightforward consequence of [13, Th. 5.1].

Proposition 5.2. Let Assumptions 3.1–3.3 hold true. Assume moreover that for every $s \in \Xi$, $D(s) = E$. Then, Eq. (16) holds true.

Proof. It is sufficient to check that the mapping h_γ satisfies the Assumption (RM) of [13, Th. 5.1]. Assumption i) in [13, As. (RM)] is satisfied by definition of h_γ. As $D(\cdot)$ is a constant equal to E, the operator $A(s,\cdot)$ is upper semi continuous as a set-valued operator [33]. Thus, $H(s,\cdot) := -A(s,\cdot) - B(s,\cdot)$ is proper, upper semi continuous with closed convex values, and μ-integrable. Hence, the assumptions iii-iv) in [13, As. (RM)] are satisfied. Assumption v) is satisfied by the natural properties of the semiflow induced by the maximal monotone map $\mathcal{A} + \mathcal{B}$, whereas Assumption vi) in [13, As. (RM)] directly follows from the present Assumptions 3.1 and 3.3 and the definition of h_γ. One should finally verify Assumption ii) in [13, As. (RM)], which states that for every converging sequence $(u_n, \gamma_n) \to (u^\star, 0)$, $h_{\gamma_n}(s, u_n) \to H(s, u^\star)$, for every $s \in \Xi$. To this end, it is sufficient to prove that

$$A_{\gamma_n}(s, u_n - \gamma_n B(s, u_n)) \to A(s, u^\star)\,. \tag{18}$$

Choose $\varepsilon > 0$. As $A(s,\cdot)$ is upper semi continuous, there exists $\eta > 0$ s.t. $\forall u$, $\|u - u^\star\| < \eta$ implies $A(s,u) \subset A(s,u^\star)_\varepsilon$. Let $v_n := J_{\gamma_n}(s, u_n - \gamma B(s,u_n))$. By the triangular inequality and the non-expansiveness of J_{γ_n},

$$\|v_n - u^\star\| \leq \|u_n - u^\star\| + \gamma_n \|B(s,u_n)\| + \|J_{\gamma_n}(u^\star) - u^\star\|\,,$$

where it is clear that each of the three terms in the right hand side tends to zero. Thus, there exists $N \in \mathbb{N}$ s.t. $\forall n \geq N$, $\|v_n - u^\star\| \leq \eta$, which in turn implies $A(s, v_n) \subset A(s, u^\star)_\varepsilon$. As $A_{\gamma_n}(s, u_n - \gamma_n B(s,u_n)) \in A(s, v_n)$, the convergence (18) is established. □

5.2 Proof of Th. 5.1

In the sequel, we prove Theorem 5.1 under the set of Assumptions 3.6-3.7. The proof in the common domain case *i.e.*, when Assumption 3.5 holds, is somewhat easier and follows from the same arguments.

In order to prove Theorem 5.1, we just have to weaken the assumptions of Proposition 5.2: for a given $s \in \Xi$, the domain $D(s)$ is not necessarily equal to E and the monotone operator $A(s, .)$ is not necessarily upper semi continuous. Up to these changes, the proof is similar to the proof of [13, Th. 3.1] and the modifications are in fact confined to specific steps of the proof.

Choose a compact set $\mathcal{K} \subset E$ s.t. $\mathcal{K} \cap \mathrm{cl}(\mathcal{D}) \neq \emptyset$. Choose $R > 0$ s.t. $\mathcal{K}$ is contained in the ball of radius R. For every $x = (x_n : n \in \mathbb{N})$ in $E^{\mathbb{N}}$, define $\tau_R(x) := \inf\{n \in \mathbb{N} : x_n > R\}$ and introduce the measurable mapping $C_R : E^{\mathbb{N}} \to E^{\mathbb{N}}$, given by

$$C_R(x) : n \mapsto x_n \mathbb{1}_{n < \tau_R(x)} + x_{\tau_R(x)} \mathbb{1}_{n \geq \tau_R(x)}\,.$$

Consider the image measure $\bar{\mathbb{P}}^{a,\gamma} := \mathbb{P}^{a,\gamma} B_R^{-1}$, which corresponds to the law of the *truncated* process $B_R(x)$. The crux of the proof consists in showing that for every $\eta > 0$ and every $M > 0$,

$$\sup_{a \in \mathcal{K} \cap \mathcal{D}_{\gamma M}} \bar{\mathbb{P}}^{a,\gamma} \left(\mathsf{d}(\mathsf{X}_\gamma, \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), \cdot)) > \eta \right) \xrightarrow[\gamma \to 0]{} 0. \tag{19}$$

Eq. (19) is the counterpart of [13, Lemma 4.3]. Once it has been proven, the conclusion follows verbatim from [13, Section 4, End of the proof]. Our aim is thus to establish Eq. (19). The proof follows the same steps as the proof of [13, Lemma 4.3] up to some confined changes. Here, the steps of the proof which do not need any modification are recalled rather briefly (we refer the reader to [13] for the details). On the other hand, the parts which require an adaptation are explicitly stated as lemmas, whose detailed proofs are provided in Appendix B.

Define $h_{\gamma,R}(s,a) := h_\gamma(s,a)\mathbb{1}_{\|a\|\le R}$. First, we recall the following decomposition, established in [13]:

$$\mathsf{X}_\gamma = \Pi_0 + \mathsf{G}_{\gamma,R}\circ \mathsf{X}_\gamma + \mathsf{X}_\gamma \circ \Delta_{\gamma,R}\,,$$

$\bar{\mathbb{P}}^{a,\gamma}$ almost surely, where $\Pi_0 : E^{\mathbb{N}} \to C(\mathbb{R}_+,E)$, $\mathsf{G}_{\gamma,R} : C(\mathbb{R}_+,E) \to C(\mathbb{R}_+,E)$ and $\Delta_{\gamma,R} : E^{\mathbb{N}} \to E^{\mathbb{N}}$ are the mappings respectively defined by

$$\begin{aligned}
&\Pi_0(x) : t \mapsto x_0\\
&\Delta_{\gamma,R}(x) : n \mapsto (x_n - x_0) - \gamma \sum_{k=0}^{n-1} \int h_{\gamma,R}(s,x_k)\mu(ds)\\
&\mathsf{G}_{\gamma,R}(\mathsf{x}) : t \mapsto \int_0^t \int h_{\gamma,R}(s,\mathsf{x}(\gamma\lfloor u/\gamma\rfloor))\,\mu(ds)du\ ,
\end{aligned}$$

for every $x = (x_n : n\in\mathbb{N})$ and every $\mathsf{x} \in C(\mathbb{R}_+,E)$.

Lemma 5.3. For all $\gamma \in (0,\gamma_0]$ and all $x \in E^{\mathbb{N}}$, define $Z^\gamma_{n+1}(x) := \gamma^{-1}(x_{n+1}-x_n)$. There exists $\varepsilon > 0$ such that:

$$\sup_{n\in\mathbb{N},a\in\mathcal{K}\cap\mathcal{D}_{\gamma M},\gamma\in(0,\gamma_0]} \bar{\mathbb{E}}^{a,\gamma}\left(\left(\|Z^\gamma_n\| + \frac{\boldsymbol{d}(x_n)}{\gamma}\mathbb{1}_{\|x_n\|\le R}\right)^{1+\varepsilon}\right) < +\infty \tag{20}$$

Using [13, Lemma 4.2], the uniform integrability condition (20) implies[1] that $\{\bar{\mathbb{P}}^{a,\gamma}\mathsf{X}_\gamma^{-1} : a \in \mathcal{K}\cap\mathcal{D}_{\gamma M}, \gamma\in(0,\gamma_0]\}$ is tight, and for any $T>0$,

$$\sup_{a\in\mathcal{K}\cap\mathcal{D}_{\gamma M}} \bar{\mathbb{P}}^{a,\gamma}(\|\mathsf{X}_\gamma\circ\Delta_{\gamma,R}\|_{\infty,T} > \varepsilon) \xrightarrow{\gamma\to 0} 0\,, \tag{21}$$

where the notation $\|\mathsf{x}\|_{\infty,T}$ stands for the uniform norm of x on $[0,T]$.

Lemma 5.4. For an arbitrary sequence (a_n,γ_n) such that $a_n \in \mathcal{K}\cap\mathcal{D}_{\gamma_n M}$ and $\gamma_n \to 0$, there exists a subsequence (still denoted as (a_n,γ_n)) such that $(a_n,\gamma_n) \to (a^*,0)$ for some $a^* \in \mathcal{K}\cap \mathrm{cl}(\mathcal{D})$, and there exists r.v. z and $(\mathsf{x}_n : n\in\mathbb{N})$ defined on some probability space $(\Omega',\mathscr{F}',\mathbb{P}')$ into $C(\mathbb{R}_+,E)$ s.t. x_n has the distribution $\bar{\mathbb{P}}^{a_n,\gamma_n}\mathsf{X}^{-1}_{\gamma_n}$ and $\mathsf{x}_n(\omega) \to \mathsf{z}(\omega)$ for all $\omega\in\Omega'$. Moreover, defining

$$u_n(t) := \mathsf{x}_n(\gamma_n\lfloor t/\gamma_n\rfloor)\ ,$$

the sequence (a_n,γ_n) and (x_n) can be chosen in such a way that the following holds $\mathbb{P}'$-a.e.

$$\sup_n \int_0^T \left(\frac{\boldsymbol{d}(u_n(t))}{\gamma_n}\mathbb{1}_{\|u_n(t)\|\le R}\right)^{1+\frac{\varepsilon}{2}} dt\ < +\infty \quad (\forall T>0)\,, \tag{22}$$

where $\varepsilon > 0$ is the constant introduced in Lem. 5.3.

The limit z satisfies the following:

Lemma 5.5. Introduce the open ball $B_R := \{u \in E : \|u\| < R\}$. The following holds $\mathbb{P}'$-a.e.:

$$\forall t\ge 0,\ \mathsf{z}(t) \in \mathrm{cl}(\mathcal{D}) \cup B_R^c\,. \tag{23}$$

[1] Lemma 4.2 of [13] was actually shown with condition $[a\in\mathcal{K}]$ instead of $[a\in\mathcal{K}\cap\mathcal{D}_{\gamma M}]$, but the proof can be easily adapted to the latter case.

Define

$$v_n(s,t) := h_{\gamma_n,R}(s, u_n(t))\,.$$

Thanks to the convergence (21), the following holds $\mathbb{P}'$-a.e.:

$$\mathsf{z}(t) = \mathsf{z}(0) + \lim_{n\to\infty} \int_0^t \int_\Xi v_n(s,u)\,\mu(ds)\,du \qquad (\forall t \geq 0)\,. \tag{24}$$

We now select an $\omega \in \Omega'$ s.t. the events (22), (23) and (24) are all realized, and omit the dependence in ω in the sequel. Otherwise stated, u_n and v_n are handled from now on as determinitic functions, and no longer as random variables. The aim of the next lemmas is to analyze the integrand $v_n(s,u)$. Let $H_R(s,a) := -A(s,a) - B(s,a)$ if $\|a\| < R$, and $H_R(s,a) := E$ otherwise. Denote the corresponding selection integral as $\mathsf{H}_R(a) = \int H_R(s,a)\,\mu(ds)$.

Lemma 5.6. For every s μ-a.e., it holds that for every $t \geq 0$, $(u_n(t), v_n(s,t)) \to \mathrm{gr}(H_R(s,\,.\,))$.

Consider some $T > 0$ and let λ_T represent the Lebesgue measure on the interval $[0,T]$. To simplify notations, we set $\mathcal{L}_E^{1+\varepsilon} := \mathcal{L}^{1+\varepsilon}(\Xi \times [0,T], \mathscr{G} \otimes \mathscr{B}([0,T]), \mu \otimes \lambda_T; E)$.

Lemma 5.7. The sequence $(v_n : n \in \mathbb{N})$ forms a bounded subset of $\mathcal{L}_E^{1+\varepsilon/2}$.

The sequence of mappings $((s,t) \mapsto (v_n(s,t), \|v_n(s,t)\|))$ is bounded in $\mathcal{L}_{E\times\mathbb{R}}^{1+\varepsilon/2}$ and therefore admits a weak cluster point in that space. We denote by (v,w) such a cluster point, where $v : \Xi \times [0,T] \to E$ and $w : \Xi \times [0,T] \to \mathbb{R}$. The following lemma is a consequence of Lem. 5.6.

Lemma 5.8. For every (s,t) $\mu \otimes \lambda_T$-a.e., $(\mathsf{z}(t), v(s,t)) \in \mathrm{gr}(H_R(s,\,.\,))$.

By Lem. 5.8 and Fubini's theorem, there is a λ_T-negligible set s.t. for every t outside this set, $v(\,.\,,t)$ is an integrable selection of $H_R(\,.\,,\mathsf{z}(t))$. Moreover, as v is a weak cluster point of v_n in $\mathcal{L}_E^{1+\varepsilon/2}$, it holds that

$$\mathsf{z}(t) = \mathsf{z}(0) + \int_0^t \int_\Xi v(s,u)\,\mu(ds)\,du\,, \qquad (\forall t \in [0,T])\,.$$

Define $\mathsf{H}_R(a) := \int H_R(\,.\,,a)d\mu$. By the above equality, z is a solution to the DI $\dot{\mathsf{x}} \in \mathsf{H}_R(\mathsf{x})$ with initial condition $\mathsf{z}(0) = a^*$. Denoting by $\Phi_R(a^*)$ the set of such solutions, this reads $\mathsf{z} \in \Phi_R(a^*)$. As $a^* \in \mathcal{K}\cap\mathrm{cl}(\mathcal{D})$, one has $\mathsf{z} \in \Phi_R(\mathcal{K}\cap\mathrm{cl}(\mathcal{D}))$ where we use the notation $\Phi_R(S) := \cup_{a\in S}\Phi_R(a)$ for every set $S \subset E$. Extending the notation $\mathsf{d}(x,\mathsf{S}) := \inf_{\mathsf{y}\in\mathsf{S}} \mathsf{d}(\mathsf{x},\mathsf{y})$, we obtain that $\mathsf{d}(\mathsf{x}_n, \Phi_R(\mathcal{K}\cap\mathrm{cl}(\mathcal{D}))) \to 0$. Thus, for every $\eta > 0$, we have shown that $\bar{\mathbb{P}}^{a_n,\gamma_n}(\mathsf{d}(X_{\gamma_n}, \Phi_R(\mathcal{K}\cap\mathrm{cl}(\mathcal{D}))) > \eta) \to 0$ as $n \to \infty$. We have thus proven the following result:

$$\forall \eta > 0,\ \lim_{\gamma\to 0} \sup_{a\in\mathcal{K}\cap\mathcal{D}_{\gamma M}} \bar{\mathbb{P}}^{a,\gamma}(\mathsf{d}(X_\gamma, \Phi_R(\mathcal{K}\cap\mathrm{cl}(\mathcal{D}))) > \eta) = 0\,.$$

Letting $T > 0$ and choosing $R > \sup\{\|\Phi(a,t)\| : t \in [0,T], a \in \mathcal{K}\cap\mathrm{cl}(\mathcal{D})\}$ (the latter quantity being finite, see *e.g.* [15]), it is easy to show that any solution to the DI $\dot{\mathsf{x}} \in \mathsf{H}_R(\mathsf{x})$ with initial condition $a \in \mathcal{K}\cap\mathrm{cl}(\mathcal{D})$ coincides with $\Phi(a,\,.\,)$ on $[0,T]$. By the same arguments as in [13, Section 4 - End of the proof], Theorem 5.1 follows.

6 Cluster points of the P_γ invariant measures. End of the proof of Th. 3.2

Lemma 6.1. Assume that there exists $x_\star \in Z(\mathcal{A}+\mathcal{B})$ that admits a $\mathcal{L}^2$ representation. Then,

$$P_\gamma(x, \|\cdot - x_\star\|^2) \leq \|x - x_\star\|^2 - 0.5\gamma\psi_\gamma(x) + \gamma^2 C,$$

where ψ_γ is the function defined in (6).

Proof. By assumption, there exists a $\mathcal{L}^2$ representation (φ, B) of $x_\star$. By expanding

$$\|x_{n+1} - x_\star\|^2 = \|x_n - x_\star\|^2 + 2\langle x_{n+1} - x_n, x_n - x_\star\rangle + \|x_{n+1} - x_n\|^2 ,$$

and by using (17), we obtain

$$\|x_{n+1} - x_\star\|^2 = \|x_n - x_\star\|^2 - 2\gamma\langle A_\gamma(\xi_{n+1}, x_n - \gamma B(\xi_{n+1}, x_n)) + B(\xi_{n+1}, x_n), x_n - x_\star\rangle \\ + \gamma^2\|A_\gamma(\xi_{n+1}, x_n - \gamma B(\xi_{n+1}, x_n)) + B(\xi_{n+1}, x_n)\|^2. \quad (25)$$

Write $x = x_n$, $A_\gamma = A_\gamma(\xi_{n+1}, x_n - \gamma B(\xi_{n+1}, x_n))$, $J_\gamma = J_\gamma(\xi_{n+1}, x_n - \gamma B(\xi_{n+1}, x_n))$, $B = B(\xi_{n+1}, x_n)$, $B_\star = (\xi_{n+1}, x_\star)$, and $\varphi = \varphi(\xi_{n+1})$ for conciseness. We write

$$\begin{aligned}\langle A_\gamma, x - x_\star\rangle &= \langle A_\gamma - \varphi, J_\gamma - x_\star\rangle + \langle A_\gamma - \varphi, x - \gamma B - J_\gamma\rangle + \gamma\langle A_\gamma - \varphi, B\rangle \\ &\quad + \langle\varphi, x - x_\star\rangle \\ &= \langle A_\gamma - \varphi, J_\gamma - x_\star\rangle + \gamma\|A_\gamma\|^2 - \gamma\langle A_\gamma, \varphi\rangle + \gamma\langle A_\gamma - \varphi, B\rangle + \langle\varphi, x - x_\star\rangle.\end{aligned}$$

We also write $\langle B, x - x_\star\rangle = \langle B - B_\star, x - x_\star\rangle + \langle B_\star, x - x_\star\rangle$ and $\gamma^2\|A_\gamma + B\|^2 = \gamma^2(\|A_\gamma\|^2 + \|B\|^2 + 2\langle A_\gamma, B\rangle)$. Plugging these identities at the right hand side of (25), we obtain

$$\begin{aligned}\|x_{n+1} - x_\star\|^2 &= \|x - x_\star\|^2 - 2\gamma\left\{\langle A_\gamma - \varphi, J_\gamma - x_\star\rangle + \langle B - B_\star, x - x_\star\rangle\right\} - \gamma^2\|A_\gamma\|^2 \\ &\quad + 2\gamma^2\langle A_\gamma, \varphi\rangle + 2\gamma^2\langle\varphi, B\rangle + \gamma^2\|B\|^2 - 2\gamma\langle\varphi + B_\star, x - x_\star\rangle \\ &\leq \|x - x_\star\|^2 - 2\gamma\left\{\langle A_\gamma - \varphi, J_\gamma - x_\star\rangle + \langle B - B_\star, x - x_\star\rangle\right\} - (\gamma^2/2)\|A_\gamma\|^2 \\ &\quad + (3\gamma^2/2)\|B\|^2 + 4\gamma^2\|\varphi\|^2 - 2\gamma\langle\varphi + B_\star, x - x_\star\rangle \\ &\leq \|x - x_\star\|^2 - 2\gamma\left\{\langle A_\gamma - \varphi, J_\gamma - x_\star\rangle + \langle B - B_\star, x - x_\star\rangle\right\} - (\gamma^2/2)\|A_\gamma\|^2 \\ &\quad + 3\gamma^2\|B - B_\star\|^2 + 3\gamma^2\|B_\star\|^2 + 4\gamma^2\|\varphi\|^2 - 2\gamma\langle\varphi + B_\star, x - x_\star\rangle\end{aligned}$$

where the first inequality is due to the fact that $2\langle a, b\rangle \leq \|a\|^2/2 + 2\|b\|^2$ and the second to the triangle inequality. Observe that the term between the braces at the right hand side of the last inequality is nonnegative thanks to the monotonicity of $A(s, \cdot)$ and $B(s, \cdot)$. Taking the conditional expectation $\mathbb{E}_n$ at each side, the contribution of the last inner product at the right hand side disappears, and we obtain

$$P_\gamma(x, \|\cdot - x_\star\|^2) \leq \|x - x_\star\|^2 - 0.5\gamma\psi_\gamma(x) + 4\gamma^2\int\|\varphi(s)\|^2\mu(ds) + 3\gamma^2\int\|B(s, x_\star)\|^2\mu(ds)$$

where ψ_γ is the function defined in (6). □

Given $k \in \mathbb{N}$, we denote by P_γ^k the kernel P_γ iterated k times. The iterated kernel is defined recursively as $P_\gamma^0(x, dy) = \delta_x(dy)$, and

$$P_\gamma^k(x, S) = \int P_\gamma^{k-1}(y, S)\, P_\gamma(x, dy)$$

for each $S \in \mathscr{B}(E)$.

Lemma 6.2. Let the assumptions of the statement of Th. 5.1 hold true. Assume that for all $\varepsilon > 0$, there exists $M > 0$ such that

$$\sup_{\gamma\in(0,\gamma_0]}\ \sup_{\pi\in\mathcal{I}(P_\gamma)} \pi((\mathcal{D}_{M\gamma})^c) \leq \varepsilon. \quad (26)$$

Then, as $\gamma \to 0$, any cluster point of $\mathcal{I}(\mathcal{P})$ is an element of $\mathcal{I}(\Phi)$.

Note that in the common domain case, (26) is trivially satisfied, since the supports of all the invariant measures are included in $\mathrm{cl}(\mathcal{D})$.

Proof. Choose two sequences (γ_i) and (π_i) such that $\gamma_i \to 0$, $\pi_i \in \mathcal{I}(P_{\gamma_i})$ for all $i \in \mathbb{N}$, and π_i converges narrowly to some $\pi \in \mathcal{M}(E)$ as $i \to \infty$.

Let f be a real, bounded, and Lipschitz function on E with Lipschitz coefficient L. By definition, $\pi_i(f) = \pi_i(P_{\gamma_i}^k f)$ for all $k \in \mathbb{N}$. Set $t > 0$, and let $k_i = \lfloor t/\gamma_i \rfloor$. We have

$$\begin{aligned}
|\pi_i f - \pi_i(f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), t))| &= \left| \int (P_{\gamma_i}^{k_i}(a, f) - f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), t)))\pi_i(da) \right| \\
&\leq \int \left| P_{\gamma_i}^{k_i}(a, f) - f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)) \right| \pi_i(da) \\
&\quad + \int \left| f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)) - f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), t)) \right| \pi_i(da) \\
&\leq \int \mathbb{E}^{a,\gamma_i} \left| f(x_{k_i}) - f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)) \right| \pi_i(da) \\
&\quad + \int \left| f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)) - f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), t)) \right| \pi_i(da) \\
&:= U_i + V_i \,.
\end{aligned}$$

By the boundedness and the Lispchitz-continuity of f,

$$U_i \leq \int \mathbb{E}^{a,\gamma_i} \left[2\|f\|_\infty \wedge L\|x_{k_i} - \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)\| \right] \pi_i(da) \,.$$

Fixing an arbitrarily small $\varepsilon > 0$, it holds by (26) that $\pi_i((\mathcal{D}_{M\gamma_i})^c) \leq \varepsilon/2$ for a large enough M. By the tightness of (π_i), we can choose a compact $\mathcal{K} \subset E$ s.t. for all i, $\pi_i(\mathcal{K}^c) \leq \varepsilon/2$. With these choices, we obtain

$$U_i \leq \sup_{a \in \mathcal{K} \cap \mathcal{D}_{M\gamma_i}} \mathbb{E}^{a,\gamma_i} \left[2\|f\|_\infty \wedge L\|x_{k_i} - \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)\| \right] + 2\|f\|_\infty \, \varepsilon \,.$$

Denoting as $(\cdot)_{[0,t]}$ the restriction of a function to the interval $[0,t]$, and observing that $\|x_{k_i} - \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)\| \leq \|(\mathsf{X}_\gamma(x) - \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), \cdot))_{[0,t]}\|_\infty$, we can now apply Th. 5.1 to obtain

$$\sup_{a \in \mathcal{K} \cap \mathcal{D}_{M\gamma_i}} \mathbb{E}^{a,\gamma_i} \left[2\|f\|_\infty \wedge L\|x_{k_i} - \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)\| \right] \xrightarrow[i\to\infty]{} 0 \,.$$

As ε is arbitrary, we obtain that $U_i \to_i 0$. Turning to V_i, fix an arbitrary $\varepsilon > 0$, and choose a compact $\mathcal{K} \subset E$ such that $\pi_i(\mathcal{K}^c) \leq \varepsilon$ for all i. We have

$$V_i \leq \sup_{a \in \mathcal{K}} \left| f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), k_i\gamma_i)) - f(\Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), t)) \right| + 2\|f\|_\infty \varepsilon \,.$$

By the uniform continuity of the function $f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), \cdot)$ on the compact $\mathcal{K} \times [0,t]$, and by the convergence $k_i\gamma_i \uparrow t$, we obtain that $\limsup_i V_i \leq 2\|f\|_\infty \varepsilon$. As ε is arbitrary, $V_i \to_i 0$. In conclusion, $\pi_i f - \pi_i(f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), t)) \to_i 0$. Moreover, $\pi_i f - \pi_i(f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), t)) \to_i \pi f - \pi(f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), t))$ since $f(\cdot) - f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), t))$ is bounded continuous. Thus, $\pi f = \pi(f \circ \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(\cdot), t))$. Since π_i converges narrowly to π, we obtain that for all $\eta > 0$, $\pi(\mathrm{cl}(\mathcal{D}_\eta)^c) \leq \liminf_i \pi_i(\mathrm{cl}(\mathcal{D}_\eta)^c) = 0$ by choosing ε arbitrarily small in (26) and making $\gamma_i \to 0$. Thus, $\mathrm{supp}(\pi) \subset \mathrm{cl}(\mathcal{D})$, and we obtain in conclusion that $\pi f = \pi(f \circ \Phi(\cdot, t))$ for an arbitrary real, bounded, and Lipschitz continuous function f. Thus, $\pi \in \mathcal{I}(\Phi)$. □

To establish (26) in the different domains case, we need the following lemma.

Lemma 6.3. Let Assumptions 3.6, 3.8, and 3.4–(a) hold true. Then, for all $\varepsilon > 0$, there exists $M > 0$ such that

$$\sup_{\gamma \in (0,\gamma_0]} \sup_{\pi \in \mathcal{I}(P_\gamma)} \pi((\mathcal{D}_{M\gamma})^c) \leq \varepsilon.$$

Proof. We start by writing

$$\boldsymbol{d}(x_{n+1}) \le \|x_{n+1} - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\| \le \|x_{n+1} - \Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n)\| + \|\Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|.$$

On the one hand, we have by Assumption 3.8 and the nonexpansiveness of the resolvent that

$$\begin{aligned}\bar{\mathbb{E}}_n^{a,\gamma}\|x_{n+1} - \Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n)\| &\le \bar{\mathbb{E}}_n^{a,\gamma}\|J_\gamma(\xi_{n+1}, x_n) - \Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n)\| + \gamma\bar{\mathbb{E}}_n^{a,\gamma}\|B(\xi_{n+1}, x_n)\| \\ &\le C\gamma(1 + \Psi(x_n)),\end{aligned}$$

on the other hand, since

$$\|\Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^2 \le \boldsymbol{d}(x_n)^2 - d(x_n, D(\xi_{n+1}))^2 \quad \text{(see (28))},$$

we can make use of Assumption 3.6 to obtain

$$\bar{\mathbb{E}}_n^{a,\gamma}\|\Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\| \le (\bar{\mathbb{E}}_n^{a,\gamma}\|\Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^2)^{1/2} \le \rho\boldsymbol{d}(x_n),$$

where $\rho \in [0, 1)$. We therefore obtain that $\bar{\mathbb{E}}_n^{a,\gamma}\boldsymbol{d}(x_{n+1}) \le \rho\boldsymbol{d}(x_n) + C\gamma(1 + \Psi(x_n))$. By iterating, we end up with the inequality

$$P_\gamma^{n+1}(a, \boldsymbol{d}) \le \rho^{n+1}\boldsymbol{d}(a) + C\gamma\sum_{k=0}^{n}\rho^{n-k}(1 + P_\gamma^k(a, \Psi)). \tag{27}$$

By Lem. 6.1, $\psi_\gamma(x) \le 2\gamma^{-1}\|x - x_\star\|^2 + \gamma C$, thus $P_\gamma(a, \psi_\gamma) \le 2\gamma^{-1}P_\gamma(a, \|\cdot - x_\star\|^2) + \gamma C \le 2\gamma^{-1}\|a - x_\star\|^2 + C < \infty$. We obtain similarly that $P_\gamma^k(a, \psi_\gamma) < \infty$, thus $P_\gamma^k(a, \Psi) < \infty$ for all $k \in \mathbb{N}$.

Since $(\mathcal{D}_{M\gamma})^c = \{x : \boldsymbol{d}(x) \ge M\gamma\}$, it holds by Markov's inequality that

$$P_\gamma^k(a, (\mathcal{D}_{M\gamma})^c) \le \frac{P_\gamma^k(a, \boldsymbol{d})}{M\gamma}$$

for all $k \in \mathbb{N}$. Let π_γ be a P_γ-invariant probability measure. From Assumption 3.4–(a) and Lem. 6.1, the inequality (13) in the statement of Prop. 3.3 is satisfied with $V(x) = \|x - x_\star\|^2$, $Q(x) = \Psi(x)$, $\alpha(\gamma) = \gamma/2$, and $\beta(\gamma) = C\gamma^2$. By the first part of this proposition, $\sup_\gamma \pi_\gamma\Psi < \infty$. In particular, noting that $\boldsymbol{d}(x) \le \|x\| + \|\Pi_{\mathrm{cl}(\mathcal{D})}(0)\|$, we obtain that $\sup_\gamma \pi_\gamma\boldsymbol{d} < \infty$. Getting back to (27), we have for all $n \in \mathbb{N}$,

$$\begin{aligned}\pi_\gamma((\mathcal{D}_{M\gamma})^c) &= P_\gamma^{n+1}(\pi_\gamma, (\mathcal{D}_{M\gamma})^c) \\ &\le \frac{P_\gamma^{n+1}(\pi_\gamma, \boldsymbol{d})}{M\gamma} \\ &\le \rho^{n+1}\frac{\pi_\gamma\boldsymbol{d}}{M\gamma} + \frac{C}{M}\sum_{k=0}^{n}\rho^{n-k}(1 + P_\gamma^k(\pi_\gamma, \Psi)) \\ &= \rho^{n+1}\frac{\pi_\gamma\boldsymbol{d}}{M\gamma} + \frac{C}{M}\sum_{k=0}^{n}\rho^{n-k}(1 + \pi_\gamma\Psi) \\ &\le \rho^{n+1}\frac{C}{M\gamma} + \frac{C}{M}.\end{aligned}$$

By making $n \to \infty$, we obtain that $\pi_\gamma((\mathcal{D}_{M\gamma})^c) \le C/M$, and the proof is concluded by taking M as large as required. □

Th. 3.2: proofs of the convergences (7), (9), and (10)

We need to check that the assumptions of Prop. 3.3 are satisfied. Lem. 6.1 shows that the inequality (13) is satisfied with $V(x) = \|x - x_\star\|^2$, $Q(x) = \Psi(x)$, $\alpha(\gamma) = \gamma/2$, and $\beta(\gamma) = C\gamma^2$, and Assumption 3.4–(a) ensures that $\Psi(x) \xrightarrow[\|x\|\to\infty]{} \infty$ as required.

When the assumptions of Th. 5.1, are satisfied, Lem. 6.2 shows with the help of Lem. 6.3 when needed that any cluster point of $\mathcal{I}(\mathcal{P})$ belongs to $\mathcal{I}(\Phi)$. The required convergences follow at once from Prop. 3.3. Theorem 3.2 is proven.

A Proofs relative to Section 4

A.1 Proof of Prop. 4.1

It is well known that the coercivity or the supercoercivity of a function $q \in \Gamma_0$ can be characterized through the study of the recession function q^∞ of q, which is the function in Γ_0 whose epigraph is the recession cone of the epigraph of q [35, §8], [29, § 6.8]. We recall the following fact.

Lemma A.1. The function $q \in \Gamma_0$ is coercive if and only if 0 is the only solution of the inequality $q^\infty(x) \le 0$. It is supercoercive if and only if $q^\infty = \iota_{\{0\}}$.

Proof. By [29, Prop. 6.8.4], $\operatorname{lev}_{\le 0} q^\infty$ is the recession cone of any level set $\operatorname{lev}_{\le a} q$ which is not empty [35, Th. 8.6]. Thus, q is coercive if and only if $\operatorname{lev}_{\le 0} q^\infty$ is the recession cone of a nonempty compact set, hence equal to $\{0\}$. The second point follows from [4, Prop. 2.16]. □

Lemma A.2. For each $\gamma > 0$, $q^\infty = (q_\gamma)^\infty$.

Proof. By [29, Th. 6.8.5], the Legendre-Fenchel transform $(q^\infty)^*$ of q^∞ satisfies $(q^\infty)^* = \iota_{\operatorname{cl}\operatorname{dom} q^*}$. Since $q_\gamma = q \,\square\, ((2\gamma)^{-1}\|\cdot\|^2)$ where $\square$ is the infimal convolution operator, $(q_\gamma)^* = q^* + (\gamma/2)\|\cdot\|^2$. Therefore, $\operatorname{dom} q^* = \operatorname{dom}(q_\gamma)^*$, which implies that $(q^\infty)^* = ((q_\gamma)^\infty)^*$, and the result follows. □

Lemma A.3 ([26, Th. II.2.1])**.** Assume that $q : \Xi \times E \to (-\infty, \infty]$ is a normal integrand such that $q(s,\cdot) \in \Gamma_0$ for almost every s. Assume that $Q(x) := \int q(s,x)\,\mu(ds)$ belongs to Γ_0. Then, $Q^\infty(x) = \int q^\infty(s,x)\,\mu(ds)$, where $q^\infty(s,\cdot)$ is the recession function of $q(s,\cdot)$.

We now enter the proof of Prop. 4.1. Denote by $g_\gamma(s,\cdot)$ the Moreau envelope of the mapping $g(s,\cdot)$ defined above.

Lemma A.4. Let Hypothesis H1 hold true. Then, for all $\gamma > 0$, the mapping

$$G^\gamma : x \mapsto \int g_\gamma(s,x)\,\mu(ds)\,,$$

is well defined on $E \to \mathbb{R}$, and is convex (hence continuous) on E. Moreover, $G^\gamma \uparrow G$ as $\gamma \downarrow 0$.

Proof. Since $x_\star \in \operatorname{dom} G$ from Hypothesis H1, it holds from the definition of the function g that $\int |g(s,x_\star)|\,\mu(ds) < \infty$. Moreover, noting that $\varphi(s) \in \partial g(s,x_\star)$, the inequality $g(s,x) \ge \langle \varphi(s), x - x_\star\rangle + g(s,x_\star)$ holds. Thus,

$$\begin{aligned} g_\gamma(s,x) &= \inf_w \Big(g(s,w) + \frac{1}{2\gamma}\,\|w-x\|^2 \Big) \ge \inf_w \Big(\langle \varphi(s), w - x_\star\rangle + g(s,x_\star) + \frac{1}{2\gamma}\,\|w-x\|^2 \Big) \\ &= \langle \varphi(s), x - x_\star\rangle \; + g(s,x_\star) - \frac{\gamma}{2}\,\|\varphi(s)\|^2. \end{aligned}$$

Writing $x = x^+ - x^-$ where $x^+ = x \vee 0$, this inequality shows that $g_\gamma(\cdot,x_\star)^-$ is integrable. Moreover, since the Moreau envelope satisfies $g_\gamma(s,x) \le g(s,x)$, we obtain that $g_\gamma(\cdot,x_\star)^+ \le g(\cdot,x_\star)^+ \le |g(\cdot,x_\star)|$ who is also integrable. Therefore, $|g_\gamma(\cdot,x_\star)|$ is integrable. For other values of x, we have

$$g_\gamma(s,x) = g_\gamma(s,x_\star) + \int_0^1 \langle x - x_\star, \nabla g_\gamma(s, x_\star + t(x-x_\star)) - \nabla g_\gamma(s,x_\star)\rangle\,dt + \langle x - x_\star, \nabla g_\gamma(s,x_\star)\rangle,$$

where $\nabla g_\gamma(s,x)$ is the gradient of $g_\gamma(s,x)$ w.r.t. x. Using the well know properties of the Yosida regularization (see Sec. 2.1), we obtain

$$|g_\gamma(s,x)| \le |g_\gamma(s,x_\star)| + \frac{\|x-x_\star\|^2}{2\gamma} + \|x-x_\star\|\,\|\varphi(s)\|^2.$$

Consequently, $g_\gamma(\cdot,x)$ is integrable, thus, $G^\gamma(x)$ is defined for all $x \in E$. The convexity and hence the continuity of G^γ follow trivially from the convexity of $g_\gamma(s,\cdot)$.

Since the integrand $g_\gamma(s,x)$ increases as γ decreases, so is the case of $G^\gamma(x)$. If $x \in \operatorname{dom}(G)$, it holds that $|g(\cdot,x)|$ is integrable. On the one hand, $g_\gamma(s,x)^+ \le |g(s,x)|$, and on the other hand, $g_\gamma(s,x)^- \le \|\varphi(s)\|\|x - x_\star\| + |g(s,x_\star)| + \|\varphi(s)\|^2$ for $\gamma \le 2$. By the dominated convergence, $G^\gamma(x) \to G(x)$ as $\gamma \to 0$. If $x \notin \operatorname{dom} G$, then $\int g_\gamma(s,x)^+ \mu(ds) \to \infty$ as $\gamma \to 0$ by monotone convergence, and $\int g_\gamma(s,x)^- \mu(ds)$ remains bound. Thus, $G^\gamma(x) \to \infty$. □

Lemma A.5. Let Hypotheses H1 and H2 hold true. Then, for all γ small enough,

$$G^\gamma(x) + F(x) - G^\gamma(x_\star) - F(x_\star) \le 2\psi_\gamma(x) + \gamma C,$$

where ψ_γ is given by (6).

Proof. By the convexity of $g_\gamma(s,\cdot)$ and $f(s,\cdot)$, we have

$$\begin{aligned} g_\gamma(s, x - \gamma\nabla f(s,x)) - g_\gamma(s,x_\star) &\le \langle \nabla g_\gamma(s, x-\gamma\nabla f(s,x)), x - \gamma\nabla f(s,x) - x_\star\rangle, \text{ and} \\ f(s,x) - f(s,x_\star) - \langle \nabla f(s,x_\star), x - x_\star\rangle &\le \langle \nabla f(s,x) - \nabla f(s,x_\star), x - x_\star\rangle. \end{aligned}$$

Write $g_\gamma = g_\gamma(s, x - \gamma\nabla f(s,x))$, $\nabla f = \nabla f(s,x)$, $\operatorname{prox}_\gamma = \operatorname{prox}_{\gamma g(s,\cdot)}(x - \gamma\nabla f(s,x))$, $\varphi = \varphi(s)$, and $\nabla f_\star = \nabla f(s,x_\star)$. From these two inequalities, we obtain

$$\begin{aligned} & g_\gamma(s, x-\gamma\nabla f(s,x)) - g_\gamma(s,x_\star) + f(s,x) - f(s,x_\star) - \langle \varphi(s) + \nabla f(s,x_\star), x - x_\star\rangle \\ & \le \langle \nabla g_\gamma, x - \gamma\nabla f - x_\star + \operatorname{prox}_\gamma - \operatorname{prox}_\gamma\rangle + \langle \nabla f - \nabla f_\star, x - x_\star\rangle - \langle \varphi, x - x_\star + \operatorname{prox}_\gamma - \operatorname{prox}_\gamma\rangle \\ & = \langle \nabla g_\gamma - \varphi, \operatorname{prox}_\gamma - x_\star\rangle + \langle \nabla f - \nabla f_\star, x - x_\star\rangle + \gamma\|\nabla g_\gamma\|^2 - \gamma\langle \varphi, \nabla g_\gamma + \nabla f\rangle. \end{aligned}$$

Again, by the convexity of $g_\gamma(s,\cdot)$, we have

$$g_\gamma(s, x - \gamma\nabla f(s,x)) \ge g_\gamma(s,x) - \gamma\langle \nabla g_\gamma(s,x), \nabla f(s,x)\rangle.$$

Thus, we obtain

$$\begin{aligned} & g_\gamma(s,x) - g_\gamma(s,x_\star) + f(s,x) - f(s,x_\star) - \langle \varphi(s) + \nabla f(s,x_\star), x - x_\star\rangle \\ & \le \langle \nabla g_\gamma - \varphi, \operatorname{prox}_\gamma - x_\star\rangle + \langle \nabla f - \nabla f_\star, x - x_\star\rangle + \gamma\|\nabla g_\gamma\|^2 - \gamma\langle\varphi, \nabla g_\gamma + \nabla f\rangle + \gamma\langle \nabla g_\gamma(s,x), \nabla f\rangle. \end{aligned}$$

We now bound the sum of the last two terms at the right hand side. By the γ^{-1}-Lipschitz continuity of the Yosida regularization, $|\langle \nabla g_\gamma(s,x) - \nabla g_\gamma, \nabla f\rangle| \le \|\nabla f\|^2$. Using in addition the inequalities $|\langle a,b\rangle| \le \|a\|^2/2 + \|b\|^2/2$ and $\|\nabla f\|^2 \le 2\|\nabla f_\star\|^2 + 2\|\nabla f - \nabla f_\star\|^2$, we obtain

$$\begin{aligned} \gamma\langle \nabla g_\gamma(s,x), \nabla f\rangle - \gamma\langle \varphi, \nabla g_\gamma + \nabla f\rangle &= \gamma\langle \nabla g_\gamma(s,x) - \nabla g_\gamma, \nabla f\rangle + \gamma\langle \nabla g_\gamma, \nabla f\rangle - \gamma\langle\varphi, \nabla g_\gamma + \nabla f\rangle \\ &\le 2\gamma\|\nabla f\|^2 + \gamma\|\nabla g_\gamma\|^2 + \gamma\|\varphi\|^2 \\ &\le 4\gamma\|\nabla f - \nabla f_\star\|^2 + 4\gamma\|\nabla f_\star\|^2 + \gamma\|\nabla g_\gamma\|^2 + \gamma\|\varphi\|^2. \end{aligned}$$

Thus,

$$\begin{aligned} & g_\gamma(s,x) - g_\gamma(s,x_\star) + f(s,x) - f(s,x_\star) - \langle \varphi(s) + \nabla f(s,x_\star), x - x_\star\rangle \\ & \le 2\left(\langle \nabla g_\gamma - \varphi, \operatorname{prox}_\gamma - x_\star\rangle + \langle \nabla f - \nabla f_\star, x - x_\star\rangle + \gamma\|\nabla g_\gamma\|^2 - 6\gamma\|\nabla f - \nabla f_\star\|^2\right) \\ & \quad + 16\gamma\|\nabla f - \nabla f_\star\|^2 - \langle \nabla f - \nabla f_\star, x - x_\star\rangle + \gamma\|\varphi\|^2 + 4\gamma\|\nabla f_\star\|^2 . \end{aligned}$$

Taking the integral with respect to $\mu(ds)$ at both sides, the contribution of the inner product $\langle \varphi + \nabla f_\star, x - x_\star\rangle$ vanishes. Recalling (6), we obtain

$$\begin{aligned} & G_\gamma(x) + F(x) - G_\gamma(x_\star) - F(x_\star) \\ & \le 2\psi_\gamma(x) - \int (\langle \nabla f - \nabla f_\star, x - x_\star\rangle - 16\gamma\|\nabla f - \nabla f_\star\|^2)\, d\mu + \gamma \int (\|\varphi\|^2 + 4\|\nabla f_\star\|^2)\, d\mu . \end{aligned}$$

Using Hypothesis H2, we obtain the desired result. □

End of the proof of Prop. 4.1. Let $\gamma_0 > 0$ be such that Lem. A.5 holds true for all $\gamma \in (0, \gamma_0]$. Denoting as $q(s,\cdot)^\infty$ the recession function of $q(s,\cdot)$, we have

$$(G^{\gamma_0} + F)^\infty \stackrel{(a)}{=} \int ((g_{\gamma_0}(s,\cdot))^\infty + f(s,\cdot)^\infty)\, \mu(ds) \stackrel{(b)}{=} \int (g(s,\cdot)^\infty + f(s,\cdot)^\infty)\, \mu(ds) \stackrel{(c)}{=} (G+F)^\infty,$$

where the equalities (a) and (c) are due to Lem. A.3, and (b) is due to Lem. A.2. Thus, by Lem. A.1, $F+G$ is coercive (resp. supercoercive) if and only if $F+G^{\gamma_0}$ is coercive (resp. supercoercive). Consequently, since G^γ increases as γ decreases by Lem. A.4, the hypotheses H1, H2, and H3–(a) (resp., H1, H2, H3–(b)) imply Assumption 3.4–(a) (resp. Assumption 3.4–(b)). Prop. 4.1 is proven.

A.2 Proof of Lem. 4.2

We first recall that $\partial G(\cdot) = \int \partial g(s,\cdot)\, \mu(ds)$, where

$$\partial g(s,\cdot) = \begin{cases} \alpha(0)^{-1}\partial h(u,\cdot) & \text{if } i = 0, \\ \partial \iota_{\mathcal{C}_i} & \text{otherwise,} \end{cases}$$

for $s = (u,i) \in \Xi$. Let ψ be an arbitrary measurable $\Sigma \to E$ function such that $\psi(u) \in \partial h(u, x_\star)$ for ζ-almost all $u \in \Sigma$ (such functions are called *measurable selections* of the set-valued function $\partial h(\cdot, x_\star)$). For each $d \in E$, it holds by the convexity of $h(u,\cdot)$ that

$$\begin{aligned} h(u, x_\star + d) &\geq h(u,x_\star) + \langle \psi(u), d\rangle, \text{ and} \\ h(u, x_\star - d) &\geq h(u,x_\star) - \langle \psi(u), d\rangle, \end{aligned}$$

for ζ-almost all $u \in \Sigma$. Equivalently,

$$h(u,x_\star) - h(u, x_\star - d) \leq \langle \psi(u), d\rangle \leq h(u, x_\star + d) - h(u, x_\star).$$

Thus, if $\|d\|$ is small enough but otherwise d is arbitrary, we get from the second assumption of the statement that $\langle \psi(u), d\rangle$ is ζ-square-integrable. Thus, $\int \|\psi(u)\|^2\, \zeta(du) < \infty$ (see [26, Th. II.4.2] for a similar argument). Now, writing $s = (u,i) \in \Xi$, every measurable selection ϕ of $\partial g(\cdot, x_\star)$ is of the form

$$\phi(s) = \begin{cases} \alpha(0)^{-1}\psi(u) & \text{if } i = 0, \\ \theta_i & \text{otherwise,} \end{cases}$$

where ψ is a measurable selection of $\partial h(\cdot, x_\star)$, and θ_i is an element of $\partial \iota_{\mathcal{C}_i}(x_\star)$. By what precedes, it is immediate that $\int \|\phi\|^2 d\mu < \infty$. By assumption, there exists a measurable selection φ of $\partial g(\cdot, x_\star)$ such that $\int (\varphi(s) + \nabla f(u, x_\star))\mu(ds) = 0$. Using the first assumption, we get that the couple $(\varphi(s), \nabla f(u, x_\star))$ is a $\mathcal{L}^2$ representation of $x_\star$.

A.3 Proof of Prop. 4.4

The assertions about $Z(\mathcal{A})$ are straightforward. A small calculation shows that

$$\begin{aligned} J_\gamma(s,x) &= (I + \gamma H(s))^{-1}(x - \gamma d(s)), \quad \text{and} \\ A_\gamma(s,x) = A(s, J_\gamma(s,x)) &= (I + \gamma H(s))^{-1}(H(s)x + d(s)). \end{aligned}$$

Using these expressions, we obtain

$$\begin{aligned} \psi_\gamma(x) &= \int \left\{ \langle A(s, J_\gamma(s,x)) - H(s)x_\star - d(s), J_\gamma(s,x) - x_\star\rangle + \gamma \|A(s, J_\gamma(s,x))\|^2 \right\} \mu(ds) \\ &= \int \left\{ (J_\gamma(s,x) - x_\star)^T \frac{H(s) + H^T(s)}{2} (J_\gamma(s,x) - x_\star) + \gamma \|A(s, J_\gamma(s,x))\|^2 \right\} \mu(ds). \end{aligned}$$

Since $(I + \gamma H(s))^{-1}$ and $H(s)(I + \gamma H(s))^{-1}$ are respectively the resolvent and the Yosida regularization of the linear, monotone and maximal operator $H(s)$, it holds that $\|(I + \gamma H(s))^{-1}\| \leq 1$, and $\|\gamma H(s)(I + \gamma H(s))^{-1}\| \leq 1$.

Denoting as $\|\cdot\|_S$ the semi norm associated with any semidefinite nonnegative matrix S, we write

$$\begin{aligned}\psi_\gamma(x) &\geq \int \|J_\gamma(s,x)-x_\star\|^2_{(H(s)+H^T(s))/2}\,\mu(ds)\\ &= \int \Big\|(I+\gamma H(s))^{-1}\Big((x-x_\star)-\gamma(H(s)x_\star+d(s))\Big)\Big\|^2_{(H(s)+H^T(s))/2}\,\mu(ds).\end{aligned}$$

Using the inequality $\|a-b\|^2 \geq 0.5\|a\|^2 - \|b\|^2$, we obtain that $\psi_\gamma(x) \geq 0.5W_\gamma(x) - U_\gamma$, with

$$\begin{aligned}W_\gamma(x) &= \int \|(I+\gamma H(s))^{-1}(x-x_\star)\|^2_{(H(s)+H^T(s))/2}\,\mu(ds), \quad \text{and}\\ U_\gamma &= \gamma^2 \int \|(I+\gamma H(s))^{-1}(H(s)x_\star+d(s))\|^2_{(H(s)+H^T(s))/2}\,\mu(ds)\\ &= \gamma\int \|H(s)x_\star+d(s)\|^2_{\gamma I_\gamma(s)}\,\mu(ds).\end{aligned}$$

with

$$I_\gamma(s) = (I+\gamma H(s))^{-T}\frac{H(s)+H^T(s)}{2}(I+\gamma H(s))^{-1}.$$

From the inequalities shown above, we have

$$\Big\|\gamma I_\gamma(s)\Big\| \leq 1.$$

Therefore,

$$0 \leq U_\gamma \leq \gamma\int \|H(s)x_\star+d(s)\|^2\,\mu(ds) \leq \gamma C.$$

Turning to $W_\gamma(x)$, it holds that

$$W_\gamma(x) = (x-x_\star)^T\Big(\int I_\gamma(s)\,\mu(ds)\Big)(x-x_\star),$$

Since $\|I_\gamma(s)\| \leq \|\frac{H(s)+H^T(s)}{2}\|$ and $I_\gamma(s) \to_{\gamma\to 0} (H(s)+H^T(s))/2$, it holds by dominated convergence that $\int I_\gamma(s)\,\mu(ds) \to_{\gamma\to 0} \boldsymbol{H}+\boldsymbol{H}^T$. If $\boldsymbol{H}+\boldsymbol{H}^T > 0$, then there exists $\gamma_0 > 0$ such that

$$\inf_{\gamma\in(0,\gamma_0]} \lambda_{\min}\Big(\int I_\gamma(s)\,\mu(ds)\Big) > 0,$$

where $\lambda_{\min}$ is the smallest eigenvalue. Thus, Assumption 3.4–(c) is verified.

A.4 Proof of Prop. 4.5

Since $A_\gamma(s,\cdot)$ is $1/\gamma$-Lipschitz, $\|A_\gamma(s,x-\gamma B(s,x))\| \geq \|A_\gamma(s,x)\| - \|B(s,x)\| \geq \|A_\gamma(s,x)\| - \|B(s,x)-B(s,x_\star)\| - \|B(s,x_\star)\|$. Therefore,

$$\begin{aligned}&\psi_\gamma(x)\\ &\geq \int\Big\{\langle B(s,x)-B(s,x_\star),x-x_\star\rangle - 6\gamma\|B(s,x)-B(s,x_\star)\|^2 + \gamma\|A_\gamma(s,x-\gamma B(s,x))\|^2\Big\}\,\mu(ds)\\ &\geq \int\Big\{\langle B(s,x)-B(s,x_\star),x-x_\star\rangle - 8\gamma\|B(s,x)-B(s,x_\star)\|^2 + (\gamma/2)\|A_\gamma(s,x)\|^2\\ &\qquad\qquad - 2\gamma\|B(s,x_\star)\|^2\Big\}\,\mu(ds)\\ &\geq \frac{\gamma}{2}\int \|A_\gamma(s,x)\|^2\,\mu(ds) - \gamma C\end{aligned}$$

for γ small enough, by Hypothesis H4. We now have

$$\gamma\int \|A_\gamma(s,x)\|^2\,\mu(ds) = \frac{1}{\gamma}\int \|x-J_\gamma(s,x)\|^2\,\mu(ds) \geq \frac{1}{\gamma}\int d(s,x)^2\,\mu(ds) \geq \frac{C}{\gamma}\boldsymbol{d}(x)^2$$

thanks to Hypothesis H2. The result follows from the boundedness of $\mathcal{D}$.

A.5 Proof of Prop. 4.6

To prove this proposition, we start with the following result.

Lemma A.6. Let $\mathsf{A} \in \mathscr{M}$ be such that

$$\exists (x_*, y_*) \in \operatorname{gr}(\mathsf{A}),\ \exists \delta > 0, \quad \boldsymbol{S}(x_*, \delta) \subset \operatorname{int}(\operatorname{dom} \mathsf{A}), \text{ and } \forall x \in \boldsymbol{S}(x_*, \delta), \inf_{y \in \mathsf{A}(x)} \langle y - y_*, x - x_* \rangle > 0.$$

Then, assuming that $\operatorname{dom} \mathsf{A}$ is unbounded,

$$\liminf_{x \in \operatorname{dom} \mathsf{A}, \|x\| \to \infty} \frac{\inf_{y \in \mathsf{A}(x)} \langle y - y_*, x - x_* \rangle}{\|x\|} > 0.$$

Proof. Given a vector $u \in E$, define the function

$$f_u(\lambda) = \inf_{y \in \mathsf{A}(x_* + \lambda u)} \langle y - y_*, u \rangle$$

for all $\lambda \geq 0$ such that $x_* + \lambda u \in \operatorname{dom} \mathsf{A}$. For all $\lambda_1 > \lambda_2$ in $\operatorname{dom} f_u$, and all $y_1 \in A(x_* + \lambda_1 u)$ and $y_2 \in A(x_* + \lambda_2 u)$, we have

$$\langle y_1 - y_*, u \rangle - \langle y_2 - y_*, u \rangle = \langle y_1 - y_2, u \rangle = \frac{1}{\lambda_1 - \lambda_2} \langle y_1 - y_2, x_* + \lambda_1 u - (x_* + \lambda_2 u) \rangle \geq 0.$$

Passing to the infima, we obtain that $f_u(\lambda_1) \geq f_u(\lambda_2)$, in other words, f_u is non decreasing.

For all $x \in \operatorname{dom} \mathsf{A}$ such that $\|x - x_*\| \geq \delta$, we have by setting $u = \delta(x - x_*)/\|x - x_*\|$

$$\inf_{y \in \mathsf{A}(x)} \langle y - y_*, x - x_* \rangle = \frac{\|x - x_*\|}{\delta} f_u(\delta^{-1} \|x - x_*\|) \geq \frac{\|x - x_*\|}{\delta} f_u(1).$$

For any $u \in \boldsymbol{S}(0, \delta)$, it holds by assumption that $f_u(1) = \inf_{y \in \mathsf{A}(x_* + u)} \langle y - y_*, u \rangle$ is positive. We shall show that $f_u(1)$ is lower semicontinuous (lsc) as a function of u on the sphere $\boldsymbol{S}(0, \delta)$. Since this sphere is compact, $f_u(1)$ attains its infimum on $\boldsymbol{S}(0, \delta)$, and the lemma will be proven.

It is well-known that A is locally bounded near any point in the interior if its domain [15, Prop. 2.9] [6, §21.4]. Thus, by the closedeness of $\operatorname{gr}(\mathsf{A})$, the inf in the expression of $f_u(1)$ is attained. Let $u_n \to u$, and write $f_{u_n}(1) = \langle y_n - y_*, u_n \rangle$. By the maximality of A, we obtain that for any accumulation point y of (y_n) (who exists by the local boundedness), it holds that $(u, y) \in \operatorname{gr}(\mathsf{A})$. Consequently, $\liminf_n f_{u_n}(1) \geq f_u(1)$, in other words, $f_u(1)$ is lsc. □

We now prove Prop. 4.6. Let us write

$$f(\gamma, s, x) = \frac{\langle A_\gamma(s, x) - \varphi(s), J_\gamma(s, x) - x_\star \rangle}{\|x\|} + \frac{\|x - J_\gamma(s, x)\|^2}{\gamma \|x\|}, \text{ and}$$
$$g(s, x) = \inf_{\gamma \in (0,1]} f(\gamma, s, x).$$

Note that $\psi_\gamma(x)/\|x\| = \int f(\gamma, s, x)\, \mu(ds)$. We shall show that $\liminf_{\|x\| \to \infty} g(s, x) > 0$ for all $s \in \Sigma$. Assume the contrary, namely, that there exist $s \in \Sigma$ and $\|x_k\| \to \infty$ such that $g(s, x_k) \to 0$. In these conditions, there exists a sequence (γ_k) in $(0, 1]$ such that $f(\gamma_k, s, x_k) \to 0$. By inspecting the second term in the expression of $f(\gamma_k, s, x_k)$, we obtain that $\|J_{\gamma_k}(s, x_k)\|/\|x_k\| \to 1$. Rewriting the first term as

$$\frac{\|J_{\gamma_k}(s, x_k)\|}{\|x_k\|} \frac{\langle A_{\gamma_k}(s, x_k) - \varphi(s), J_{\gamma_k}(s, x_k) - x_\star \rangle}{\|J_{\gamma_k}(s, x_k)\|},$$

and recalling that $A_{\gamma_k}(s, x_k) \in A(s, J_{\gamma_k}(s, x_k))$, Lem. A.6 shows that the lim inf of this term is positive, which raises a contradiction.

Note that

$$\inf_{\gamma \in (0,1]} \frac{\psi_\gamma(x)}{\|x\|} \geq \int g(s, x)\, \mu(ds).$$

Using Fatou's lemma, we obtain Assumption 3.4–(a).

B Proofs relative to Section 5

B.1 Proof of Lem. 5.3

Let ε be the smallest of the three constants (also named ε) in Assumptions 3.1, 3.3 and 3.7 respectively where $\mathcal{K} = B_R$. For every a, γ, the following holds for $\bar{\mathbb{P}}^{a,\gamma}$-almost all $x = (x_n : n \in \mathbb{N})$:

$$\begin{aligned}
\boldsymbol{d}(x_{n+1})\mathbb{1}_{\|x_{n+1}\|\le R} &= \boldsymbol{d}(x_{n+1})\mathbb{1}_{\|x_{n+1}\|\le R}(\mathbb{1}_{\|x_n\|\le R} + \mathbb{1}_{\|x_n\|>R}) = \boldsymbol{d}(x_{n+1})\mathbb{1}_{\|x_{n+1}\|\le R}\mathbb{1}_{\|x_n\|\le R} \\
&\le \boldsymbol{d}(x_{n+1})\mathbb{1}_{\|x_n\|\le R} \\
&= \|x_{n+1} - \Pi_{\mathcal{D}}(x_{n+1})\|\mathbb{1}_{\|x_n\|\le R} \\
&\le \|x_{n+1} - \Pi_{\mathcal{D}}(x_n)\|\mathbb{1}_{\|x_n\|\le R}\,.
\end{aligned}$$

Using the notation $\bar{\mathbb{E}}_n^{a,\gamma} = \bar{\mathbb{E}}^{a,\gamma}(\,.\,|x_0,\dots,x_n)$, we thus obtain:

$$\bar{\mathbb{E}}_n^{a,\gamma}(\boldsymbol{d}(x_{n+1})^{1+\varepsilon}\mathbb{1}_{\|x_{n+1}\|\le R}) \le \int \|J_\gamma(s, x_n - \gamma B(s,x_n)) - \Pi_{\mathcal{D}}(x_n)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\, d\mu(s)\,.$$

By the convexity of $\|\cdot\|^{1+\varepsilon}$, for all $\alpha \in (0,1)$,

$$\|x+y\|^{1+\varepsilon} = \frac{1}{\alpha^{1+\varepsilon}}\left\|\alpha x + (1-\alpha)\frac{\alpha}{1-\alpha}y\right\|^{1+\varepsilon} \le \alpha^{-\varepsilon}\|x\|^{1+\varepsilon} + (1-\alpha)^{-\varepsilon}\|y\|^{1+\varepsilon}\,.$$

Therefore, by setting $\delta_\gamma(s,a) := \|J_\gamma(s, a - \gamma B(s,a)) - \Pi_{D(s)}(a)\|$,

$$\begin{aligned}
\bar{\mathbb{E}}_n^{a,\gamma}(\boldsymbol{d}(x_{n+1})^{1+\varepsilon}\mathbb{1}_{\|x_{n+1}\|\le R}) &\le \alpha^{-\varepsilon}\int \delta_\gamma(s,x_n)^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\, d\mu(s) \\
&\quad + (1-\alpha)^{-\varepsilon}\int \|\Pi_{D(s)}(x_n) - \Pi_{\mathcal{D}}(x_n)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\, d\mu(s)\,.
\end{aligned}$$

Note that for every $s \in \Xi$, $a \in E$,

$$\|\delta_\gamma(s,a)\| \le \|J_\gamma(s,a) - \Pi_{D(s)}(a)\| + \gamma\|B(s,a)\|\,.$$

Hence, by Assumptions 3.7 and 3.3, there exists a deterministic constant $C > 0$ s.t.

$$\sup_n \int \delta_\gamma(s,x_n)^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\, d\mu(s) \le C\gamma^{1+\varepsilon}\,.$$

Moreover, since $\Pi_{\mathrm{cl}(D(s))}$ is a firmly non expansive operator [6, Chap. 4], it holds that for all $u \in \mathrm{cl}(\mathcal{D})$, and for μ-almost all s,

$$\|\Pi_{\mathrm{cl}(D(s))}(x_n) - u\|^2 \le \|x_n - u\|^2 - \|\Pi_{\mathrm{cl}(D(s))}(x_n) - x_n\|^2.$$

Taking $u = \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)$, we obtain that

$$\|\Pi_{\mathrm{cl}(D(s))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^2 \le \boldsymbol{d}(x_n)^2 - d(x_n, D(s))^2. \tag{28}$$

Making use of Assumption 3.6, and assuming without loss of generality that $\varepsilon \le 1$, we obtain

$$\begin{aligned}
\int \|\Pi_{\mathrm{cl}(D(s))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^{1+\varepsilon}\, d\mu(s) &\le \left(\int \|\Pi_{\mathrm{cl}(D(s))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^2\, d\mu(s)\right)^{(1+\varepsilon)/2} \\
&\le \alpha'\boldsymbol{d}(x_n)^{1+\varepsilon}\,,
\end{aligned}$$

for some $\alpha' \in [0,1)$. Choosing α close enough to zero, we obtain that there exists $\rho \in [0,1)$ such that

$$\bar{\mathbb{E}}_n^{a,\gamma}\left(\frac{\boldsymbol{d}(x_{n+1})^{1+\varepsilon}}{\gamma^{1+\varepsilon}}\mathbb{1}_{\|x_{n+1}\|\le R}\right) \le \rho\frac{\boldsymbol{d}(x_n)^{1+\varepsilon}}{\gamma^{1+\varepsilon}}\mathbb{1}_{\|x_n\|\le R} + C.$$

Taking the expectation at both sides, iterating, and using the fact that $\boldsymbol{d}(x_0) = \boldsymbol{d}(a) < M\gamma$, we obtain that

$$\sup_{n\in\mathbb{N},a\in\mathcal{K}\cap\mathcal{D}_{\gamma M},\gamma\in(0,\gamma_0]} \bar{\mathbb{E}}^{a,\gamma}\left(\left(\frac{\boldsymbol{d}(x_n)}{\gamma}\right)^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\right) < +\infty\,. \tag{29}$$

Since $A_\gamma(s,\cdot)$ is γ^{-1}-Lipschitz continuous, $\|A_\gamma(s, x-\gamma B(s,x))\| \le \|A_\gamma(s,x)\| + \|B(s,x)\|$. Moreover, choosing measurably $\tilde{x} \in \mathcal{D}$ in such a way that $\|x - \tilde{x}\| \le 2\boldsymbol{d}(x)$, we obtain $\|A_\gamma(s,x)\| \le \|A_0(s,\tilde{x})\| + 2\frac{\boldsymbol{d}(x)}{\gamma}$. Therefore, there exists R' depending only on R and $\mathcal{D}$ s.t.

$$\|A_\gamma(s,x)\|\mathbb{1}_{\|x\|\le R} \le \|A_0(s,\tilde{x})\|\mathbb{1}_{\|\tilde{x}\|\le R'} + 2\frac{\boldsymbol{d}(x)}{\gamma}\mathbb{1}_{\|x\|\le R}\,.$$

Thus,

$$\begin{aligned}\bar{\mathbb{E}}_n^{a,\gamma}(\|Z_{n+1}^\gamma\|^{1+\varepsilon}) &= \int \|h_{\gamma,R}(s,x_n)\|^{1+\varepsilon}d\mu(s)\\ &= \int \|B(s,x_n) + A_\gamma(s, x_n - \gamma B(s,x_n))\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\,d\mu(s)\\ &\le \int \left(2\|B(s,x_n)\| + \|A_0(s,\tilde{x}_n)\| + 2\frac{\boldsymbol{d}(x_n)}{\gamma}\right)^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R'}\,d\mu(s)\,. \end{aligned}\tag{30}$$

By Assumption 3.3, $\int\|B(s,x_n)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\,d\mu(s) \le C$ where the constant C depends only on ε and R. By Assumption 3.1, we also have $\int\|A_0(s,x_n)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\,d\mu(s) \le C$ for some (other) constant C. The third term is controlled by Eq. (29). Taking expectations, the bound (20) is established.

B.2 Proof of Lem. 5.4

The first point can be obtained by straightforward application of Prokhorov and Skorokhod's theorems. However, to verify the second point, we need to construct the sequences more carefully. Choose $\varepsilon > 0$ as in Lem. 5.3. We define the process $Y^\gamma : E^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ s.t. for every $n \in \mathbb{N}$,

$$Y_n^\gamma(x) := \sum_{k=0}^{n-1}\frac{\boldsymbol{d}(x_k)^{1+\varepsilon/2}}{\gamma^{\varepsilon/2}}\mathbb{1}_{\|x_k\|\le R}\,,$$

and we denote by $(X, Y^\gamma) : E^{\mathbb{N}} \to (E\times\mathbb{R})^{\mathbb{N}}$ the process given by $(X,Y^\gamma)_n(x) := (x_n, Y_n^\gamma(x))$. We define for every n, $\tilde{Z}_{n+1}^\gamma := \gamma^{-1}((X,Y^\gamma)_{n+1} - (X,Y^\gamma)_n)$. By Lem. 5.3, it is easily seen that

$$\sup_{n\in\mathbb{N},a\in\mathcal{K}\cap\mathcal{D}_{\gamma M},\gamma\in(0,\gamma_0]} \bar{\mathbb{E}}^{a,\gamma}\left(\|\tilde{Z}_n^\gamma\|\mathbb{1}_{\|\tilde{Z}_n^\gamma\|>A}\right) \xrightarrow{A\to+\infty} 0\,.$$

We now apply [13, Lemma 4.2], only replacing E by $E\times\mathbb{R}$ and $\bar{\mathbb{P}}^{a,\gamma}$ by $\bar{\mathbb{P}}^{a,\gamma}(X,Y^\gamma)^{-1}$. By this lemma, the family $\{\bar{\mathbb{P}}^{a,\gamma}(X,Y^\gamma)^{-1}\overline{\mathsf{X}}_\gamma^{-1} : a \in \mathcal{K}\cap\mathcal{D}_{\gamma M}, \gamma\in(0,\gamma_0]\}$ is tight, where $\overline{\mathsf{X}}_\gamma^{-1} : (E\times\mathbb{R})^{\mathbb{N}} \to C(\mathbb{R}_+, E\times\mathbb{R})$ is the piecewise linear interpolated process, defined in the same way as X_γ only substituting $E\times\mathbb{R}$ with E in the definition. By Prokhorov's theorem, one can choose the subsequence (a_n,γ_n) s.t. $\bar{\mathbb{P}}^{a_n,\gamma_n}(X,Y^{\gamma_n})^{-1}\overline{\mathsf{X}}_{\gamma_n}^{-1}$ converges narrowly to some probability measure Υ on $E\times\mathbb{R}$. By Skorokhod's theorem, we can define a stochastic process $((\mathsf{x}_n,\mathsf{y}_n) : n\in\mathbb{N})$ on some probability space $(\Omega',\mathscr{F}',\mathbb{P}')$ into $C(\mathbb{R}_+, E\times\mathbb{R})$, whose distribution for a fixed n coincides with $\bar{\mathbb{P}}^{a_n,\gamma_n}(X,Y^{\gamma_n})^{-1}\overline{\mathsf{X}}_{\gamma_n}^{-1}$, and s.t. for every $\omega\in\Omega'$, $(\mathsf{x}_n(\omega),\mathsf{y}_n(\omega)) \to (\mathsf{z}(\omega),\mathsf{w}(\omega))$, where (z,w) is a r.v. defined on the same space. In particular, the first marginal distribution of $\bar{\mathbb{P}}^{a_n,\gamma_n}(X,Y^{\gamma_n})^{-1}\overline{\mathsf{X}}_{\gamma_n}^{-1}$ coincides with $\bar{\mathbb{P}}^{a_n,\gamma_n}\mathsf{X}_{\gamma_n}^{-1}$. Thus, the first point is proven.

For every $\gamma\in(0,\gamma_0]$, introduce the mapping

$$\begin{aligned}\Gamma_\gamma : C(\mathbb{R}_+,E) &\to C(\mathbb{R}_+,\mathbb{R})\\ \mathsf{x} &\mapsto \left(t\mapsto \int_0^t (\gamma^{-1}\boldsymbol{d}(\mathsf{x}(\gamma\lfloor u/\gamma\rfloor)))^{1+\varepsilon/2}\mathbb{1}_{\|\mathsf{x}(\gamma\lfloor u/\gamma\rfloor)\|\le R}du\right)\,.\end{aligned}$$

We denote by $\underline{\mathsf{X}}_\gamma^{-1} : \mathbb{R}^{\mathbb{N}} \to C(\mathbb{R}_+, \mathbb{R})$ the piecewise linear interpolated process, defined in the same way as X_γ only substituting $\mathbb{R}$ with E in the definition. It is straightforward to show that $\underline{\mathsf{X}}_\gamma \circ Y^{\gamma_n} = \Gamma_\gamma \circ \mathsf{X}_\gamma$. For every n, by definition of the couple $(\mathsf{x}_n, \mathsf{y}_n)$, the distribution under $\mathbb{P}'$ of the r.v. $\Gamma_{\gamma_n}(\mathsf{x}_n) - \mathsf{y}_n$ is equal to the distribution of $\Gamma_{\gamma_n} \circ \mathsf{X}_{\gamma_n} - \underline{\mathsf{X}}_{\gamma_n} \circ Y^{\gamma_n}$ under $\bar{\mathbb{P}}^{a_n, \gamma_n}$. Therefore, $\mathbb{P}'$-a.e. and for every n, $\mathsf{y}_n = \Gamma_{\gamma_n}(\mathsf{x}_n)$. This implies that, $\mathbb{P}'$-a.e., $\Gamma_{\gamma_n}(\mathsf{x}_n)$ converges (uniformly on compact set) to w. On that event, this implies that for every $T \geq 0$, $\Gamma_{\gamma_n}(\mathsf{x}_n)(T) \to \mathsf{w}(T)$, which is finite. Hence, $\sup_n \Gamma_{\gamma_n}(\mathsf{x}_n)(T) < \infty$ on that event, which proves the second point.

B.3 Proof of Lem. 5.5

For every $t \geq 0$, notice that $u_n(t) \to \mathsf{z}(t)$ $\mathbb{P}'$-a.e. Thus, $\boldsymbol{d}(\mathsf{z}(t))\mathbb{1}_{\|\mathsf{z}(t)\|<R} \leq \liminf_n \boldsymbol{d}(u_n(t))\mathbb{1}_{\|u_n(t)\|\leq R}$. By Fatou's lemma,

$$\mathbb{E}'(\boldsymbol{d}(\mathsf{z}(t))\mathbb{1}_{\|\mathsf{z}(t)\|<R}) \leq \liminf_n \mathbb{E}'(\boldsymbol{d}(u_n(t))\mathbb{1}_{\|u_n(t)\|\leq R}).$$

Define $k_n = \lfloor \frac{t}{\gamma_n} \rfloor$ and notice that

$$\begin{aligned}\mathbb{E}'(\boldsymbol{d}(u_n(t))\mathbb{1}_{\|u_n(t)\|\leq R}) &= \bar{\mathbb{E}}^{a_n,\gamma_n}(\boldsymbol{d}(x_{k_n})\mathbb{1}_{\|x_{k_n}\|\leq R}) \\ &\leq \sup_{k\in\mathbb{N}} \bar{\mathbb{E}}^{a_n,\gamma_n}(\boldsymbol{d}(x_k)\mathbb{1}_{\|x_k\|\leq R}).\end{aligned}$$

By Lem. 5.3 and since $\gamma_n \to 0$, the supremum in the above inequality converges to zero as $n \to \infty$. As a consequence, $\mathbb{E}'(\boldsymbol{d}(\mathsf{z}(t))\mathbb{1}_{\|\mathsf{z}(t)\|<R}) = 0$. This means that, $\mathbb{P}'$-a.e., $\mathsf{z}(t) \in \mathrm{cl}(\mathcal{D}) \cup B_R^c$. As $\mathrm{cl}(\mathcal{D}) \cup B_R^c$ is closed and z is continuous, the probability-one event on which the above inclusion holds can be made independent from t, and the conclusion follows.

B.4 Proof of Lem. 5.6

Consider any $t \geq 0$ and any s s.t. $\mathcal{D} \subset D(s)$. We prove that $(u_n(t), v_n(s,t)) \to \mathrm{gr}(H_R(s, .))$. It is clear that $u_n(t) \to \mathsf{z}(t)$. If $\|\mathsf{z}(t)\| \geq R$, the result is trivial. We now assume that $\|\mathsf{z}(t)\| < R$. In this case, note that $\mathsf{z}(t) \in \mathrm{cl}(\mathcal{D})$ by Lem. 5.5. This also implies that $\mathsf{z}(t) \in \mathrm{cl}(D(s))$.

To simplify notations, we now omit the dependence in (s,t) and write $u_n := u_n(t)$, $v_n := v_n(s,t)$, $A := A(s, .)$, $B := B(s, .)$, $\gamma := \gamma_n$, $J_\gamma := J_\gamma(s, .)$, $A_\gamma := A_\gamma(s, .)$, $D := D(s)$, $H_R = H_R(s, .)$, $\mathsf{z} := \mathsf{z}(t)$. We also define $\tilde{u}_n := J_\gamma(u_n - \gamma B(u_n))$.

As $\|\mathsf{z}\| < R$, it holds that $\|u_n\| < R$ for every n large enough. Thus, $-v_n = B(u_n) + A_\gamma(u_n - \gamma B(u_n))$. We decompose:

$$(u_n, -v_n) = (\tilde{u}_n, B(\tilde{u}_n) + A_\gamma(u_n - \gamma B(u_n))) + (u_n - \tilde{u}_n, B(u_n) - B(\tilde{u}_n)).$$

As $A_\gamma(u_n - \gamma B(u_n)) \in A(\tilde{u}_n)$, the first term in the right hand side belongs to $\mathrm{gr}(A + B)$. It remains to show that the second term converges to zero, and we deduce that $(u_n, v_n) \to \mathrm{gr}(H_R)$, as obviously $\mathrm{gr}(-A - B) \subset \mathrm{gr}(H_R)$. One has

$$\begin{aligned}\|\tilde{u}_n - u_n\| &\leq \|J_\gamma(u_n - \gamma B(u_n)) - (u_n - \gamma B(u_n))\| + \gamma\|B(u_n)\| \\ &= \gamma\|A_\gamma(u_n - \gamma B(u_n))\| + \gamma\|B(u_n)\| \\ &\leq \gamma\|A_\gamma(\mathsf{z})\| + \gamma\|A_\gamma(u_n - \gamma B(u_n)) - A_\gamma(\mathsf{z})\| + \gamma\|B(u_n)\| \\ &\leq \|J_\gamma(\mathsf{z}) - \mathsf{z}\| + \|u_n - \gamma B(u_n) - \mathsf{z}\| + \gamma\|B(u_n)\|,\end{aligned}$$

where, for the last inequality, we used the γ^{-1}-Lipschitz continuity of A_γ. As $\mathsf{z} \in \mathrm{cl}(D(s))$, it holds that $\|J_\gamma(\mathsf{z}) - \mathsf{z}\| \to 0$. Using the continuity of B and the convergence $u_n \to \mathsf{z}$, we conclude that $\|\tilde{u}_n - u_n\| \to 0$ as $n \to \infty$. Thus, $(u_n - \tilde{u}_n, B(u_n) - B(\tilde{u}_n)) \to 0$ and the lemma is shown.

B.5 Proof of Lem. 5.7

Define $c_a := \sup_{a\in B_R\cap\mathcal{D}} \int \|A_0(s,a)\|^{1+\varepsilon/2} d\mu(s)$ and $c_b := \sup_{a:\|a\|\le R} \int \|B(s,a)\|^{1+\varepsilon/2} d\mu(s)$ (these constants being finite by Assumptions 3.1 and 3.3). By the same derivations as those leading to Eq. (30), we obtain

$$\int \|v_n(s,t)\|^{1+\varepsilon/2} d\mu(s) \le C\left(\frac{\boldsymbol{d}(u_n(t))^{1+\varepsilon/2}}{\gamma^{1+\varepsilon/2}}\mathbb{1}_{\|u_n(t)\|\le R} + c_a + c_b\right).$$

The proof is concluded by applying Lem. 5.4.

B.6 Proof of Lemma 5.8

The sequence $((v_n, \|v_n(\,.\,,\,.\,)\|))$ converges weakly to (v,w) in $\mathcal{L}^{1+\varepsilon/2}_{E\times\mathbb{R}}$ along some subsequence (*n.b.*: compactness and sequential compactness are the same notions in the weak topology of $\mathcal{L}^{1+\varepsilon/2}_{E\times\mathbb{R}}$). We still denote by $((v_n, \|v_n(\,.\,,\,.\,)\|))$ this subsequence. By Mazur's theorem, there exists a function $J : \mathbb{N}\to\mathbb{N}$ and a sequence of sets of weights $\{\alpha_{k,n} : n\in\mathbb{N}, k = n\ldots, J(n) \,:\, \alpha_{k,n}\ge 0, \sum_{k=n}^{J(n)}\alpha_{k,n} = 1\}$ such that the sequence of functions

$$(\bar v_n, \bar w_n) \,:\, (s,t)\mapsto \sum_{k=n}^{J(n)} \alpha_{k,n}(v_k(s,t), \|v_k(s,t)\|)$$

converges strongly to (v,w) in that space, as $n\to\infty$. Taking a further subsequence (which we still denote by $(\bar v_n,\bar w_n)$) we obtain the $\mu\otimes\lambda_T$-almost everywhere convergence of $(\bar v_n,\bar w_n)$ to (v,w). Consider a negligible set $\mathscr{N}\in\mathscr{B}([0,T])\otimes\mathscr{G}$ such that for all $(s,t)\notin\mathscr{N}$, the following assertions are true: *i)* $(\bar v_n(s,t),\bar w_n(s,t))\to(v(s,t),w(s,t))$; *ii)* $(u_n(t), v_n(s,t))\to_n \mathrm{gr}(H_R(s,\,.\,))$; *iii)* $w(s,t)$ is finite. The point *ii)* is made possible by Lem. 5.6. Let $\varepsilon>0$. By conditions *ii)* and the fact that $u_n(t)\to\mathsf{z}(t)$, there exists $n = n_\varepsilon$ s.t. for all $k\ge n$, there exists $(a_k,b_k)\in\mathrm{gr}(H_R(s,\,.\,))$ satisfying $\|a_k-\mathsf{z}(t)\|<\varepsilon$ and $\|b_k - v_k(s,t)\|<\varepsilon$. If $\|z(t)\|\ge R$, obviously $(\mathsf{z}(t), v(s,t))\in\mathrm{gr}(H_R(s,\,.\,))$. We just need to consider the case where $\|z(t)\|<R$, in which case the condition $(\mathsf{z}(t), v(s,t))\in\mathrm{gr}(H_R(s,\,.\,))$ is equivalent to:

$$(\mathsf{z}(t), -v(s,t))\in\mathrm{gr}(A(s,\,.\,)+B(s,\,.\,))\,. \tag{31}$$

To show Eq. (31), consider an arbitrary $(p,q)\in\mathrm{gr}(A(s,\,.\,)+B(s,\,.\,))$. Decompose:

$$\langle q+\bar v_n(s,t), p-\mathsf{z}(t)\rangle = A_n + B_n + C_n\,, \tag{32}$$

where

$$\begin{aligned}
A_n &= \sum_{k=n}^{J(n)}\alpha_{k,n}\langle q+b_k, p-a_k\rangle\\
B_n &= \sum_{k=n}^{J(n)}\alpha_{k,n}\langle -b_k+v_k(s,t), p-a_k\rangle\\
C_n &= \sum_{k=n}^{J(n)}\alpha_{k,n}\langle q+v_k(s,t), a_k-\mathsf{z}(t)\rangle\,.
\end{aligned}$$

The left hand side of (32) converges to $\langle q+v(s,t), p-\mathsf{z}(t)\rangle$. The term A_n is positive by monotonicity of $A(s,\,.\,)+B(s,\,.\,)$. Moreover,

$$B_n \ge -\varepsilon\sum_{k=n}^{J(n)}\alpha_{k,n}\|p-a_k\| \ge -\varepsilon(\|p\| + \sup_{k\ge n}\|a_k\|)\ \ge -\varepsilon(C+\varepsilon)\,,$$

where the constant $C := \|p\| + \sup_n \|u_n(t)\|$ is finite, since $u_n(t)$ converges. Similarly,

$$C_n \geq -\varepsilon \sum_{k=n}^{J(n)} \alpha_{k,n}(\|q\| + \|v_k(s,t)\|)\,,$$

and the right hand side converges to $-\varepsilon(\|q\| + w(s,t))$. Letting $\varepsilon \to 0$, we conclude that $\langle q + v(s,t), p - \mathsf{z}(t)\rangle \geq 0$. As $A(s,\,.\,) + B(s,\,.\,) \in \mathscr{M}$, this implies that Eq. (31) holds.

References

[1] Y. F. Atchade, G. Fort, and E. Moulines. On stochastic proximal gradient algorithms. *ArXiv e-prints, 1402.2365*, February 2014.

[2] H. Attouch. Familles d'opérateurs maximaux monotones et mesurabilité. *Annali di Matematica Pura ed Applicata*, 120(1):35–111, 1979.

[3] J.-P. Aubin and A. Cellina. *Differential inclusions*, volume 264 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1984. Set-valued maps and viability theory.

[4] H. H. Bauschke and J. M. Borwein. Legendre functions and the method of random bregman projections. *Journal of Convex Analysis*, 4(1):27–67, 1997.

[5] H. H. Bauschke, J. M. Borwein, and W. Li. Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization. *Mathematical Programming*, 86(1):135–160, 1999.

[6] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011.

[7] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.

[8] M. Benaïm and M. W. Hirsch. Stochastic approximation algorithms with constant step size whose average is cooperative. *Ann. Appl. Probab.*, 9(1):216–241, 1999.

[9] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM J. Control Optim.*, 44(1):328–348 (electronic), 2005.

[10] D. P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Math. Program.*, 129(2, Ser. B):163–195, 2011.

[11] P. Bianchi. Ergodic convergence of a stochastic proximal point algorithm. *SIAM J. Optim.*, 26(4):2235–2260, 2016.

[12] P. Bianchi and W. Hachem. Dynamical behavior of a stochastic Forward-Backward algorithm using random monotone operators. *J. Optim. Theory Appl.*, 171(1):90–120, 2016.

[13] P. Bianchi, W. Hachem, and A. Salim. Constant step stochastic approximations involving differential inclusions: Stability, long-run convergence and applications. *arXiv preprint arXiv:1612.03831*, 2016.

[14] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3. Springer, New York, second edition, 2006. Theory and examples.

[15] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North-Holland mathematics studies. Elsevier Science, Burlington, MA, 1973.

[16] R. E. Bruck, Jr. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Funct. Anal.*, 18:15–26, 1975.

[17] C. Castaing and M. Valadier. *Convex analysis and measurable multifunctions.* Lecture Notes in Mathematics, Vol. 580. Springer-Verlag, Berlin-New York, 1977.

[18] P. L. Combettes and J.-C. Pesquet. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. *SIAM J. Optim.*, 25(2):1221–1248, 2015.

[19] P. L. Combettes and J.-C. Pesquet. Stochastic approximations and perturbations in forward-backward splitting for monotone operators. *Pure Appl. Funct. Anal.*, 1(1):13–37, 2016.

[20] J. de Vries. *Elements of topological dynamics*, volume 257 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1993.

[21] A. Dieuleveut, A. Durmus, and F. Bach. Bridging the gap between constant step size stochastic gradient descent and markov chains. *arXiv preprint arXiv:1707.06386*, 2017.

[22] M. Faure and G. Roth. Stochastic approximations of set-valued dynamical systems: convergence with positive probability to an attractor. *Math. Oper. Res.*, 35(3):624–640, 2010.

[23] J.-C. Fort and G. Pagès. Asymptotic behavior of a Markovian stochastic algorithm with constant step. *SIAM J. Control Optim.*, 37(5):1456–1482 (electronic), 1999.

[24] R.Z. Ha'sminskii. The average principle for parabolic and elliptic differential equations and Markov processes with small diffusions. *Theor. Probab. Appl.*, 8:1–21, 1963.

[25] F. Hiai and H. Umegaki. Integrals, conditional expectations, and martingales of multivalued functions. *Journal of Multivariate Analysis*, 7(1):149 – 182, 1977.

[26] J.-B. Hiriart-Urruty. *About properties of the mean value functional and of the continuous infimal convolution in stochastic convex analysis*, pages 763–789. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976.

[27] J.-B. Hiriart-Urruty. *Contributions à la programmation mathématique: cas déterministe et stochastique.* Université de Clermont-Ferrand II, Clermont-Ferrand, 1977. Thèse présentée à l'Université de Clermont-Ferrand II pour obtenir le grade de Docteur ès Sciences Mathématiques, Série E, No. 247.

[28] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York).* Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[29] P.-J. Laurent. *Approximation et optimisation.* Hermann, Paris, 1972. Collection Enseignement des Sciences, No. 13.

[30] I. Molchanov. *Theory of random sets.* Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005.

[31] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson et Cie, Éditeurs, Paris, 1964.

[32] J. Peypouquet and S. Sorin. Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Anal.*, 17(3-4):1113–1163, 2010.

[33] R. R. Phelps. Lectures on maximal monotone operators. *Extracta Math.*, 12(3):193–230, 1997.

[34] R. T. Rockafellar. Measurable dependence of convex sets and functions on parameters. *J. Math. Anal. Appl.*, 28:4–25, 1969.

[35] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[36] R. T. Rockafellar and R. J.-B. Wets. On the interchange of subdifferentiation and conditional expectations for convex functionals. *Stochastics*, 7(3):173–182, 1982.

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

[38] L. Rosasco, S. Villa, and B. C. Vũ. A stochastic inertial forward–backward splitting algorithm for multivariate monotone inclusions. *Optimization*, 65(6):1293–1314, 2016.

[39] G. Roth and W. H. Sandholm. Stochastic approximations with constant step size and differential inclusions. *SIAM J. Control Optim.*, 51(1):525–555, 2013.

[40] D. W. Walkup and R. J.-B. Wets. Stochastic programs with recourse. II: On the continuity of the objective. *SIAM J. Appl. Math.*, 17:98–103, 1969.