A constant step stochastic Douglas Rachford algorithm

Pascal Bianchi* Walid Hachem† Adil Salim*

October 30, 2017

Notations

For every set T, we denote by $\mathbb{R}^T$ the set of functions on $T \to \mathbb{R}$. If $\mathcal{X}$ is a Euclidean space, we denote by $\langle\, .\, ,\, .\, \rangle$ the standard inner product on $\mathcal{X}$, and by $\|\, .\, \|$ the Euclidean norm.

We denote by $\Gamma_0(\mathcal{X})$ the set of convex, proper and lower semicontinuous functions on $\mathcal{X} \to (-\infty, +\infty]$. For every $h \in \Gamma_0(\mathcal{X})$, $\gamma > 0$, we introduce the *proximity operator*

$$\mathrm{prox}_{\gamma,h}(x) = \arg\min_{w\in\mathcal{X}} h(w) + \frac{\|w-x\|^2}{2\gamma}$$

and the Moreau envelope

$$h_\gamma(x) = \min_{w\in\mathcal{X}} h(w) + \frac{\|w-x\|^2}{2\gamma}$$

for every $x \in \mathcal{X}$. We denote by $\partial h(x)$ the subdifferential of h at the point x, by $\partial h_0(x)$ the least norm element of $\partial h(x)$. Recall that h_γ is differentiable and $\nabla h_\gamma(x) = \frac{1}{\gamma}(x - \mathrm{prox}_{\gamma,h}(x))$. If A is a set, the notation ι_A stands for the indicator function of the set A, equal to zero on that set and to $+\infty$ elsewhere. If R is a linear operator, we denote by R^* the adjoint operator. When $D \subset E$, $d(x, D)$ denote the distance from the point $x \in E$ to D and if D is closed and convex, Π_D denote the projection onto D. The set of minimizers of h is the set of zeros of ∂h denoted $Z(\partial h)$.

1 Statement of the Problem

1.1 Douglas-Rachford algorithm

Consider the Problem

$$\min_{x\in\mathcal{X}} F(x) + G(x) \tag{1}$$

where $F, G \in \Gamma_0(\mathcal{X})$. The Douglas-Rachford algorithm writes

$$\begin{aligned} u_{n+1} &= \mathrm{prox}_{\gamma,F}(x_n) \\ z_{n+1} &= \mathrm{prox}_{\gamma,G}(2u_{n+1} - x_n) \\ x_{n+1} &= x_n + z_{n+1} - u_{n+1}\,. \end{aligned}$$

Under the standard qualification condition $0 \in \mathrm{ri}(\mathrm{dom}(F) - \mathrm{dom}(G))$ and assuming that the set of minimizers of problem (1) is non empty, the iterates $\mathrm{prox}_{\gamma,F}(x_n)$ converge to a minimizer of problem (1) as $n \to \infty$.

*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

1.2 Adaptive Scenario

Let ξ be a random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into an arbitrary measurable space $(\Xi, \mathcal{G})$ with distribution μ. We say that a mapping $f : \mathcal{X} \times \Xi \to (-\infty, +\infty]$ is a normal convex integrand if $f(\,.\,, s) \in \Gamma_0(\mathcal{X})$ for every $s \in \Xi$ and if $f(x, \,.\,)$ is measurable for every $x \in \mathcal{X}$. The expectation $\mathbb{E}(|f(x,\xi)|)$ according to the random variable ξ is supposed to be finite.

From now on, assume that the mapping F and G are of the form

$$F(x) = \mathbb{E}(f(x,\xi))$$
$$G(x) = \mathbb{E}(g(x,\xi)) \ ,$$

where f, g are normal convex integrands. Denote by $(\xi_n : n \in \mathbb{N})$ a sequence of iid copies of the r.v. ξ. In the sequel, we use the notation $f_n := f(\,.\,, \xi_n)$ and $g_n := g(\,.\,, \xi_n)$. The adaptive Douglas-Rachford algorithm is given by

$$\begin{aligned} u_{n+1} &= \operatorname{prox}_{\gamma, f_{n+1}}(x_n) \\ z_{n+1} &= \operatorname{prox}_{\gamma, g_{n+1}}(2u_{n+1} - x_n) \\ x_{n+1} &= x_n + z_{n+1} - u_{n+1} \, . \end{aligned}$$

We denote by $D(s)$ the domain of $g(\cdot, s)$, and by $\mathcal{D}$ the set defined by the relation $x \in \mathcal{D} \iff x \in D(\xi)$ a.s. We denote by $\boldsymbol{d}(x) = d(x, \mathcal{D})$. We also denote $F^\gamma(x) = \int f_\gamma(x,s)\mu(ds)$ and $G^\gamma(x) = \int g_\gamma(x,s)\mu(ds)$. We assume that $f(\cdot,\xi)$ has a.s a full domain (equal to $\mathcal{X}$) and is continuously differentiable. Under these assumptions, $Z(\partial(G+F)) = Z(\partial G + \nabla F) = Z(\mathbb{E}(\partial g(\cdot,\xi)) + \mathbb{E}(\nabla f(\cdot,\xi)))$ [33].

1.3 Useful facts

We first observe that the process (x_n) described by Eq. (**??**) is a homogeneous Markov chain with transition kernel denoted by P_γ. 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.

Then, we 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. (**??**) shows that $P_\gamma(\cdot, f) \in C_b(E)$ by the continuity of $\operatorname{prox}_{\gamma g(\cdot,s)}$ 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 $\gamma \in (0, \gamma_0]$.

Finally, it is a standard fact of the monotone operator theory that for any x_0 in the domain of $F+G$, the Differential Inclusion (DI)

$$\begin{cases} \dot{\mathsf{x}}(t) & \in & -(\partial F + \partial G)(\mathsf{x}(t)) \\ \mathsf{x}(0) & = & x_0 \end{cases} \tag{2}$$

admits a unique absolutely continuous solution on $\mathbb{R}_+ := [0, \infty)$.

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

2 Theorem

H1 There exists $x_\star \in Z(\partial G + \nabla F)$ admitting a $\mathcal{L}^2(\mu)$ representation (φ, ψ) *i.e* $\exists \varphi, \psi \in \mathcal{L}^2(\mu)$, such that $\varphi(\xi) \in \partial g(x_\star, \xi)$ a.s, $\psi(\xi) = \nabla f(x_\star, \xi)$ a.s and $\mathbb{E}(\varphi(\xi) + \psi(\xi)) = 0$.

H2 There exists $L > 0$ s.t. $\nabla f(\cdot, \xi)$ is a.s L-Lipschitz continuous.

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

(a) $F + G$ is coercive.

(b) $F + G$ is supercoercive.

H4 For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

$$\sup_{x \in \mathcal{K} \cap \mathcal{D}} \int \|\partial g_0(x,s)\|^{1+\varepsilon} \, \mu(ds) < \infty,$$

H5 There exists a closed ball in E such that $\|\nabla f(x,s)\| \le M(s)$ for all x in this ball, where $M(s)$ is μ-integrable. Moreover, for every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

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

H6 For all $\gamma \in (0, \gamma_0]$ and all $x \in E$,

$$\int \left(\|\nabla f_\gamma(x,s)\| + \frac{1}{\gamma} \| \operatorname{prox}_{\gamma g(\cdot,s)}(x) - \Pi_{\mathrm{cl}(D(s))}(x) \| \right) \mu(ds) \le C(1 + |F^\gamma(x) + G^\gamma(x)|) \, .$$

H7 $\forall x \in E$, $\displaystyle\int d(x, D(s))^2 \, \mu(ds) \ge C \boldsymbol{d}(x)^2$.

H8 For every compact set $\mathcal{K} \subset E$, there exists $\varepsilon > 0$ such that

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

Theorem 2.1. Let Hypotheses H1–H8 hold true. Then, for each probability measure ν having a finite second moment, for any $\varepsilon > 0$,

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

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

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

where $\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$.

3 Proof of Theorem 2.1

In this section, we study the iterations given by the adaptive Douglas Rachford algorithm. Let $\gamma_0 > 0$, $a \in E$ and $(\xi_n)_{n\in\mathbb{N}}$ be an i.i.d sequence of random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\Xi, \mathcal{G})$ with distribution μ. The adaptive Douglas Rachford algorithm with step size $\gamma > 0$ writes $x_0 = a$ and for all $n \in \mathbb{N}$,

$$x_{n+1} = x_n - \gamma\nabla f_\gamma(x_n, \xi_{n+1}) - \gamma\nabla g_\gamma(x_n - 2\gamma\nabla f_\gamma(x_n, \xi_{n+1}), \xi_{n+1}). \tag{3}$$

Define

$$h_\gamma(x, s) := -\nabla f_\gamma(x, s) - \nabla g_\gamma(x - 2\gamma\nabla f_\gamma(x, s), s).$$

The algorithm (3) can be rewritten as

$$x_{n+1} = x_n + \gamma h_\gamma(x_n, \xi_{n+1}). \tag{4}$$

In Sec. 3.1, we show that the linearly interpolated process constructed from the sequence (x_n) with time frame γ converges narrowly as $\gamma \to 0$ to a Differential Inclusion (DI) solution in the topology of uniform convergence on compact sets. The main result of this section is Th. 3.1, which has its own interest. To prove this theorem, we establish the tightness of the linearly interpolated process (Lem. 3.2), then we show that the limit points coincide with the DI solution (Lem. 3.3–3.5). In Sec. 3.2, we start by establishing the inequality (17), which implies the tightness of the set of invariant measures $\mathcal{I}(\mathcal{P})$ in Lem 3.7. Then, we show that the cluster points of $\mathcal{I}(\mathcal{P})$ are invariant measures for the flow induced by the DI (Lem 3.9). 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. 3.8.

3.1 Weak APT

For every $\gamma > 0$, we introduce the linearly interpolated process

$$\mathsf{X}_\gamma : (E^{\mathbb{N}}, \mathscr{B}(E)^{\otimes\mathbb{N}}) \to (C(\mathbb{R}_+, E), \mathscr{B}(C(\mathbb{R}_+, E)))$$

, 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 $\mathbb{P}^{\nu,\gamma}\mathsf{X}_\gamma^{-1}$.

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

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

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 beginning section is to establish the following result:

Theorem 3.1. Let Assumptions H4, H5, H7 and H8 hold true. Then, for every $\eta > 0$, for every compact set $\mathcal{K} \subset E$ s.t. $\mathcal{K} \cap \mathcal{D} \neq \emptyset$,

$$\forall M \geq 0, \sup_{a \in \mathcal{K} \cap \mathcal{D}_{\gamma M}} \mathbb{P}^{a,\gamma}\left(\mathsf{d}(\mathsf{X}_\gamma, \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), \cdot)) > \eta\right) \xrightarrow[\gamma \to 0]{} 0. \tag{5}$$

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 $\bar{\mathbb{P}}^{a,\gamma} := \mathbb{P}^{a,\gamma} C_R^{-1}$, which corresponds to the law of the *truncated* process $C_R(x)$ and denote by $\bar{\mathbb{E}}^{a,\gamma}$ the corresponding mathematical expectation. The crux of the proof consists in showing that for every $\eta > 0$ and every $M > 0$,

$$\sup_{a \in \mathcal{K} \cap \mathcal{D}_{\gamma M}} \bar{\mathbb{P}}^{a,\gamma}\left(\mathsf{d}(\mathsf{X}_\gamma, \Phi(\Pi_{\mathrm{cl}(\mathcal{D})}(a), \cdot)) > \eta\right) \xrightarrow[\gamma \to 0]{} 0. \tag{6}$$

Eq. (6) 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. (6). 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 at the end of this section.

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

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

$\bar{\mathbb{P}}^{a,\gamma}$ 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

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

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

Lemma 3.2. For all $\gamma \in (0, \gamma_0]$ and all $x \in E^{\mathbb{N}}$, define $Z_{n+1}^\gamma(x) := \gamma^{-1}(x_{n+1} - x_n)$. There exists $\varepsilon > 0$ such that:

$$\sup_{n \in \mathbb{N}, a \in \mathcal{K} \cap \mathcal{D}_{\gamma M}, \gamma \in (0,\gamma_0]} \bar{\mathbb{E}}^{a,\gamma}\left(\left(\|Z_n^\gamma\| + \frac{\boldsymbol{d}(x_n)}{\gamma}\mathbb{1}_{\|x_n\| \leq R}\right)^{1+\varepsilon}\right) < +\infty \tag{7}$$

Proof. Let ε be the smallest of the three constants (also named ε) in Assumptions H4, H5 and H8 respectively where $\mathcal{K} = B_R$. For every a, γ, the following holds for $\bar{\mathbb{P}}^{a,\gamma}$-almost all $x = (x_n : n \in \mathbb{N})$:

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

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

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

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

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

Therefore, by setting $\delta_\gamma(x,s) := \|x + \gamma h_\gamma(x,s) - \Pi_{D(s)}(x)\|$,

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

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

$$\delta_\gamma(x,s) = \operatorname{prox}_{\gamma g(\cdot,s)}(x - 2\gamma\nabla f_\gamma(x,s)) + \gamma\nabla f_\gamma(x,s) - \Pi_{D(s)}(x) + \operatorname{prox}_{\gamma g(\cdot,s)}(x) - \operatorname{prox}_{\gamma g(\cdot,s)}(x)$$

Hence,

$$\|\delta_\gamma(x,s)\| \leq 3\gamma\|\nabla f_\gamma(x,s)\| + \|\operatorname{prox}_{\gamma g(\cdot,s)}(x) - \Pi_{D(s)}(x)\|$$

And, by Assumptions H4 and H5, there exists a deterministic constant $C>0$ s.t.

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

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,

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

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

$$\|\Pi_{\mathrm{cl}(D(s))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^2 \leq \boldsymbol{d}(x_n)^2 - d(x_n, D(s))^2. \tag{8}$$

Making use of Assumption H7, and assuming without loss of generality that $\varepsilon \leq 1$, we obtain

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

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

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

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

$$\sup_{n\in\mathbb{N}, a\in\mathcal{K}\cap\mathcal{D}_{\gamma M}, \gamma\in(0,\gamma_0]} \bar{\mathbb{E}}^{a,\gamma}\left(\left(\frac{\boldsymbol{d}(x_n)}{\gamma}\right)^{1+\varepsilon}\mathbb{1}_{\|x_n\|\leq R}\right) < +\infty . \tag{9}$$

Since $\nabla g_\gamma(\cdot,s)$ is γ^{-1}-Lipschitz continuous, $\|h_\gamma(x,s)\| \leq \|\nabla g_\gamma(x,s)\| + 3\|\nabla f_\gamma(x,s)\|$. Moreover, choosing measurably $\tilde{x} \in \mathcal{D}$ in such a way that $\|x - \tilde{x}\| \leq 2\boldsymbol{d}(x)$, we obtain $\|\nabla g_\gamma(x,s)\| \leq \|\partial g_0(\tilde{x},s)\| + 2\frac{\boldsymbol{d}(x)}{\gamma}$. Therefore, there exists R' depending only on R and $\mathcal{D}$ s.t.

$$\|\nabla g_\gamma(x,s)\|\mathbb{1}_{\|x\|\leq R} \leq \|\partial g_0(\tilde{x},s)\|\mathbb{1}_{\|\tilde{x}\|\leq R'} + 2\frac{\boldsymbol{d}(x)}{\gamma}\mathbb{1}_{\|x\|\leq R} .$$

In the following, C is a positive constant that can change from a line to another. Choosing $\varepsilon > 0$ enough small and using Assumption H4, H5 and Eq. (9), we have

$$
\begin{aligned}
\bar{\mathbb{E}}_n^{a,\gamma}(\|Z_{n+1}^\gamma\|^{1+\varepsilon}) = & \int \|h_\gamma(x_n,s)\|^{1+\varepsilon} \mathbb{1}_{\|x_n\|\le R}\, d\mu(s) \\
\le & \int \left(\|\nabla g_\gamma(x_n,s)\| + 3\|\nabla f_\gamma(x_n,s)\|\right)^{1+\varepsilon} \mathbb{1}_{\|x_n\|\le R}\, d\mu(s) \\
\le & C \int \|\nabla g_\gamma(x_n,s)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R} + \|\nabla f_\gamma(x_n,s)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\, d\mu(s) \\
\le & C \int \|\partial g_0(\tilde{x}_n,s)\|^{1+\varepsilon}\mathbb{1}_{\|\tilde{x}_n\|\le R'}\, d\mu(s) + C\int \|\nabla f_\gamma(x_n,s)\|^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\, d\mu(s) \\
& + C\frac{\boldsymbol{d}(x_n)}{\gamma}^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R} \\
\le & C + C\frac{\boldsymbol{d}(x_n)}{\gamma}^{1+\varepsilon}\mathbb{1}_{\|x_n\|\le R}\,.
\end{aligned} \tag{10}
$$

Taking expectations, the bound (7) is established.

□

Using [13, Lemma 4.2], the uniform integrability condition (7) 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 $T>0$,

$$
\sup_{a\in\mathcal{K}\cap\mathcal{D}_{\gamma M}} \bar{\mathbb{P}}^{a,\gamma}(\|\mathsf{X}_\gamma \circ M_{\gamma,R}\|_{\infty,T} > \varepsilon) \xrightarrow{\gamma\to 0} 0\,, \tag{11}
$$

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

Lemma 3.3. 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 $\bar{\mathbb{P}}^{a_n,\gamma_n}\mathsf{X}_{\gamma_n}^{-1}$ 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.

$$
\sup_n \int_0^T \left(\frac{\boldsymbol{d}(u_n(t))}{\gamma_n}\mathbb{1}_{\|u_n(t)\|\le R}\right)^{1+\frac{\varepsilon}{2}} dt \; < +\infty \quad (\forall T>0)\,, \tag{12}
$$

where $\varepsilon>0$ is the constant introduced in Lem. 3.2.

Proof. 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. 3.2. 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. 3.2, it is easily seen that

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

[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.

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 $T \ge 0$, $\Gamma_{\gamma_n}(\mathsf{x}_n)(T) \to \mathsf{w}(T)$, which is finite. Hence, $\sup_n \Gamma_{\gamma_n}(\mathsf{x}_n)(T) < \infty$ on that event, which proves the second point.

□

Define

$$v_n(s,t) := -\nabla f_{\gamma_n}(u_n(t), s)\mathbb{1}_{\|u_n(t)\| \le R} .$$

and

$$w_n(s,t) := -\nabla g_{\gamma_n}(u_n(t) - 2\gamma\nabla f_{\gamma_n}(u_n(t), s), s)\mathbb{1}_{\|u_n(t)\| \le R} .$$

Thanks to the convergence (11), the following holds $\mathbb{P}'$-a.e.:

$$\mathsf{z}(t) = \mathsf{z}(0) + \lim_{n\to\infty} \int_0^t \int_\Xi v_n(s,u) + w_n(s,u)\, \mu(ds)\, du \qquad (\forall t \ge 0) . \tag{13}$$

We now select an $\omega \in \Omega'$ s.t. the events (12) and (13) are all realized, and omit the dependence in ω in the sequel. Otherwise stated, u_n, v_n and w_n are handled from now on as deterministic functions, and no longer as random variables. The aim of the next lemmas is to analyze the integrand $v_n(s,u) + w_n(s,u)$. 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 3.4. The sequences $(v_n : n \in \mathbb{N}), (w_n : n \in \mathbb{N})$ form bounded subsets of $\mathcal{L}_E^{1+\varepsilon/2}$.

Proof. By the same derivations as those leading to Eq. (10), there exists $C > 0$ such that

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

The proof is concluded by applying Lem. 3.3.

□

The sequence of mappings $((s,t) \mapsto (v_n(s,t), w_n(s,t)))$ is bounded in $\mathcal{L}_{E^2}^{1+\varepsilon/2}$ 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 $w : \Xi \times [0,T] \to E$. Let $H_R(x,s) := \nabla f(x,s) + \partial g(x,s))$ if $\|x\| < R$, $\{0\}$ if $\|a\| > R$, and $H_R(x,s) := E$ otherwise. vérifier

Denote the corresponding selection integral as $\mathsf{H}_R(a) = \int H_R(s,a)\, \mu(ds)$. à définir plus haut

Lemma 3.5. For every (s,t) $\mu \otimes \lambda_T$-a.e., $(\mathsf{z}(t), (v+w)(s,t)) \in \mathrm{gr}(H_R(s,\,.))$.

Proof. To simplify notations, we now omit the dependence in (s,t) in the sequel and write $u_n := u_n(t)$, $v_n := v_n(s,t)$, $w_n := w_n(s,t)$, $h_\gamma := h_\gamma(\cdot,s)$, $\partial g := \partial g(\cdot,s)$, $\nabla f := \nabla f(\cdot,s)$, $\gamma := \gamma_n$, $\mathrm{prox}_{\gamma f} := \mathrm{prox}_{\gamma f(\cdot,s)}$, $\nabla f_\gamma := \nabla f_\gamma(\cdot,s)$, $\mathsf{z} := \mathsf{z}(t)$. Moreover, we write $\widetilde{\mathrm{prox}}_{\gamma g}(x) := \mathrm{prox}_{\gamma g(\cdot,s)}(x - 2\gamma\nabla f_\gamma(x,s))$ and $\widetilde{\nabla} g_\gamma := \nabla g_\gamma(x - 2\gamma \nabla f_\gamma(x,s), s)$, for all $x \in E$.

There exists a subsequence of (γ_n) that is decreasing and such that $d_n := \sup_{t\in[0,T]} \|u_n(t) - \mathsf{z}(t)\|$ is decreasing to zero. We still denote by (γ_n) such a subsequence. The sequence

$$((v_n, w_n, \|v_n\|, \|w_n\|))_n$$

converges weakly to $(v, w, \tilde v, \tilde w)$ in $\mathcal{L}^{1+\varepsilon/2}_{E^2\times\mathbb{R}^2}$ 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 $((v_n, w_n, \|v_n\|, \|w_n\|))_n$ 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

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

converges strongly to $(v, w, \tilde v, \tilde w)$ in that space, as $n \to \infty$. Taking a further subsequence (which we still denote by $(\bar v_n, \bar w_n, \tilde v_n, \tilde w_n)$) we obtain the $\mu \otimes \lambda_T$-almost everywhere convergence of $(\bar v_n, \bar w_n, \tilde v_n, \tilde w_n)$ to $(\bar v, \bar w, \tilde v, \tilde w)$. Consider a negligible set $\mathscr{N} \in \mathscr{G} \otimes \mathscr{B}([0,T])$ such that for all $(s,t) \notin \mathscr{N}$, $(\bar v_n, \bar w_n, \tilde v_n, \tilde w_n) \to (v, w, \tilde v, \tilde w)$ and $\tilde v, \tilde w$ are finite.

If $\|z(t)\| \ge R$, obviously $(\mathsf{z}(t), (v+w)(s,t)) \in \mathrm{gr}(H_R(\cdot,s))$. We just need to consider the case where $\|z(t)\| < R$. Besides, the condition $(\mathsf{z}(t), (v+w)(s,t)) \in \mathrm{gr}(H_R(\cdot,s))$ is equivalent to:

$$(\mathsf{z}(t), -(v+w)(s,t)) \in \mathrm{gr}(\partial(f(\cdot,s) + g(\cdot,s))) = \mathrm{gr}(\nabla f(\cdot,s) + \partial g(\cdot,s))\,. \tag{14}$$

To show Eq. (14), consider an arbitrary $(p,q) \in \mathrm{gr}(\nabla f(\cdot,s) + \partial g(\cdot,s))$. There exists $(q_f, q_g) \in E^2$ such that $q = q_f + q_g$, $(p, q_f) \in \mathrm{gr}(\nabla f(\cdot,s))$ and $(p, q_g) \in \mathrm{gr}(\partial g(\cdot,s))$.

Recall that $-h_\gamma(x) = \nabla f_\gamma(x) + \nabla g_\gamma(x - 2\gamma\nabla f_\gamma(x))$. We start by decomposing $\langle x-p, -h_\gamma(x)-q\rangle$ for any $x \in E$. On the one hand,

$$\langle x + \gamma h_\gamma(x) - p, -h_\gamma(x) - q\rangle = -\gamma\langle h_\gamma(x), q\rangle - \gamma\|h_\gamma(x)\|^2 + \langle x - p, -h_\gamma(x) - q\rangle$$

On the other hand,

$$\begin{aligned}
&\langle x - \gamma\nabla f_\gamma(x) - \gamma\widetilde{\nabla} g_\gamma(x) - p, \nabla f_\gamma(x) + \widetilde{\nabla} g_\gamma(x) - (q_f + q_g)\rangle \\
=&\langle \mathrm{prox}_{\gamma f}(x) - \gamma\widetilde{\nabla} g_\gamma(x) - p, \nabla f_\gamma(x) - q_f\rangle + \langle \widetilde{\mathrm{prox}}_{\gamma g}(x) + \gamma\nabla f_\gamma(x) - p, \widetilde{\nabla} g_\gamma(x) - q_g\rangle \\
=&\langle \mathrm{prox}_{\gamma f}(x) - p, \nabla f_\gamma(x) - q_f\rangle + \langle \widetilde{\mathrm{prox}}_{\gamma g}(x) - p, \widetilde{\nabla} g_\gamma(x) - q_g\rangle \\
&- \gamma\langle \widetilde{\nabla} g_\gamma(x), \nabla f_\gamma(x) - q_f\rangle + \gamma\langle \nabla f_\gamma(x), \widetilde{\nabla} g_\gamma(x) - q_g\rangle \\
=&\langle \mathrm{prox}_{\gamma f}(x) - p, \nabla f_\gamma(x) - q_f\rangle + \langle \widetilde{\mathrm{prox}}_{\gamma g}(x) - p, \widetilde{\nabla} g_\gamma(x) - q_g\rangle \\
&+ \gamma\langle \widetilde{\nabla} g_\gamma(x), q_f\rangle - \gamma\langle \nabla f_\gamma(x), q_g\rangle.
\end{aligned}$$

Using the monotonicity of ∇f and ∂g, we finally have

$$\begin{aligned}
0 \le &\langle x - p, -h_\gamma(x) - q\rangle \\
&- \gamma\langle \tilde\nabla g_\gamma(x), q_f\rangle + \gamma\langle \nabla f_\gamma(x), q_g\rangle - \gamma\langle h_\gamma(x), q\rangle.
\end{aligned} \tag{15}$$

As $\|\mathsf{z}\| < R$, it holds that $\|u_n\| < R$ for every n large enough. Thus, $-v_n = \nabla f_{\gamma_n}(u_n)$ and $-w_n = \nabla g_{\gamma_n}(u_n - 2\gamma_n \nabla f_{\gamma_n}(u_n))$. Using (15) with u_n instead of x and γ_n instead of γ, we have

$$\begin{aligned}
0 \leq & \sum_{k=n}^{J(n)} \alpha_{k,n} \left(\langle z - p, -(v_k + w_k) - q \rangle + \langle u_k - z, -(v_k + w_k) - q \rangle \right) \\
& + \sum_{k=n}^{J(n)} \alpha_{k,n} \gamma_k \left(\langle w_k, q_f \rangle - \langle v_k, q_g \rangle - \langle h_{\gamma_k}(u_k), q \rangle \right) \\
\leq & \langle z - p, -(\bar{v}_n + \bar{w}_n) - q \rangle + \sum_{k=n}^{J(n)} \alpha_{k,n} d_k \left(\|v_k\| + \|w_k\| + \|q\| \right) \\
& + \sum_{k=n}^{J(n)} \alpha_{k,n} \gamma_k \left(\|w_k\|\|q_f\| + \|v_k\|\|q_g\| + \|v_k\|\|q\| + \|w_k\|\|q\| \right) \\
\leq & \langle z - p, -(\bar{v}_n + \bar{w}_n) - q \rangle + d_n \left(\tilde{v}_n + \tilde{w}_n + \|q\| \right) \\
& + \gamma_n \left(\tilde{w}_n \|q_f\| + \tilde{v}_n \|q_g\| + \tilde{v}_n \|q\| + \tilde{w}_n \|q\| \right) . \qquad (16)
\end{aligned}$$

Letting $n \to +\infty$, since (v_n) and (w_n) are a.e bounded sequences in E, we conclude that $\langle \mathsf{z} - p, -(v + w) - q \rangle \geq 0$ a.e. As $\nabla f + \partial g \in \mathscr{M}$, this implies that $(\mathsf{z}, (v + w)) \in \nabla f + \partial g$ a.e.

□

By Lem. 3.5 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

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

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:

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

Let $T > 0$ and $R > \sup\{\|\Phi(a, t)\| : t \in [0, T], a \in \mathcal{K} \cap \mathrm{cl}(\mathcal{D})\}$ (the latter quantity being finite, see *e.g.* [15]). Consider any solution x to the DI $\dot{\mathsf{x}} \in \mathsf{H}_R(\mathsf{x})$ with initial condition $a \in \mathcal{K} \cap \mathrm{cl}(\mathcal{D})$. Consider the set $F = \{t \in [0, T], \mathsf{x}(t) = \Phi(a, t)\}$. Then, $0 \in F$. Let $\bar{t} = \sup F$ and assume that $\bar{t} < T$. Since F is closed, $\bar{t} \in F$ and we have $\|\mathsf{x}(\bar{t})\| < R$, hence there exists $\varepsilon > 0$ such that $\|\mathsf{x}(t)\| < R$ for all $t \in [\bar{t}, \bar{t} + \varepsilon]$. Then, x and $\Phi(a, \cdot)$ are solutions to the DI $\dot{\mathsf{x}} \in \mathsf{H}(\mathsf{x})$ over $[\bar{t}, \bar{t} + \varepsilon]$ and $\mathsf{x}(\bar{t}) = \Phi(a, \bar{t})$, therefore $\mathsf{x}(t) = \Phi(a, t)$ for all $t \in [\bar{t}, \bar{t} + \varepsilon]$. Hence, $\bar{t} + \varepsilon \in F$. Finally, $\bar{t} = T$ and $F = [0, T]$. By the same arguments as in [13, Section 4 - End of the proof], Theorem 3.1 follows.

3.2 Stability

Theorem 3.6. Assume hypotheses H1 and H2. Let $x_\star \in Z(\partial G + \nabla F)$ that admits a $\mathcal{L}^2$ representation. Then, there exists $\alpha, C > 0$ such that

$$\mathbb{E}_n^{\gamma, a} \|x_{n+1} - x_\star\|^2 \leq \|x_n - x_\star\|^2 - \alpha\gamma(F^\gamma(x_n) + G^\gamma(x_n)) + \gamma C. \qquad (17)$$

for γ enough close to 0.

Proof. To simplify notations, we now omit the dependence in (s,t) in the sequel and write $u_n := u_n(t)$, $v_n := v_n(s,t)$, $w_n := w_n(s,t)$, $h_\gamma := h_\gamma(\cdot,s)$, $\partial g := \partial g(\cdot,s)$, $\nabla f := \nabla f(\cdot,s)$, $\gamma := \gamma_n$, $\mathrm{prox}_{\gamma f} := \mathrm{prox}_{\gamma f(\cdot,s)}$, $\nabla f_\gamma := \nabla f_\gamma(\cdot,s)$, $\mathsf{z} := \mathsf{z}(t)$. Moreover, we write $\widetilde{\mathrm{prox}}_{\gamma g}(x) := \mathrm{prox}_{\gamma g(\cdot,s)}(x - 2\gamma\nabla f_\gamma(x,s))$ and $\widetilde{\nabla} g_\gamma := \nabla g_\gamma(x - 2\gamma\nabla f_\gamma(x,s),s)$, for all $x \in E$.

By assumption, there exists a $\mathcal{L}^2$ representation (φ,ψ) of $x_\star$. We write

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

We also write

$$\begin{aligned}\langle \widetilde{\nabla} g_\gamma(x), x - x_\star\rangle &= \langle \widetilde{\nabla} g_\gamma(x) - \varphi, x - x_\star\rangle + \langle \varphi, x - x_\star\rangle\\ &= \langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\mathrm{prox}}_{\gamma g}(x) - x_\star\rangle + \langle \widetilde{\nabla} g_\gamma(x) - \varphi, x - \widetilde{\mathrm{prox}}_{\gamma g}(x) - 2\gamma\nabla f_\gamma(x)\rangle\\ &\quad + \langle \varphi, x - x_\star\rangle + \langle \widetilde{\nabla} g_\gamma(x) - \varphi, 2\gamma\nabla f_\gamma(x)\rangle\\ &= \langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\mathrm{prox}}_{\gamma g}(x) - x_\star\rangle - \gamma\langle \varphi, \widetilde{\nabla} g_\gamma(x)\rangle\\ &\quad + 2\gamma\langle \widetilde{\nabla} g_\gamma(x), \nabla f_\gamma(x)\rangle + \langle \varphi, x - x_\star\rangle + \gamma\|\widetilde{\nabla} g_\gamma(x)\|^2 - 2\gamma\langle \varphi, \nabla f_\gamma(x)\rangle.\end{aligned}$$

Hence,

$$\begin{aligned}&\langle \nabla f_\gamma(x) + \widetilde{\nabla} g_\gamma(x), x - x_*\rangle\\ =&\langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\mathrm{prox}}_{\gamma g}(x) - x_\star\rangle + \langle \nabla f_\gamma(x) - \psi, \mathrm{prox}_{\gamma f}(x) - x_\star\rangle\\ &+ \gamma\|\widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\|^2 - \gamma\left\{\langle \varphi + \psi, \nabla f_\gamma(x)\rangle + \langle \varphi, \widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\rangle\right\}\\ &+ \langle \varphi + \psi, x - x_\star\rangle\end{aligned} \tag{18}$$

By expanding

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

we obtain

$$\begin{aligned}\|x_{n+1} - x_\star\|^2 = \|x_n - x_\star\|^2 - 2\gamma\langle \widetilde{\nabla} g_\gamma(x_n), x_n - x_\star\rangle - 2\gamma\langle \nabla f_\gamma(x_n), x_n - x_\star\rangle\\ + \gamma^2\|\widetilde{\nabla} g_\gamma(x_n) + \nabla f_\gamma(x_n)\|^2.\end{aligned} \tag{19}$$

Using (18), we obtain

$$\begin{aligned}\|x_{n+1} - x_\star\|^2 = \|x_n - x_\star\|^2&\\ &- 2\gamma\left\{\langle \widetilde{\nabla} g_\gamma(x_n) - \varphi, \widetilde{\mathrm{prox}}_{\gamma g}(x_n) - x_\star\rangle + \langle \nabla f_\gamma(x_n) - \psi, \mathrm{prox}_{\gamma f}(x_n) - x_\star\rangle\right\}\\ &- \gamma^2\|\widetilde{\nabla} g_\gamma(x_n) + \nabla f_\gamma(x_n)\|^2 + 2\gamma^2\left\{\langle \varphi + \psi, \nabla f_\gamma(x_n)\rangle + \langle \varphi, \widetilde{\nabla} g_\gamma(x_n) + \nabla f_\gamma(x_n)\rangle\right\}\\ &- 2\gamma\langle \varphi + \psi, x_n - x_\star\rangle\end{aligned}$$

where we used $\|a+b\|^2 = \|a\|^2 + \|b\|^2 + 2\langle a,b\rangle$. Then, since $2\langle a,b\rangle \le \|a\|^2/2 + 2\|b\|^2$,

$$\begin{aligned}\|x_{n+1} - x_\star\|^2 \le \|x_n - x_\star\|^2&\\ &- 2\gamma\left\{\langle \widetilde{\nabla} g_\gamma(x_n) - \varphi, \widetilde{\mathrm{prox}}_{\gamma g}(x_n) - x_\star\rangle + \langle \nabla f_\gamma(x_n) - \psi, \mathrm{prox}_{\gamma f}(x_n) - x_\star\rangle\right\}\\ &- \gamma^2/2\|\widetilde{\nabla} g_\gamma(x_n) + \nabla f_\gamma(x_n)\|^2 + \gamma^2/2\|\nabla f_\gamma(x_n)\|^2\\ &+ 2\gamma^2\|\varphi\|^2 + 2\gamma^2\|\varphi + \psi\|^2 - 2\gamma\langle \varphi + \psi, x_n - x_\star\rangle.\end{aligned} \tag{20}$$

Observe that the term between the braces at the right hand side of the last inequality is nonnegative thanks to the monotonicity of $\nabla f(\cdot, s)$ and $\partial g(\cdot, s)$.

Let $x \in E$. By the convexity of g_γ and f_γ, we have

$$g_\gamma(x - 2\gamma\nabla f_\gamma(x)) - g_\gamma(x_\star) \le \langle \widetilde{\nabla} g_\gamma(x), x - 2\gamma \nabla f_\gamma(x) - x_\star\rangle \tag{21}$$

and

$$f_\gamma(x) - f_\gamma(x_\star) \le \langle \nabla f_\gamma(x), x - x_\star\rangle. \tag{22}$$

Using the $1/\gamma$-Lipschitz continuity of ∇g_γ we have

$$g_\gamma(x) - g_\gamma(x - 2\gamma\nabla f_\gamma(x)) \le \langle \widetilde{\nabla} g_\gamma(x), 2\gamma \nabla f_\gamma(x)\rangle + 2\gamma\|\nabla f_\gamma(x)\|^2. \tag{23}$$

Summing the inequalities (21), (22) and (23) we obtain

$$f_\gamma(x) - f_\gamma(x_*) + g_\gamma(x) - g_\gamma(x_*) \le \langle \nabla f_\gamma(x) + \widetilde{\nabla} g_\gamma(x), x - x_*\rangle + 2\gamma\|\nabla f_\gamma(x)\|^2 \tag{24}$$

Using (18),

$$\begin{aligned}
& f_\gamma(x) - f_\gamma(x_*) + g_\gamma(x) - g_\gamma(x_*) \\
\le\ & 2\gamma\|\nabla f_\gamma(x)\|^2 \\
& + \langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\operatorname{prox}}_{\gamma g}(x) - x_\star\rangle + \langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle \\
& + \gamma\|\widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\|^2 - \gamma\left\{\langle \varphi + \psi, \nabla f_\gamma(x)\rangle + \langle \varphi, \widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\rangle\right\} \\
& + \langle \varphi + \psi, x - x_\star\rangle \\
\le\ & \frac{3}{2}\gamma\|\nabla f_\gamma(x)\|^2 \\
& + \langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\operatorname{prox}}_{\gamma g}(x) - x_\star\rangle + \langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle \\
& + \frac{3}{2}\gamma\|\widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\|^2 + \langle \varphi + \psi, x - x_\star\rangle + \frac{\gamma}{2}\|\varphi\|^2 + \frac{\gamma}{2}\|\varphi + \psi\|^2 \\
\le\ & -\frac{3}{2}\gamma\|\nabla f_\gamma(x)\|^2 + 3\left\{\gamma\|\nabla f_\gamma(x)\|^2 - \langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle\right\} \\
& + 6\left\{\langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\operatorname{prox}}_{\gamma g}(x) - x_\star\rangle + \langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle\right\} \\
& + \frac{3}{2}\gamma\|\widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\|^2 + \langle \varphi + \psi, x - x_\star\rangle + \frac{\gamma}{2}\|\varphi\|^2 + \frac{\gamma}{2}\|\varphi + \psi\|^2
\end{aligned}$$

Since $\langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\operatorname{prox}}_{\gamma g}(x) - x_\star\rangle$ and $\langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle$ are nonnegative. Using $\psi = \nabla f(x_\star)$, $\nabla f_\gamma(x) = \nabla f(\operatorname{prox}_{\gamma f}(x))$ and Assumption H2, there exists by Baillon-Haddad theorem $c > 0$ such that $c\|\nabla f_\gamma(x) - \psi\|^2 \le \langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle$. Then,

$$-\langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle \le -c\|\nabla f_\gamma(x) - \psi\|^2 \le -c/2\|\nabla f_\gamma(x)\|^2 + c\|\psi\|^2.$$

If $\gamma < c/2$ we finally have

$$\begin{aligned}
& f_\gamma(x) - f_\gamma(x_*) + g_\gamma(x) - g_\gamma(x_*) \\
& \le -\frac{3}{2}\gamma\|\nabla f_\gamma(x)\|^2 + 3c\|\psi\|^2 \\
& \quad + 6\left\{\langle \widetilde{\nabla} g_\gamma(x) - \varphi, \widetilde{\operatorname{prox}}_{\gamma g}(x) - x_\star\rangle + \langle \nabla f_\gamma(x) - \psi, \operatorname{prox}_{\gamma f}(x) - x_\star\rangle\right\} \\
& \quad + \frac{3}{2}\gamma\|\widetilde{\nabla} g_\gamma(x) + \nabla f_\gamma(x)\|^2 + \langle \varphi + \psi, x - x_\star\rangle + \frac{\gamma}{2}\|\varphi\|^2 + \frac{\gamma}{2}\|\varphi + \psi\|^2.
\end{aligned}$$

Using 20, there exists $\alpha, C, C' > 0$ such that

$$\begin{aligned}\|x_{n+1} - x_\star\|^2 \le \|x_n - x_\star\|^2 \\ - \alpha\gamma \{f_\gamma(x_n) + g_\gamma(x_n) - f_\gamma(x_\star) - g_\gamma(x_\star)\} \\ + C\gamma \left\{\|\varphi\|^2 + \|\psi\|^2 + \|\varphi + \psi\|^2\right\} + C'\gamma\langle \varphi + \psi, x_n - x_\star\rangle.\end{aligned}$$

Taking the conditional expectation, the last inner product vanishes, and we get the result. □

Lemma 3.7. If Eq (17) hold, then the set of measures $\mathcal{I}(\mathcal{P})$ is tight.

Proof. The following inequalities hold $F^{\gamma_0}(x) + G^{\gamma_0}(x) \le F^{\gamma}(x) + G^{\gamma}(x) \le F^{\gamma'}(x) + G^{\gamma'}(x) \le F(x) + G(x)$ for all $0 \le \gamma' \le \gamma \le \gamma_0$. Moreover H3 $\iff F^{\gamma_0} + G^{\gamma_0}$coercive $\iff F^{\gamma_0} + G^{\gamma_0}$ coercive (see [?]). Hence, condition (PH) in [13] holds. □

Lemma 3.8. Let Assumptions H7, H6, and H3 hold true. Then, for all $\varepsilon > 0$, there exists $M > 0$ such that

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

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 H6 and the nonexpansiveness of the resolvent that

$$\begin{aligned}\bar{\mathbb{E}}_n^{a,\gamma}\|x_{n+1} - \Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n)\| &\le \bar{\mathbb{E}}_n^{a,\gamma}\|\operatorname{prox}_{\gamma g(\cdot,\xi_{n+1})}(x_n) - \Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n)\| + \gamma\bar{\mathbb{E}}_n^{a,\gamma}\|\nabla f_\gamma(x_n, \xi_{n+1})\| \\ &\le C\gamma(1 + F^\gamma(x_n) + G^\gamma(x_n))\,,\end{aligned}$$

On the other hand, since

$$\|\Pi_{\mathrm{cl}(D(\xi_{n+1}))}(x_n) - \Pi_{\mathrm{cl}(\mathcal{D})}(x_n)\|^2 \le \boldsymbol{d}(x_n)^2 - d(x_n, D(\xi_{n+1}))^2 \quad \text{(see (8))},$$

we can make use of Assumption H7 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 $\bar{\mathbb{E}}_n^{a,\gamma}\boldsymbol{d}(x_{n+1}) \le \rho\boldsymbol{d}(x_n) + C\gamma(1 + F^\gamma(x_n) + G^\gamma(x_n))$. By iterating, we end up with the inequality

$$\bar{\mathbb{E}}^{a,\gamma}(\boldsymbol{d}(x_{n+1})) \le \rho^{n+1}\boldsymbol{d}(a) + C\gamma\sum_{k=0}^{n}\rho^{n-k}(1 + \bar{\mathbb{E}}^{a,\gamma}(F^\gamma(x_k) + G^\gamma(x_k))). \tag{25}$$

By taking the expectation and iterating in Th. 3.6,

$$\alpha\sum_{k=0}^{n}\bar{\mathbb{E}}^{a,\gamma}(F^\gamma(x_k) + G^\gamma(x_k)) \le \frac{1}{\gamma}\|a - x_\star\|^2 + (n+1)C$$

for all $n \ge 0$. Thus, by induction, for all $k \ge 0$, $\mathbb{E}^{a,\gamma}(F^\gamma(x_k) + G^\gamma(x_k)) < \infty$. By Markov's inequality,

vérifier ce point. On a vraiment besoin de vérifier les integrabilités?

$$\bar{\mathbb{P}}^{a,\gamma}(\boldsymbol{d}(x_k) \ge M\gamma) \le \frac{\bar{\mathbb{E}}^{a,\gamma}(\boldsymbol{d}(x_k))}{M\gamma}$$

for all $k \in \mathbb{N}$. Let $\pi_\gamma \in \mathcal{I}(P_\gamma)$. From Th. 3.6, and Lem. 3.7 we have

$$\mathbb{E}_n^{\gamma,a}\|x_{n+1} - x_\star\|^2 \le \|x_n - x_\star\|^2 - \alpha\gamma(F^{\gamma_0}(x_n) + G^{\gamma_0}(x_n)) + \gamma C. \tag{26}$$

for all $\gamma \in (0, \gamma_0]$. Using [13] it implies

$$\sup_{\gamma\in(0,\gamma_0]} \sup_{\pi\in\mathcal{I}(P_\gamma)} \pi(F^{\gamma_0} + G^{\gamma_0}) < \infty.$$

In particular, noting that $\boldsymbol{d}(x) \leq \|x\| + \|\Pi_{\mathrm{cl}(\mathcal{D})}(0)\|$, we obtain that $\sup_{\gamma\in(0,\gamma_0]} \sup_{\pi\in\mathcal{I}(P_\gamma)} \pi(\boldsymbol{d}) < \infty$. Let $\gamma \in (0, \gamma_0]$ and $\pi \in \mathcal{I}(P_\gamma)$. Getting back to (25), we have for all $n \in \mathbb{N}$,

$$
\begin{aligned}
\pi(\{x : \boldsymbol{d}(x) \geq M\gamma\}) &= \bar{\mathbb{P}}^{\pi,\gamma}(\boldsymbol{d}(x_{n+1}) \geq M\gamma) \\
&\leq \frac{\mathbb{E}^{\pi,\gamma}(\boldsymbol{d}(x_{n+1}))}{M\gamma} \\
&\leq \rho^{n+1}\frac{\pi(\boldsymbol{d})}{M\gamma} + \frac{C}{M}\sum_{k=0}^{n} \rho^{n-k}(1 + \mathbb{E}^{\pi,\gamma}(F^{\gamma_0}(x_k) + G^{\gamma_0}(x_k))) \\
&= \rho^{n+1}\frac{\pi(\boldsymbol{d})}{M\gamma} + \frac{C}{M}\sum_{k=0}^{n} \rho^{n-k}(1 + \pi(F^{\gamma_0} + G^{\gamma_0})) \\
&\leq \rho^{n+1}\frac{C}{M\gamma} + \frac{C}{M}\,.
\end{aligned}
$$

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

Lemma 3.9. Let the assumptions of the statement of Th. 3.1 hold true. Assume that for all $\varepsilon > 0$, there exists $M > 0$ such that

$$
\sup_{\gamma\in(0,\gamma_0]} \sup_{\pi\in\mathcal{I}(P_\gamma)} \pi((\mathcal{D}_{M\gamma})^c) \leq \varepsilon. \tag{27}
$$

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

Proof. See [?] □

3.3 Proof of Th. 2.1

Assume H1-H6. By 3.7, $\bigcup_{\gamma\in(0,\gamma_0]} \mathcal{I}(P_\gamma)$ is tight and by Lem. 3.8 and Lem. 3.9 any cluster point of $\mathcal{I}(\mathcal{P})$ is an element of $\mathcal{I}(\Phi)$ as $\gamma \to 0$. The rest of the proof follows word-for-word from [13].

préciser que π intègre ψ

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. Asymptotics of constant step stochastic approximations involving differential inclusions. *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] J. de Vries. *Elements of topological dynamics*, volume 257 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1993.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

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

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

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

[29] 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.

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

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

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

[33] 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.

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

[35] 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.

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

[37] 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.