A closed-measure approach to stochastic approximation

P. Bianchi[a] and R. Rios-Zertuche[b]

[a]LTCI, Telecom Paris, Institut Polytechnique de Paris
[b]UiT The Arctic University of Norway

ARTICLE HISTORY
Compiled December 2, 2023

Abstract
This paper introduces a new method to tackle the issue of the almost sure convergence of stochastic approximation algorithms defined from a differential inclusion. Under the assumption of slowly decaying step-sizes, we establish that the set of essential accumulation points of the iterates belongs to the Birkhoff center associated with the differential inclusion. Unlike previous works, our results do not rely on the notion of asymptotic pseudotrajectories introduced by Benaïm–Hofbauer–Sorin, which is the predominant technique to address the convergence problem. They follow as a consequence of Young's superposition principle for closed measures. This perspective bridges the gap between Young's principle and the notion of invariant measure of set-valued dynamical systems introduced by Faure and Roth. Also, the proposed method allows to obtain sufficient conditions under which the velocities locally compensate around any essential accumulation point.

KEYWORDS
Stochastic approximation: Closed measures; Weak convergence; Differential inclusions

1. Introduction

Let $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be an upper semi-continuous set-valued map, and let $\pi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be the projection $\pi(x,v) = x$.

In this paper we are concerned with stochastic processes of the following form:

$$x_{i+1} = x_i + \epsilon_i \theta_i + \epsilon_i \eta_{i+1},$$

where θ_i is in a δ_i-neighborhood of the set $H(x_i)$, both sequences $(\delta_i) \subset [0, +\infty)$ and $(\eta_i) \subset \mathbb{R}^n$ are random, $\delta_i \to 0$ as $i \to +\infty$, and η_{i+1} is a martingale increment. We assume that the step size sequence (ϵ_i) converges to 0; contrary to [5, 12] but similarly to [8, 17], we allow this convergence to happen arbitrarily slowly, hence covering a broad set of practical situations. The sequences (x_i) model a discrete process with drift θ_i, noise η_i, and step-size ϵ_i; we will describe some of the situations where these arise in Section 5 below. An important example is the classical Robbins-Monro algorithm [26], in which $\theta_i = h(x_i)$ for some vector field h; when h is Lipschitz continuous, the so-called ordinary differential equation method can be used to characterize the set of accumulation points of (x_i) [4, 12, 23, 24], and the analysis of more general versions of the algorithm require more sophisticated techniques [5].

Email: bianchi@telecom-paris.fr

We investigate the dynamics of the sequences (x_i) using the method of "closed measures," introduced in [10] in a more restrictive context. Let us briefly describe the approach. We consider the probability measures

$$\mu_i = \frac{\sum_{j=0}^{i} \epsilon_j \delta_{(x_j, \theta_j + \eta_{j+1})}}{\sum_{j=0}^{i} \epsilon_j}$$

on $\mathbb{R}^n \times \mathbb{R}^n$. These are a sort of "occupation measures" of the sequence (μ_i) in the sense that they encode the "position" x_j and "velocity" $\theta_j + \eta_{j+1} = (x_{i+1} - x_i)/\epsilon_i$ of the sequence (x_i). We then consider the set $\mathcal{A}$ of weak* accumulation points of the sequence (μ_i). The elements $\mu \in \mathcal{A}$ are probability measures that encode the long-term, recurring dynamics of (x_i). With mild assumptions 4.1, 4.4, and 4.5, $\mathcal{A}$ is non-empty and the measures μ it contains have the interesting property of being closed, that is, the integrals of gradients vanish: for all $\phi \in C^\infty(\mathbb{R}^n)$, we have

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \langle \nabla \phi(x), v \rangle \, d\mu(x, v) = 0.$$

The Young superposition principle, Th. 3.2, is available for measures with this closedness property; its statement implies that the measures $\mu \in \mathcal{A}$ can be understood as being composed of occupation measures of solutions of the differential inclusion H in a sense that is made precise in Definition 3.5. In particular, this means that the measures μ are invariant for H, and that the projections $\pi(\operatorname{supp} \mu)$ of their supports are contained in the Birkhoff center of H, defined as the closure of all the recurrent points of the solutions of the differential inclusion $x'(t) \in H(x(t))$; in this way, we recover results of [8, 16, 17], formerly obtained using the theory of asymptotic pseudotrajectories [6, 7].

The projections $\pi(\operatorname{supp} \mu)$, $\mu \in \mathcal{A}$, are also shown to compose the essential accumulation set $\operatorname{ess\,acc}(x_i)$ of the sequence (x_i); this is a subset of the accumulation set in which we know that (x_i) spends substantial time. We thus clarify the link between $\operatorname{ess\,acc}(x_i)$ and the Birkhoff center of H.

We also focus on the stable zeros $\operatorname{zer}_s(H)$ of H, which are defined to be the points $x \in \mathbb{R}^n$ such that $0 \in H(x)$ and the only solution of the differential inclusion $x'(t) \in H(x(t))$ is $x(t) = x$. We are able to show, roughly speaking, that the average of the "velocities" $v_i := \theta_i + \eta_{i+1}$ of (x_i) vanishes on $\operatorname{zer}_s(H)$. Moreover, this vanishing implies the oscillation compensation property: if $\operatorname{zer}_s(H) = \operatorname{BC}(H)$, then for all bounded, continuous $\psi \colon \mathbb{R}^n \to (0, +\infty)$ we have (for some subsequence i_k)

$$\lim_{k \to +\infty} \frac{\sum_{j \leqslant i_k} \epsilon_j \psi(x_j) v_j}{\sum_{j \leqslant i_k} \epsilon_j \psi(x_j)} = 0,$$

as long as $\liminf_k \sum_{j \leqslant i_k} \epsilon_j \psi(x_j) / \sum_{j \leqslant i_k} \epsilon_j > 0$. Using ψ to approximate the indicator function of a ball B, the compensation of the oscillations can be seen to be roughly equivalent (up to some technical assumptions) to saying that, on any ball B centered at any point $x \in \operatorname{ess\,acc}(x_i)$, the weighted average of the velocities v_i,

$$\frac{\sum_{\substack{j \leqslant i \\ x_j \in B}} \epsilon_j v_j}{\sum_{\substack{j \leqslant i \\ x_j \in B}} \epsilon_j} = \frac{\sum_{\substack{j \leqslant i \\ x_j \in B}} \epsilon_j (\theta_j + \eta_{j+1})}{\sum_{\substack{j \leqslant i \\ x_j \in B}} \epsilon_j} = \frac{1}{\sum_{\substack{j \leqslant i \\ x_j \in B}} \epsilon_j} \sum_k (x_k^+ - x_k^-),$$

vanishes asymptotically; here we have denoted x_k^+ and x_k^- the points at which the sequence (x_i) enters and exits the ball B for the k-th time, respectively. Observe that the asymptotic vanishing of this quantity implies a "slow down" of the sequence (x_i) (i.e., if $x_k^+ = x_{i+N}$ and $x_k^- = x_i$, the distance traversed $\|x_k^+ - x_k^-\| = \|x_{i+N} - x_i\|$ grows smaller with respect to the time $\sum_{j=i}^{i+N} \epsilon_j$ this takes), at least when we know that $\|x_k^+ - x_k^-\|$ is uniformly bounded away from 0. This generalizes some of the results of [10].

The framework considered in this paper fits, in particular, the case in which H is the Clarke subdifferential of a Lyapunov function V; see Corollary 4.15. Other interesting applications are presented in Section 5, and include the stochastic descent and heavyball algorithms, and the best-response dynamics in games.

The plan of the paper is the following. Section 2 presents known facts about differential inclusions and subdifferentials. Section 3 introduces Young's superposition principle for closed measures, along with useful consequences. Section 4 gives the main results of the paper and is devoted to the development of the analysis of the long-term dynamics of the iterates and their convergence. Section 5 gives some applications to optimization and game theory.

2. Preliminaries

2.1. Notations

If (E, d) is a Polish space equipped with its Borel σ-field, we denote by $\mathcal{P}(E)$ the set of probability measures on E. We denote by $\mathrm{supp}(\nu)$ the support of a measure ν. We denote by $\mathcal{C}(I, E)$ the set of continuous functions on $I \to E$, where I is a real interval. The set $\mathcal{C}(I, \mathbb{R}^n)$ (where n is an integer) will always be equipped with the topology of uniform convergence on compact intervals of I. We denote by $\mathcal{C}_b(\mathbb{R}^n, \mathbb{R})$ the set of bounded continuous functions on $\mathbb{R}^n \to \mathbb{R}$, by $\mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ the set of infinitely differentiable functions on $\mathbb{R}^n \to \mathbb{R}$. The notation 1_A represents the indicator function of a set A, equal to one on that set, zero otherwise. The closure of a set A is denoted by $\mathrm{cl}(A)$. For every set $A \subset \mathbb{R}^n$, define $A^\perp = \{u : \langle u, a\rangle = 0, \forall a \in A\}$. We denote by $d(x, A) = \inf\{d(x, y) : y \in A\}$ the distance between a point $x \in E$ and a non-empty subset $A \subset E$. We say that X is a *set-valued map* on E into E', denoted by $X : E \rightrightarrows E'$, if $X(x)$ is a subset of E' for every $x \in E$. A point $x \in E$ is called a *zero* of X if $0 \in X(x)$. We denote by $\mathrm{zer}(X)$ the set of zeroes of X. We denote by $\mathrm{graph}\,(X) \coloneqq \{(x, y) : y \in X(x)\}$ the graph of X. We say that X is *upper semicontinuous* (u.s.c.) if for every $x \in E$ and every neighborhood V of $X(x)$, there exists a neighborhood U of x such that $X(U) \subset V$. The projection onto position space will be denoted $\pi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $\pi(x, v) = x$.

2.2. Differential inclusions

Let $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a u.s.c. set-valued map with non-empty compact convex values. We say that an absolutely continuous mapping $\gamma\colon \mathbb{R}_+ \to \mathbb{R}^n$ is a solution to the differential inclusion

$$\gamma'(t) \in H(\gamma(t)), \tag{DI}$$

with initial condition $x \in \mathbb{R}^n$, if $\gamma(0) = x$ and if (DI) holds for almost every $t \in \mathbb{R}_+$. We denote by $S_H(x)$ the subset of $\mathcal{C}(\mathbb{R}^+, \mathbb{R}^n)$ of all solutions to (DI) issued from x. In the particular case, where H satisfies the following linear growth condition:

$$\exists C > 0,\ \forall x \in \mathbb{R}^n,\ \|H(x)\| \leqslant C(1 + \|x\|)\ , \tag{1}$$

the set $S_H(x)$ is non-empty for every $x \in \mathbb{R}^n$ [2].

The limit set of a function $\gamma \in C(\mathbb{R}_+, \mathbb{R}^n)$ is defined as

$$L(\gamma) \coloneqq \bigcap_{t \geqslant 0} \mathrm{cl}\,(\gamma([t, +\infty))\,. \tag{2}$$

It coincides with the set of points of the form $\lim_n \gamma(t_n)$ for some sequence $t_n \to \infty$.

Definition 2.1. A point $x \in \mathbb{R}^n$ is said *recurrent* w.r.t (DI) if there exists $\gamma \in S_H(x)$ such that $x \in L(\gamma)$. The *Birkhoff center*, denoted by $\mathrm{BC}(H)$, is the closure of the set of all recurrent points:

$$\mathrm{BC}(H) \coloneqq \mathrm{cl}\,\{x \in \mathbb{R}^n\ :\ \exists \gamma \in S_H(x),\ x \in L(\gamma)\}\ .$$

Definition 2.2. Given $\Lambda \subset \mathbb{R}^n$, we say that a continuous function $V\colon \mathbb{R}^n \to \mathbb{R}$ is a *Lyapunov function* for Λ if for all $\gamma\colon \mathbb{R}_+ \to \mathbb{R}^n$ solution to (DI), we have for all $t > 0$:

$$\begin{cases} V(\gamma(t)) \leqslant V(\gamma(0)) \\ V(\gamma(t)) < V(\gamma(0)), \quad \text{if } \gamma(0) \notin \Lambda\,. \end{cases}$$

Proposition 2.3. *Let $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a u.s.c. set-valued map with non-empty compact convex values. Consider $\Lambda \subset \mathbb{R}^n$ and let $V\colon \mathbb{R}^n \to \mathbb{R}$ be a Lyapunov function for Λ. Then,* $\mathrm{BC}(H) \subset \Lambda$.

Proof. Consider $x \in \mathrm{BC}(H)$ and $\gamma \in S_H(x)$ such that $x \in L(\gamma)$. Assume by contradiction that $x \notin \Lambda$. Then, $V(\gamma(1)) < V(x)$. Set $\varepsilon \coloneqq V(x) - V(\gamma(1))$. As V is continuous, there exists $\delta > 0$ such that for every y, $\|y - x\| < \delta$ implies $V(y) \geqslant V(x) - \varepsilon/2$. As $x \in L(\gamma)$, there exists $t > 1$ such that $\|\gamma(t) - x\| < \delta$. Thus, $V(\gamma(t)) \geqslant V(x) - \varepsilon/2$. As $V(\gamma(t)) \leqslant V(\gamma(1))$, we obtain a contradiction. □

2.3. Clarke subdifferential

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a locally-Lipschitz continuous function. Denote by $\mathrm{Reg}(f)$ the set of its differentiability points. Recall that $\mathrm{Reg}(f)$ is dense in $\mathbb{R}^n$ by Rademacher's theorem.

Definition 2.4 (Clarke subdifferential)**.** The Clarke subdifferential of f at a point $x \in \mathbb{R}^n$ is the set

$$\partial f(x) \coloneqq \overline{\mathrm{conv}}\{v \in \mathbb{R}^n : \exists (y_i)_i \in \mathrm{Reg}(f)^{\mathbb{N}},\ y_i \to x \text{ and } \nabla f(y_i) \to v\}\,.$$

The corresponding set-valued map ∂f is u.s.c. with non-empty compact convex values.

Definition 2.5 (Path differentiability)**.** A locally-Lipschitz function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said *path-differentiable* if, for every locally-Lipschitz curve $\gamma\colon \mathbb{R} \to \mathbb{R}^n$, the composition $f \circ \gamma$ is differentiable at almost every point t, and satisfies:

$$(f \circ \gamma)'(t) = \langle v, \gamma'(t) \rangle$$

for every $v \in \partial f(\gamma(t))$.

3. Closed measures

3.1. The superposition principle

Let n be an integer.

Consider a probability measure $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$. Let $\pi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be the projection, $\pi(x, v) \coloneqq x$. The disintegration theorem states that there exists a probability transition kernel $\mathbb{R}^n \times \mathscr{B}(\mathbb{R}^n) \to [0, 1]$, denoted by $(x, A) \mapsto \mu_x(A)$ such that for every bounded continuous function $\varphi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$\mu(\varphi) = \int \left(\int \varphi(x, v) d\mu_x(v) \right) d(\pi_* \mu)(x)\,, \tag{3}$$

where $\pi_*\mu$ is the pushforward measure of μ through π *i.e.*, $\pi_*\mu(A) = \mu(A \times \mathbb{R}^n)$ for every $A \in \mathscr{B}(\mathbb{R}^n)$. Note that:

$$\operatorname{supp}(\pi_* \mu) = \pi(\operatorname{supp} \mu)\,.$$

We shall say that μ has the *disintegration* $\mu = \int \mu_x d(\pi_*\mu)(x)$ if Eq. (3) holds for every bounded continuous function φ. Whenever it is well defined, we will refer to the quantity

$$v_\mu(x) \coloneqq \int v d\mu_x(v)\,,$$

as the *centroid field* associated with μ.

Definition 3.1. A compactly supported probability measure $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ is said *closed*, if for every $g \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$,

$$\int \langle \nabla g(x), v \rangle d\mu(x, v) = 0\,.$$

We define the shift operator $\Theta\colon \mathcal{C}(\mathbb{R}, \mathbb{R}^n) \to \mathcal{C}(\mathbb{R}, \mathcal{C}(\mathbb{R}, \mathbb{R}^n))$ s.t. for every $\gamma \in \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$, $\Theta(\gamma)\colon t \mapsto \gamma(t + \cdot)$. A measure $\vartheta \in \mathcal{P}(\mathcal{C}(\mathbb{R}, \mathbb{R}^n))$ is said Θ-invariant if $\vartheta = \vartheta \Theta_t^{-1}$ for all t, where $\Theta_t(\gamma) \coloneqq \Theta(\gamma)(t)$.

Theorem 3.2 (Young's superposition principle)**.** *Let $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ be a closed probability measure. There exists $\vartheta \in \mathcal{P}(\mathcal{C}(\mathbb{R}, \mathbb{R}^n))$ such that:*

i) ϑ is Θ-invariant.
ii) ϑ-almost every curve γ is Lipschitz continuous and differentiable at 0.

iii) For every bounded measurable function $\varphi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$\int \varphi(x, v_\mu(x)) d(\pi_* \mu)(x) = \int \varphi(\gamma(0), \gamma'(0)) d\vartheta(\gamma)\,. \tag{4}$$

In addition, ϑ*-almost every* γ *satisfies* $\gamma'(t) = v_\mu(\gamma(t))$ *for almost every* $t \in \mathbb{R}$.

Proof. The statement can be established as a consequence of [9, Th. 6.2]. In order to make the paper self-contained, we provide a direct and constructive proof, which is inspired of the proof of [1, Th. 8.2.1].

Smooth case. Consider a closed measure $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that $v_\mu(x)$ is a smooth function. Consider a compact set $U \subset \mathbb{R}^n$ containing the supports of ρ and v_μ. Consider the flow $\Phi\colon U \times \mathbb{R} \to U$ associated to the ODE $x'(t) = v_\mu(x(t))$, so that

$$\Phi(x, 0) = x \qquad \text{and} \qquad \frac{\partial \Phi}{\partial t}(x, t) = v_\mu(\Phi(x, t)),$$

and denote $\Phi_t(x) = \Phi(x, t)$. Consider any smooth function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$. Define $g(t, x) = \varphi(\Phi_t(x))$. Obviously, $s \mapsto g(t + s, \Phi_{-s}(x))$ is constant. Computing the derivative of this function at the point $s = 0$, we obtain the transport equation:

$$0 = \partial_t g(t, x) - \langle \nabla_x g(t, x), v_\mu(x) \rangle\,.$$

Integrating w.r.t. $\pi_* \mu$,

$$\frac{d}{dt} \int g(t, x) d\pi_* \mu(x) = \int \langle \nabla_x g(t, x), v_\mu(x) \rangle d\pi_* \mu(x) = \int \langle \nabla_x g(t, x), v \rangle d\mu(x, v) = 0\,,$$

as μ is closed. Therefore, $\int \varphi(\Phi_t(x)) d\pi_* \mu(x)$ is constant w.r.t. t. In other words, $\pi_* \mu$ is an invariant measure of Φ.

We define ϑ as the probability measure on $\mathcal{C}(\mathbb{R}, \mathbb{R}^n)$, given for every $g\colon \mathcal{C}(\mathbb{R}, \mathbb{R}^n) \to \mathbb{R}_+$ by:

$$\vartheta(g) \coloneqq \int g(\Phi(x, \cdot)) d\pi_* \mu(x)\,.$$

As Φ preserves $\pi_* \mu$, it follows that ϑ is Θ-invariant. Let $\phi(x, v)$ be a bounded continuous function, and set $g(\gamma) = \phi(\gamma(0), \gamma'(0))$. Note that g is well-defined ϑ-a.e., because ϑ-almost every γ is differentiable at 0. Moreover, $\gamma'(0) = v_\mu(\gamma(0))$ almost everywhere. By definition of ϑ, Eq. (4) follows.

General case. We follow the same steps as in the proof of [1, Th. 8.2.1]. Consider a closed measure μ. Consider a smooth and compactly supported $\psi : \mathbb{R}^n \to \mathbb{R}_+$ such that $\int \psi(x) dx = 1$. For every $\varepsilon > 0$, set $\psi_\varepsilon(x) = \psi(x/\varepsilon)/\varepsilon^n$. Define for every $x \in \mathbb{R}^n$,

$$\rho_\varepsilon(x) \coloneqq \int \psi_\varepsilon(x - y) \pi_* \mu(dy) \quad \text{and} \quad v_\epsilon(x) \coloneqq \begin{cases} \rho_\varepsilon(x)^{-1} \int \psi_\varepsilon(x - y) v d\mu(y, v), & \rho_\varepsilon(x) \neq 0, \\ 0, & \rho_\varepsilon(x) = 0. \end{cases}$$

It is an easy exercise to verify that the measure defined by:

$$\mu_\varepsilon(dx, dv) := \rho_\varepsilon(x) dx \otimes \delta_{v_\varepsilon(x)}(dv)$$

is closed, and satisfies the smoothness assumption of the previous paragraph. Thus, there exists a Θ-invariant measure ϑ_ε on $C(\mathbb{R},\mathbb{R}^n)$ satisfying Eq. (4). In particular, setting $\mathrm{ev}_t(\gamma) := \gamma(t)$ for every $\gamma \in C(\mathbb{R},\mathbb{R}^n)$, it holds that, for all t,

$$(\mathrm{ev}_t)_*\vartheta_\varepsilon = \rho_\varepsilon(x)dx\,. \tag{5}$$

We prove that the family $(\vartheta_\varepsilon : 0<\varepsilon<1)$ is tight in $\mathcal{P}(C(\mathbb{R},\mathbb{R}^n))$. As $\pi_*\mu$ and ψ are compactly supported, there exists a compact set $K \subset \mathbb{R}^n$ such that, for all $\varepsilon \in (0,1)$ and for ϑ_ε-almost every γ, $\gamma(0) \in K$. Moreover, for every $T>0$,

$$\begin{aligned}\int \left(\int_0^T \|\gamma'(t)\|^2 dt\right) d\vartheta_\varepsilon(\gamma) &= T\int \|v_\varepsilon(\gamma(0))\|^2 d\vartheta_\varepsilon(\gamma)\\ &= T\int \|v_\varepsilon(x)\|^2\rho_\varepsilon(x)dx\\ &\leqslant T\int \|v_\mu(x)\|^2\pi_*\mu(dx)\,,\end{aligned} \tag{6}$$

where the last inequality follows from [1, Lemma 8.1.10]. For every $c>0$, the set $\mathcal{K}$ of absolutely continuous curves γ such that $\gamma(0)\in K$ and $\int_0^T\|\gamma'(t)\|^2dt < c$ is relatively compact in $C([0,T],\mathbb{R}^n)$. The tightness of $(\vartheta_\varepsilon)_\varepsilon$ follows from (6) and Markov's inequality.

As a consequence, there exists a measure ϑ such that $\vartheta_{\varepsilon_i}$ converges weak* to ϑ, along some subsequence (ε_i). The measure $\vartheta_{\varepsilon_i}$ being Θ-invariant, the same holds for ϑ, and the first point is proved. Also, letting ε_i tend to zero in Eq. (5), we obtain that $(\mathrm{ev}_t)_*\vartheta = \pi_*\mu$ for every t.

Consider a fixed $T>0$. For any smooth and compactly supported vector field $w \in C_c^\infty(\mathbb{R}^n,\mathbb{R}^n)$, consider the map $F_w : C(\mathbb{R},\mathbb{R}^n)\to\mathbb{R}_+$ given by:

$$F_w(\gamma) \coloneqq \int_0^T \left\|\gamma(t)-\gamma(0)-\int_0^t w(\gamma(s))ds\right\| dt\,. \tag{7}$$

Observe that $F_{v_\varepsilon}(\gamma)=0$ for ϑ_ε-a.e. γ.

Let $\delta>0$. From the Stone-Weierstrass theorem and the existence of C^∞ bump functions, it follows that compactly-supported, C^∞ functions are dense in $L^1(\pi_*\mu)$. In particular, there exists $w\in C_c^\infty(\mathbb{R}^n,\mathbb{R}^n)$ such that $\int\|w(x)-v_\mu(x)\|\pi_*\mu(dx)<\delta$. Define:

$$w_\varepsilon(x) := \rho_\varepsilon(x)^{-1}\int \psi_\varepsilon(x-y)w(y)\pi_*\mu(dy)\,.$$

Using that F_{v_ε} is equal to zero ϑ_ε-a.e., and using the fact that $(\mathrm{ev}_s)_*\vartheta_\varepsilon = \rho_\varepsilon(x)dx$, we obtain:

$$\begin{aligned}&\int F_w(\gamma)\,d\vartheta_\varepsilon(\gamma) \leqslant \int F_{w_\varepsilon}(\gamma)\,d\vartheta_\varepsilon(\gamma) + \int\int_0^T\int_0^t \|w_\varepsilon(\gamma(s))-w(\gamma(s))\|ds dt\,d\vartheta_\varepsilon(\gamma)\\ &\leqslant T^2\int\|w_\varepsilon(x)-v_\varepsilon(x)\|\rho_\varepsilon(x)dx + T^2\int\|w_\varepsilon(x)-w(x)\|\rho_\varepsilon(x)dx\,.\end{aligned} \tag{8}$$

The integral in the first term in the right-hand side of Eq. (8) satisfies:

$$\begin{aligned}\int \|w_\varepsilon(x) - v_\varepsilon(x)\|\rho_\varepsilon(x)dx &\leqslant \int\int \psi_\varepsilon(x-y)\|w(y) - v_\mu(y)\|\pi_*\mu(dy)dx \\ &= \int \|w(y) - v_\mu(y)\|\pi_*\mu(dy) \leqslant \delta\,.\end{aligned}$$

The second term in Eq. (8) can be handled by noting that $\|w(x) - w(y)\| \leqslant C\|x - y\|$ for some Lipschitz constant C (which may depend on δ). By straightforward algebra,

$$\int \|w_\varepsilon(x) - w(x)\|\rho_\varepsilon(x)dx \leqslant C\varepsilon \int \|x\|\psi(x)dx\,.$$

Finally, letting ε tend to zero in Eq. (8) along the sequence (ε_i), and by using the continuity of the map F_w, we obtain that $\int F_w d\vartheta \leqslant T^2\delta$. Define F_{v_μ} as in Eq. (7), with v_μ in place of w. The function F_{v_μ} is well defined ϑ-almost everywhere, and satisfies:

$$\begin{aligned}\int F_{v_\mu}(\gamma)d\vartheta(\gamma) &\leqslant T^2\delta + \int\int_0^T\int_0^t \|w(\gamma(s)) - v_\mu(\gamma(s))\|dsdtd\vartheta(\gamma) \\ &\leqslant T^2\delta + T^2\int \|w(x) - v_\mu(x)\|\pi_*\mu(x)\,,\end{aligned}$$

where the last inequality is due to the fact that $(\mathrm{ev}_s)_*\vartheta = \pi_*\mu$ for all s. Therefore, $\int F_{v_\mu} d\vartheta$ is bounded by $2T^2\delta$. As δ was arbitrary, the former is equal to zero. We conclude that for ϑ-almost every γ,

$$\forall t \in \mathbb{R},\ \gamma(t) = \gamma_0 + \int_0^t v_\mu(\gamma(s))ds\,.$$

This proves the second point and the last point of the theorem. Moreover, note that for ϑ-almost every γ, $\gamma'(t) = v_\mu(\gamma(t))$ for almost every t. Hence, the third point follows by the Θ-invariance of ϑ. □

3.2. Consequences

We review some consequences of Young's superposition principle.

Proposition 3.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ a path-differentiable function and $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ be a closed measure. Then, for every measurable map $\xi\colon \mathbb{R}^n \to \mathbb{R}^n$ such that $\xi(x) \in \partial f(x)$ for $\pi_*\mu$-almost every x, $\int\langle \xi(x), v\rangle d\mu(x,v) = 0$.*

Proof. By Prop. 3.2, there exists a measure ϑ on $\mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ such that:

$$\int\langle \xi(x), v\rangle d\mu(x,v) = \int\langle \xi(x), v_\mu(x)\rangle\pi_*\mu(dx) = \int\langle \xi(\gamma(0)), \gamma'(0)\rangle\vartheta(d\gamma)\,.$$

As ϑ is shift-invariant, the above value coincides with $\int\langle \xi(\gamma(t)), \gamma'(t)\rangle\vartheta(d\gamma)$, for every

t. Thus, for every $T > 0$

$$\int \langle \xi(x), v\rangle d\mu(x,v) = \frac{1}{T}\int_0^T \int \langle \xi(\gamma(t)), \gamma'(t)\rangle \vartheta(d\gamma)dt\,.$$

By Fubini's theorem, the right-hand side is equal to $\frac{1}{T}\int (f(\gamma(T)) - f(\gamma(0)))\vartheta(d\gamma)$, which is equal to zero, using the fact that ϑ is Θ-invariant.

□

Theorem 3.4. *Consider an u.s.c. map $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with non-empty compact convex values. Let $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ be a closed probability measure. Assume that for $\pi_*\mu$-almost every x,*

$$v_\mu(x) \in H(x)\,. \tag{9}$$

Then, $\pi(\operatorname{supp}\mu) \subset \mathrm{BC}(H)$. Moreover, $v_\mu = 0$, $\pi_*\mu$-a.e. on the set $\mathrm{zer}_s(H) \coloneqq \{x \in \mathrm{zer}(H) : S_H(x) = \{t \mapsto x\}\}$ i.e., the set of zeroes x of H such that any solution to (DI) issued from x is constant.

Proof. Define the measurable map $L\colon \mathcal{C}(\mathbb{R}, \mathbb{R}^n) \rightrightarrows \mathbb{R}^n$ given by Eq. (2). Note that $L = L \circ \Theta$. As ϑ is Θ-invariant, $(\mathrm{ev}_0, L)_*\vartheta = (\mathrm{ev}_t, L)_*\vartheta$ for every t, where ev_t stands for the projection $\mathrm{ev}_t(\gamma) = \gamma(t)$. Moreover, for every $\gamma \in \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$, $d(\gamma(t), L(\gamma)) \to 0$. By the dominated convergence theorem,

$$\int 1 \wedge d(\gamma(0), L(\gamma))d\vartheta(\gamma) = \lim_{t\to\infty} \int 1 \wedge d(\gamma(t), L(\gamma))d\vartheta(\gamma) = 0\,.$$

Therefore, ϑ-almost every γ satisfies $\gamma(0) \in L(\gamma)$. Moreover, γ is a solution to (DI) on $[0, +\infty)$ by the last statement of Th. 3.2. Therefore, $\gamma(0)$ is a recurrent point of (DI). As the distribution of $\gamma(0)$ is equal to $\pi_*\mu$, the first point is proved.

If, moreover, $\gamma(0) \in \mathrm{zer}_s(H)$, then $\gamma \equiv \gamma(0)$. In particular, $\gamma'(0) = 0$. Using the superposition principle,

$$0 = \int \|\gamma'(0)\| 1_{\mathrm{zer}_s(H)}(\gamma(0))d\vartheta(\gamma) = \int \|v_\mu(x)\| 1_{\mathrm{zer}_s(H)}(x) d\pi_*\mu(x)\,.$$

Here, we used the fact that $\mathrm{zer}_s(H)$ is measurable. Indeed, if $S_I\colon \mathbb{R}^n \to \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ is the map such that $S_I(x)$ is the constant function $t \mapsto x$, the set $\mathrm{zer}_s(H)$ coincides with $\{x \in \mathbb{R}^n : (S_H - S_I)(x) = \{t \mapsto 0\}\}$. By [2], $S_H\colon \mathbb{R}^n \rightrightarrows C(\mathbb{R}, \mathbb{R}^n)$ is an upper semicontinuous map with closed graph and is therefore measurable by [14, Th. 18.20]. The same holds for $S_H - S_I$. Thus, $\mathrm{zer}_s(H) = (S_H - S_I)^{-1}(t \mapsto 0)$ is also measurable. □

Definition 3.5 (H-invariant measure [17, 25])**.** A probability measure $\nu \in \mathcal{P}(\mathbb{R}^n)$ is *H-invariant (for (DI))* if there exists a Θ-invariant measure $\vartheta \in \mathcal{P}(C(\mathbb{R}, \mathbb{R}^n))$ which is supported by the set of (complete) solutions to (DI), and such that $\nu = \vartheta(\mathrm{ev}_0)^{-1}$, where ev_0 is the projection $\mathrm{ev}_0(\gamma) \coloneqq \gamma(0)$.

Remark 3.6. Consider a closed measure μ satisfying Eq. (9), and denote by ϑ the measure of Th. 3.2. By the last point of Th. 3.2, ϑ is supported by the set of (complete) solutions to (DI). The superposition principle also shows that $\pi_*\mu = \vartheta(\mathrm{ev}_0)^{-1}$. As a consequence, $\pi_*\mu$ is an invariant measure to (DI) in the sense of Faure and Roth [17],

and the two measures ϑ represent the same object in both cases. Compare also with Corollary 4.10.

Theorem 3.7. *Consider an u.s.c. map $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with non-empty compact convex values. Consider the differential inclusion* (DI) *and let $V\colon \mathbb{R}^n \to \mathbb{R}$ be a path-differentiable Lyapunov function for some $\Lambda \subset \mathbb{R}^n$. Let $\mu \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ be a closed measure satisfying Eq.* (9) *$\pi_*\mu$-almost everywhere. Then, $v_\mu(x) \in \partial V(x)^\perp$ for $\pi_*\mu$-almost every x.*

Proof. Let ϑ be the measure given by Th. 3.2. By Prop. 2.3 and Th. 3.4, for ϑ-almost every γ,

$$\gamma(\mathbb{R}) \subset \pi(\mathrm{supp}(\mu)) \subset \mathrm{BC}(H) \subset \Lambda\,.$$

Therefore, $V \circ \gamma$ is constant. As V is path-differentiable, this implies that for almost every t and for every $v \in \partial V(\gamma(t))$, $\langle v, \gamma'(t)\rangle = 0$. In other words, $\gamma'(t) \in \partial V(\gamma(t))^\perp$. As ϑ is Θ-invariant, $\gamma'(0) \in \partial V(\gamma(0))^\perp$, ϑ-a.e. The result follows from Eq. (4) where $\phi(x, v)$ is taken as the indicator function of $\mathrm{graph}(\partial V^\perp)$. □

4. Stochastic approximation

This is the main section of the paper. In Section 4.1, we give the precise framework of the random sequences (x_i) that interest us, we define the measures μ_i. Then, in Section 4.2, we study the accumulation points of the sequence (μ_i) establishing their closedness in Proposition 4.9 and their invariance for $x'(t) = H(x(t))$ in Corollary 4.10, as well as their link to the Birkhoff center. We then proceed, in Section 4.3 to define and study the essential accumulation set and to link it to the Birkhoff center of H. Finally, we collect in Section 4.4 our discussion regarding oscillation compensation properties for general differential inclusions H and for those associated to Lyapunov functions V.

4.1. The Framework

We introduce a set-valued map $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, which satisfies the following assumption.

Assumption 4.1. The map $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is u.s.c. with non-empty compact convex values.

We define:

$$H^\delta(x) \coloneqq \{y \in \mathbb{R}^n : \exists z \text{ s.t. } \|z - x\| \leqslant \delta \text{ and } d(y, H(z)) \leqslant \delta\}\,. \tag{10}$$

Let $(\Omega, \mathcal{F}, (\mathcal{F}_i)_{i\in\mathbb{N}}, \mathbb{P})$ be a filtered probability space. Consider a real sequence $(\epsilon_i) \in (0, +\infty)^{\mathbb{N}}$. Assume that there exists a $(\mathcal{F}_i)$-adapted stochastic process (x_i, η_i, δ_i) on Ω to $(\mathbb{R}^n \times \mathbb{R}^n \times [0, \infty))^{\mathbb{N}}$, equipped with the Borel σ-field associated with the product topology, such that the following inclusion holds almost everywhere:

$$x_{i+1} \in x_i + \epsilon_i H^{\delta_i}(x_i) + \epsilon_i \eta_{i+1}. \tag{11}$$

Remark 4.2. Examples of processes satisfying Eq. (11) are provided in Section 5. In

the general setting, one can construct such a process by setting $\Omega \coloneqq (\mathbb{R}^n \times [0,\infty))^{\mathbb{N}}$, $\mathcal{F}$ equal to the Borel σ-field associated with the product topology, $(\eta_i, \delta_i)_i$ equal to the canonical process on Ω, and $(\mathcal{F}_i)$ equal to the natural filtration. By Assumption 4.1 and the measurable selection theorem [13, Theorem III.9], there exists a Borel map $\varphi : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n$, such that $\varphi(x,\delta) \in H^\delta(x)$ for every (x,δ). Given an arbitrary $x_0 \in \mathbb{R}^n$, the sequence of r.v. (x_i) iteratively defined by $x_{i+1} = x_i + \epsilon_i \varphi(x_i, \delta_i) + \epsilon_i \eta_{i+1}$ is adapted, and satisfies the inclusion (11).

We introduce the following event:

$$\Gamma \coloneqq \{(x_i) \text{ is bounded and } \delta_i \to 0\}\,.$$

Remark 4.3. The set Γ is measurable, because it is the intersection of the measurable set $\{\limsup_i \delta_i = 0\}$ and the countable union of the sets $\{\sup_i \|x_i\| \leqslant q\}$ for $q \in \mathbb{N}$.

We make the following assumptions.

Assumption 4.4. For every $i \in \mathbb{N}$, the r.v. $\|\eta_i\|$ is $\mathbb{P}$- integrable, and satisfies $\mathbb{P}$-a.e.:

$$\mathbb{E}(\eta_{i+1}|\mathcal{F}_i) = 0$$

Moreover, there exists $q > 1$, such that

$$\sup_i \mathbb{E}\left(\|\eta_{i+1}\|^q|\mathcal{F}_i\right) < \infty, \quad \mathbb{P}\text{-a.e. on } \Gamma\,. \tag{12}$$

Assumption 4.5. The step size sequence (ϵ_i) is positive and satisfies:

i) $\sum_i \epsilon_i = +\infty$.
ii) $\epsilon_i \to 0$.

We define the velocity

$$v_{i+1} = \frac{x_{i+1} - x_i}{\epsilon_i}$$

We introduce the random measure μ_i on $\mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^n)$ by

$$\mu_i = \frac{\sum_{j=0}^i \epsilon_j \delta_{(x_j, v_{j+1})}}{\sum_{j=0}^i \epsilon_j}\,. \tag{13}$$

Remark 4.6. Note that μ_i is a random variable on Ω to the space of Borel probability measures on $\mathbb{R}^n \times \mathbb{R}^n$, equipped with Borel σ-field associated with the weak* topology.

4.2. Weak accumulation points*

We start with a preliminary lemma.

Lemma 4.7. *Let Assumptions 4.1, 4.4 and 4.5-i) hold. Define $\bar{v}_i \coloneqq v_i - \eta_i$ and define $\bar{\mu}_i$ in the same way as μ_i, only replacing v_{i+1} by $\bar{v}_{i+1}$ in Eq. (13). The following holds*

$\mathbb{P}$-*a.e. on* Γ*:*

$$\sup_i \int \|v\|^q d\mu_i(x,v) < \infty \tag{14}$$

$$\sup_i \int \|v\|^q d\bar{\mu}_i(x,v) < \infty \tag{15}$$

$$\lim_{i\to\infty} \frac{\sum_{j\leqslant i} \epsilon_j h(x_j)\eta_{j+1}}{\sum_{j\leqslant i}\epsilon_j} = 0\,, \tag{16}$$

for every measurable function $h\colon \mathbb{R}^n \to \mathbb{R}$ *which is bounded on bounded sets. In particular,* (μ_i) *and* $(\bar{\mu}_i)$ *are tight,* $\mathbb{P}$*-a.e. on* Γ*.*

Proof. Define $\bar{v}_i \coloneqq v_i - \eta_i$. By definition, $\bar{v}_{i+1} \in H^{\delta_i}(x_i)$.

We first establish Eq. (16). Denote by M_{i+1} the numerator of the ratio in Eq. (16). Note that (M_i) is a martingale. Set $\alpha_i \coloneqq \mathbb{E}(\|M_{i+1} - M_i\|^q \,|\mathcal{F}_i)$ and $s_i \coloneqq \sum_{j\leqslant i}\alpha_j$. We use Lemma A.1. On the event $\{\sup_i s_i < \infty\}$, M_i converges a.s. to a finite random variable. As $\sum_i \epsilon_i = \infty$, we obtain that $M_{i+1}/\sum_{j\leqslant i}\epsilon_j \to 0$ a.s. on that event. On the event $\{s_i \to \infty\}$, we have that, almost surely,

$$\frac{M_{i+1}}{(s_i \log^2(s_i))^{1/q}} \to 0\,.$$

Note that $\alpha_i = \epsilon_i^q |h(x_i)|^q \mathbb{E}(\|\eta_{i+1}\|^q \,|\mathcal{F}_i)$. Thus, by Assumption 4.4, α_i/ϵ_i^q is uniformly bounded a.e. on Γ. This implies that, for some random variable c, which is finite a.e. on Γ,

$$\frac{M_{i+1}}{\sum_{j\leqslant i}\epsilon_j} \leqslant c \frac{M_{i+1}}{(s_i \log^2(s_i))^{1/q}} U_i^{1/q},,$$

where

$$U_i \coloneqq \frac{\left(\sum_{j\leqslant i}\epsilon_j^q\right)\log^2\left(\sum_{j\leqslant i}\epsilon_j^q\right)}{\left(\sum_{j\leqslant i}\epsilon_j\right)^q}\,.$$

Using that $\sum_{j\leqslant i}\epsilon_j^q \leqslant \sum_{j\leqslant i}\epsilon_j$ for large i, we obtain that U_i tends to zero. Thus, $M_{i+1}/\sum_{j\leqslant i}\epsilon_j \to 0$ a.s. on the event $\{s_i \to \infty\} \cap \Gamma$.

We now establish Eq. (14). We decompose:

$$\frac{\sum_{j\leqslant i}\epsilon_j \|v_{j+1}\|^q}{\sum_{j\leqslant i}\epsilon_j} = \frac{\sum_{j\leqslant i}\epsilon_j \mathbb{E}(\|v_{j+1}\|^q|\mathcal{F}_j)}{\sum_{j\leqslant i}\epsilon_j} + \frac{\sum_{j\leqslant i}\epsilon_j\left(\|v_{j+1}\|^q - \mathbb{E}(\|v_{j+1}\|^q|\mathcal{F}_j)\right)}{\sum_{j\leqslant i}\epsilon_j}\,.$$

The second term in the right-hand side of the above equality tends to zero a.s. on Γ as a consequence of Eq. (16), used with the martingale increment $\|v_{j+1}\|^q - \mathbb{E}(\|v_{j+1}\|^q|\mathcal{F}_j)$ instead of η_{j+1} (the proof is identical). Consider the first term. The sequence $\mathbb{E}(\|v_{j+1}\|^q|\mathcal{F}_j)$ is uniformly bounded a.e. on Γ. Consequently, Eq. (14) is shown. The proof of Eq. (15) follows the same line. □

Recall that $\pi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the projection, $\pi(x,v) \coloneqq x$.

Proposition 4.8. *Let Assumptions 4.1, 4.4 and 4.5-i) hold. Then, $\mathbb{P}$-a.e. $((x_i),(\delta_i))$ on Γ, any weak* accumulation point μ of (μ_i) satisfies $v_\mu(x) \in H(x)$ for $\pi_*\mu$-almost all x.*

Proof. Let μ be an accumulation point of (μ_i), that is, $\mu_i \to \mu$ along some subsequence. Extracting a further subsequence, one can assume as well that $\bar{\mu}_i \to \bar{\mu}$ along the same subsequence, for some other limiting measure $\bar{\mu}$.

For every bounded continuous function h,

$$\int h(x)v \, d\mu_i(x,v) - \int h(x)v \, d\bar{\mu}_i(x,v) = \frac{\sum_{j\leqslant i} \epsilon_j h(x_j)\eta_{j+1}}{\sum_{j\leqslant i}\epsilon_j}\,. \tag{17}$$

The right-hand side tends a.s. to zero on Γ by Lem. 4.7. Eq. (14) and (15) of Lem. 4.7 ensure uniform integrability, and thus imply that the left-hand side of Eq. (17) converges to $\int h(x)vd\mu(x,v) - \int h(x)vd\bar{\mu}(x,v)$. Thus, for every bounded continuous function h,

$$\int h(x)v \, d\mu(x,v) = \int h(x)v \, d\bar{\mu}(x,v)\,.$$

By disintegration, $\int h(x)v_\mu(x)d\pi_*\mu(x) = \int h(x)v_{\bar{\mu}}(x)d\pi_*\bar{\mu}(x)$. As $\pi_*\mu = \pi_*\bar{\mu}$, we conclude that $v_\mu = v_{\bar{\mu}}$ $\pi_*\mu$-a.e.

Let $\delta > 0$ be arbitrary. As $\delta_i \to 0$ a.e. on Γ, $\bar{v}_{i+1} \in H^\delta(x_i)$ for every i large enough. Equivalently, $1_{\operatorname{graph} H^\delta}(x_i, \bar{v}_{i+1}) = 1$ for large i. Thus,

$$\int 1_{\operatorname{graph} H^\delta}(x,v) d\bar{\mu}_i(x,v) \to 1\,.$$

As $\operatorname{graph} H^\delta$ is closed, $\int 1_{\operatorname{graph} H^\delta}(x,v)d\bar{\mu}(x,v) = 1$. Desintegrate $\bar{\mu} = \int \bar{\mu}_x d\pi_*\mu(x)$. By Fubini's theorem, $\int 1_{\operatorname{graph} H^\delta}(x,v)d\bar{\mu}_x(v) = 1$ for $\pi_*\mu$-a.e x. This proves that $\operatorname{supp}(\mu_x) \subset H^\delta(x)$. Using that $H^\delta(x)$ is convex, it follows that $v_{\bar{\mu}}(x) \in H^\delta(x)$ for $\pi_*\mu$-a.e. x. As δ is arbitrary, $v_{\bar{\mu}}(x) \in \bigcap_{\delta>0} H^\delta(x) = H(x)$. As v_μ coincides with $v_{\bar{\mu}}$ π_*-a.e., the conclusion follows. □

Proposition 4.9. *Let Assumptions 4.1, 4.4 and 4.5 hold. Then, $\mathbb{P}$-a.e. $((x_i),(\delta_i))$ on Γ, any accumulation point μ of (μ_i) satisfies, for every $g \in \mathcal{C}^\infty(\mathbb{R}^n,\mathbb{R})$, $\int\langle\nabla g(x), v\rangle d\mu(x,v) = 0$.*

Proof. Consider a fixed $\omega \in \Gamma$ (we omit the dependency in ω) such that (x_i) is bounded and $\int \|v\|^q d\mu_i(x,v)$ is uniformy bounded. Let μ be an accumulation point of (μ_i). Consider $g \in \mathcal{C}^\infty(\mathbb{R}^n,\mathbb{R})$. By uniform integrability,

$$\int\langle\nabla g(x), v\rangle d\mu(x,v) = \lim_{i\to\infty}\int\langle\nabla g(x), v\rangle d\mu_i(x,v)\,.$$

Set $t_i \coloneqq \sum_{j=0}^{i-1}\epsilon_j$ for every $i \geqslant 0$. Equivalently, $\epsilon_i = t_{i+1} - t_i$. Let $\mathsf{x}\colon \mathbb{R}_+ \to \mathbb{R}^n$ be the piecewise linear interpolated process associated with (x_i), defined by:

$$\mathsf{x}(t) \coloneqq x_i + v_{i+1}(t - t_i)$$

for every $t \in [t_i, t_{i+1})$. For every i, define the measure $\tilde{\mu}_i \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ given by:

$$\tilde{\mu}_i \coloneqq \frac{1}{t_{i+1}} \int_0^{t_{i+1}} \delta_{(\mathsf{x}(t), \mathsf{x}'(t))} dt \,.$$

Define $\varphi(x, v) \coloneqq \langle \nabla g(x), v \rangle$. After straightforward algebra,

$$\mu_i(\varphi) - \tilde{\mu}_i(\varphi) = \frac{1}{t_{i+1}} \sum_{j=0}^{i} \int_{t_j}^{t_{j+1}} (\varphi(x_j, v_{j+1}) - \varphi(\mathsf{x}(t), v_{j+1}))\, dt \,.$$

Note that $\|\nabla g(x)\|$ is Lipschitz continuous and bounded on K. Thus, there exists a constant $C > 0$ such that:

$$\begin{aligned}
|\mu_i(\varphi) - \tilde{\mu}_i(\varphi)| &\leqslant \frac{1}{t_{i+1}} \sum_{j=0}^{i} \int_{t_j}^{t_{j+1}} \|\nabla g(x_j) - \nabla g(\mathsf{x}(t))\| \|v_{j+1}\| dt \\
&\leqslant \frac{C}{t_{i+1}} \sum_{j=0}^{i} \|v_{j+1}\| \int_{t_j}^{t_{j+1}} (1 \wedge \|x_j - \mathsf{x}(t)\|) dt \\
&\leqslant \frac{C}{t_{i+1}} \sum_{j=0}^{i} \epsilon_j \|v_{j+1}\| (1 \wedge \epsilon_j \|v_{j+1}\|) \,.
\end{aligned}$$

For every $\varepsilon > 0$, one has $\epsilon_i \leqslant \varepsilon$ for every i large enough. Taking limits,

$$\limsup_{i \to \infty} |\mu_i(\varphi) - \tilde{\mu}_i(\varphi)| \leqslant C \int \|v\| (1 \wedge \varepsilon \|v\|) d\mu(x, v) \,.$$

As ε is arbitrary, and using that $\int \|v\| d\mu(x, v) < \infty$, we obtain $|\mu_i(\varphi) - \tilde{\mu}_i(\varphi)| \to 0$. Moreover,

$$\tilde{\mu}_i(\varphi) = \frac{1}{t_{i+1}} \int_0^{t_{i+1}} \langle \nabla g(\mathsf{x}(t)), \mathsf{x}'(t) \rangle dt = \frac{g(x_{i+1}) - g(x_0)}{t_{i+1}} \,.$$

Thus, $\tilde{\mu}_i(\varphi) \to 0$, and the result is proved. □

Corollary 4.10. *Let Assumptions 4.1, 4.4 and 4.5 hold. Then, $\mathbb{P}$-a.e. $((x_i), (\delta_i))$ on Γ, any accumulation point μ of (μ_i) is H-invariant in the sense of Definition 3.5.*

Proof. This follows immediately from the superposition principle, Th. 3.2 (which may be applied by Proposition 4.9), together with Prop. 4.8. □

We recall the definition of $\mathrm{zer}_s(H)$ in Th. 3.4.

Theorem 4.11. *Let Assumptions 4.1, 4.4 and 4.5 hold. Then, $\mathbb{P}$-a.e. on Γ, any accumulation point μ of (μ_i) is such that:*

i) $\pi(\mathrm{supp}\, \mu) \subset \mathrm{BC}(H)$.
ii) $v_\mu = 0$, $\pi_* \mu$*-a.e. on* $\mathrm{zer}_s(H)$.

Proof. Let μ be chosen as in Prop. 4.9. Define $\mu' \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ as:

$$\mu'(\varphi) = \int \varphi(x, v_\mu(x)) d\pi_* \mu(x)\,,$$

for every $\varphi \in \mathcal{C}_b(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$. By Prop. 4.9, $\int \langle \nabla g(x), v \rangle d\mu'(x,v) = 0$ for every $g \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$. As H is locally bounded, μ' is compactly supported. In sum, we conclude that μ' is closed. By Prop. 4.8, $v_\mu(x) \in H(x)$ for $\pi_*\mu$-almost every x. As $v_\mu = v_{\mu'}$, the conclusion of Th. 4.11 follows from Prop. 3.4. □

4.3. Essential accumulation set

Following the exposition of [10], we introduce the notion of essential accumulation points of the sequence (x_i). For every $U \subset \mathbb{R}^n$, define:

$$\tau_i^U \coloneqq \mu_i(U \times \mathbb{R}^n) = \frac{\sum_{j=0}^i \epsilon_j 1_U(x_j)}{\sum_{j=0}^i \epsilon_j}.$$

The quantity τ_i^U should be interpreted as the average amount of time during which the iterates are in the set U, up to iteration i.

Definition 4.12. For every $\omega \in \Omega$, a point x is an *essential accumulation* point of the sequence $(x_i(\omega))$ if for every neighborhood U of x, $\limsup_i \tau_i^U(\omega) > 0$.

We denote by $\operatorname{ess\,acc}(x_i)$ the set of essential accumulation points. Th. 4.11 has the following consequence.

Corollary 4.13. *Let Assumptions 4.1, 4.4 and 4.5 hold. Then, $\mathbb{P}$-a.e. on Γ,*

$$\operatorname{ess\,acc}(x_i) \subset \mathrm{BC}(H)\,.$$

In particular, if (DI) admits a Lyapunov function $V\colon \mathbb{R}^n \to \mathbb{R}$ for some $\Lambda \subset \mathbb{R}^n$, then $\operatorname{ess\,acc}(x_i) \subset \Lambda$.

Proof. The last statement is immediate from Prop. 2.3. We show the first statement. The proof follows from Th. 4.11 along with the following equality:

$$\operatorname{ess\,acc}(x_i) = \mathrm{cl} \bigcup_\mu \pi(\operatorname{supp}\mu)\,, \tag{18}$$

where the union is taken over all weak* accumulation points of (μ_i). This point is established in [10]. We provide a shorter proof, for completeness. Consider $x \in \operatorname{ess\,acc}(x_i)$. Choose $k > 0$ and set U equal to the closed ball of center x and radius $1/k$. It holds that $\limsup_i \pi_*\mu_i(U) > 0$. Thus there exists $\varepsilon > 0$, a closed measure μ and a subsequence (μ_{φ_i}) s.t. $\mu_{\varphi_i} \to \mu$ and $\pi_*\mu_i(U) \to \varepsilon$. As U is closed, $\pi_*\mu(U) \geqslant \varepsilon > 0$ by the Portmanteau Theorem. Thus, $\operatorname{supp}(\pi_*\mu) \cap U \neq \emptyset$. There exists $y_k \in \pi(\operatorname{supp}(\mu))$ s.t. $\|y_k - x\| \leqslant 1/k$. This proves the first inclusion.

Conversely, consider x in the right-hand side of (18), and let U be an open neighborhood of x. There exists a point $y \in U$ and an accumulation point μ of (μ_i) s.t. $y \in \operatorname{supp}(\pi_*\mu)$. As U is an open neighborhood of y, by definition of the support of a

measure, it holds that $\pi_*\mu(U) > 0$. There exists a subsequence (μ_{φ_i}) s.t. $\mu_{\varphi_i} \to \mu$. As U is open, $\liminf_i \pi_*\mu_{\varphi_i}(U) \geqslant \pi_*\mu(U) > 0$. Thus, $\limsup_i \pi_*\mu_i(U) > 0$, which proves that x is an essential accumulation point. □

4.4. Oscillation compensation

Definition 4.14. We say that (x_i) has the *oscillation compensation* property if for every $\psi \in \mathcal{C}_b(\mathbb{R}^n, \mathbb{R})$,

$$\lim_{i\to\infty} \frac{\sum_{j\leqslant i} \epsilon_j \psi(x_j) v_{j+1}}{\sum_{j\leqslant i} \epsilon_j} = 0\,. \tag{19}$$

As explained in the introduction, the oscillation compensation property implies a slow down of the sequence (x_i). Alternatively, it is not difficult to show that, $\mathbb{P}$-a.e. on Γ, (x_i) has the oscillation compensation property if, and only if:

$$\lim_{k\to\infty} \frac{\sum_{j\leqslant i_k} \epsilon_j \psi(x_j) v_{j+1}}{\sum_{j\leqslant i_k} \epsilon_j \psi(x_j)} = 0\,. \tag{20}$$

for every $\psi \in \mathcal{C}_b(\mathbb{R}^n, \mathbb{R}_+)$ and every subsequence (i_k) such that

$$\liminf_{k\to+\infty} \frac{\sum_{j\leqslant i_k} \epsilon_j \psi(x_j)}{\sum_{j\leqslant i_k} \epsilon_j} > 0. \tag{21}$$

Corollary 4.15. *Let Assumptions 4.1, 4.4 and 4.5 hold. Assume that the following condition holds:*

i) $\mathrm{BC}(H) = \mathrm{zer}_s(H)$,

or the stronger condition:

ii) There exists a path-differentiable Lyapunov function $V\colon \mathbb{R}^n \to \mathbb{R}$ for some $\Lambda \subset \mathbb{R}^n$, such that $H(x) \cap \partial V(x)^\perp = \{0\}$ for all $x \in \Lambda$.

Then, $\mathbb{P}$-a.e. on Γ, (x_i) has the oscillation compensation property.

Proof. We first establish that the second condition implies the first one. The second condition implies that $\Lambda \subset \mathrm{zer}(H)$. Moreover, consider $\gamma \in S_H(x)$ for an arbitrary $x \in \Lambda$. By Def. 2.2, $\gamma([0, +\infty)) \in \Lambda$, and $V \circ \gamma$ is constant. By the path-differentiability, $0 = \langle v, \gamma'(t)\rangle$ for all $v \in \partial V(\gamma(t))$ and for almost all t. Thus, $\gamma'(t)$ is orthogonal to $\partial V(\gamma(t))$. As $\gamma(t) \in \Lambda$, we obtain that $\gamma'(t) = 0$. Thus, $\gamma(t) = x$ for all x. This proves that $x \in \mathrm{zer}_s(H)$. Thus, $\mathrm{BC}(H) \subset \Lambda \subset \mathrm{zer}_s(H)$.

Now consider the case where the first condition holds. By Th. 4.11, it holds that every accumulation point μ of (μ_i) satisfies $v_\mu = 0$ $\pi_*\mu$-a.e. This directly implies the statement. □

5. Applications

5.1. Stochastic subgradient descent

The following result generalizes results of [10] to the stochastic setting.

Proposition 5.1. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a path-differentiable function. Consider a random sequence (x_i) satisfying:*

$$x_{i+1} \in x_i - \epsilon_i \partial f(x_i) + \epsilon_i \eta_{i+1}\,.$$

Let Assumptions 4.4, 4.5 hold, with $\delta_i \coloneqq 0$. The following holds a.e. on the event $\{\sup_i \|x_i\| < \infty\}$:

i) $\operatorname{ess\,acc}(x_i) \subset \{x \in \mathbb{R}^n \,:\, 0 \in \partial f(x)\}$.
ii) *(x_i) has the oscillation compensation property.*

The reason why we require the function f to be path-differentiable is to ensure that f is a Lyapunov function for its (negative) Clarke subdifferential $-\partial f$, and for Λ equal to the Clarke critical set. In contrast, weaker assumptions, like for example f Lipschitz, can potentially yield dynamics that are too unwieldy. In addition, the path-differentiability assumption is weak enough to cover most cases of interest in applications; see [10] for a discussion.

Proof. Observing that $V \coloneqq f$ is a path-differentiable Lyapunov function for the differential inclusion $x'(t) \in -\partial f(x(t))$, the conclusion follows from Cor. 4.13 and Cor. 4.15. □

Remark 5.2. More generally, the subdifferential ∂f can be replaced by any *conservative field* $H\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (see [11] for definitions), hence extending the results to other algorithms of the flavour of the stochastic subgradient descent, such as the celebrated back-propagation algorithm.

5.2. Stochastic heavy ball

Let $f\colon \mathbb{R}^m \to \mathbb{R}$ be a path-differentiable function, for some integer $m \geqslant 1$. We consider a random sequence (x_i) on $\mathbb{R}^m \times \mathbb{R}^m$, satisfying $x_i = (q_i, p_i)$, where:

$$\begin{aligned} q_{i+1} &= q_i + \alpha_i p_{i+1}\,, \\ p_{i+1} &\in (1-\beta_i) p_i - \beta_i \partial f(q_i) + \beta_i \eta_{i+1}\,, \end{aligned} \tag{22}$$

where (α_i), (β_i) are two sequences of positive step sizes. This method is sometimes refered to as Stochastic Heavy Ball (SHB) [3, 18]. It can be rewritten as a function of the variable (q_i) in the following way:

$$q_{i+1} \in q_i + \beta_i' \left(-\partial f(q_i) + \eta_{i+1}\right) + \alpha_i'(q_i - q_{i-1})\,,$$

with the change of variable $\alpha_i' \coloneqq \alpha_i \alpha_{i-1}^{-1}(1-\beta_i)$ and $\beta_i' \coloneqq \alpha_i \beta_i$. We make the following assumption:

Assumption 5.3. The sequences (α_i), (β_i) are positive, and satisfy $\sum \alpha_i = +\infty$ and $\alpha_i \to 0$. Moreover, there exists $c > 0$ s.t. $\alpha_i/\beta_i \to c$.

Proposition 5.4. *Let $f\colon \mathbb{R}^m \to \mathbb{R}$ be a path-differentiable function. Consider the sequence (x_i), $x_i = (q_i, p_i)$ given by Eq. (22). Let Assumptions 4.4 and 5.3 hold. Then,*

i) $\operatorname{ess\,acc}(x_i) \subset \{(x, 0) \,:\, 0 \in \partial f(x)\}$.
ii) *(x_i) has the oscillation compensation property.*

Proof. In order to keep the exposition concise, assume $\alpha_i = c\beta_i$. The iterates (x_i) have the form (11), where $\delta_i = 0$, where η_{i+1} is replaced with $(0, \eta_{i+1})$, and where the map $H\colon \mathbb{R}^m \times \mathbb{R}^m \rightrightarrows \mathbb{R}^m \times \mathbb{R}^m$ is given by:

$$H(q,p) \coloneqq (-cp, \partial f(q) - p)\,.$$

Considering (DI), the function $V\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ given by $V(q,p) \coloneqq f(q) + c\frac{\|p\|^2}{2}$, is a path differentiable Lyapunov function for the set $\mathbb{R}^m \times \{0\}$. Indeed, any solution $\gamma(t) = (q(t), p(t))$ to (DI) satisfies, for almost every t,

$$(V \circ \gamma)'(t) = -c\|p(t)\|^2\,. \tag{23}$$

Thus, $\mathrm{BC}(H) \subset \mathbb{R}^m \times \{0\}$. Consider an arbitrary recurrent point x (if it exists). Then, $x = (q, 0)$ for some $q \in \mathbb{R}^m$. Consider $\gamma \in S_H(x)$ s.t. $x \in L(\gamma)$. Set $\gamma(t) = (q(t), p(t))$. As the function $V \circ \gamma$ is constant, Eq. (23) implies that $p(t) = 0$ a.e. As $q'(t) = -cp(t) = 0$, we conclude that $q(t) = x$ for all t. Lastly, the inclusion $p'(t) \in \partial f(q(t)) - p(t)$ reads $0 \in \partial f(x)$. Thus $x \in \mathrm{zer}(-\partial f)$. This proves that $\mathrm{BC}(H) \subset \{(x,0) : 0 \in \partial f(x)\}$. As γ is constant, this also proves that $\mathrm{BC}(H) = \mathrm{zer}_s(H)$. The conclusion follows from Cor. 4.13 and Cor. 4.15. □

5.3. Fictitious play and best-response game dynamics

There are many possible dynamics for games; see for example [21]. The setting in this section roughly corresponds to the one developed in [6, 7]. We consider the dynamics of fictitious play, which in this case means that the players play simultaneously at each turn, and they choose their respective actions with knowledge of the sequence (ξ_n) of averages of the actions of a game (to be specified below). The dynamics of interest will be that of the sequence (ξ_n).

Consider a game with m players and, for each $i = 1, \dots, m$, let X^i be a compact, convex subsets of $\mathbb{R}^n$ that will play the role of the possible (mixed) actions of player i. Typically, X^i is a simplex, so that there are finitely many pure actions available to player i, corresponding to the vertices of the simplex. Here are some useful notations: we will denote the set of actions available to the opponents of player i by

$$X^{-i} = X^1 \times X^2 \times \cdots \times X^{i-1} \times X^{i+1} \times \cdots \times X^m \subset \mathbb{R}^{n(m-1)},$$

and for $x = (x^1, \dots, x^m) \in X^1 \times \cdots \times X^m$,

$$x^{-i} = (x^1, x^2, \dots, x^{i-1}, x^{i+1}, \dots, x^m) \in X^{-i}.$$

Each player also has a payoff function $U^i\colon X^i \times X^{-i} \to \mathbb{R}$. Player i wants to maximize her payoff U^i, so for each given vector $\xi^{-i} \in X^{-i}$ of actions of its opponents, it makes sense to define the *best response set* by

$$BR^i(\xi^{-i}) = \operatorname*{arg\,max}_{\xi^i \in X^i} U^i(\xi^i, \xi^{-i}) \subseteq X^i,$$

and we will assume that the functions U^i are such that each $BR^i(\xi^{-i})$ is convex and compact. It follows from the Berge maximum theorem that, if U^i is continuous, then also the maps BR^i are upper semicontinuous.

Let $x_\ell = (x_\ell^1, \dots, x_\ell^m) \in X^1 \times \cdots \times X^m$ be the profile of actions taken by all players at stage ℓ. We denote the average of the actions up to stage n by

$$\xi_n = \frac{1}{n} \sum_{\ell=0}^{n} x_\ell \tag{24}$$

Player i chooses, at each stage $n+1$, a point

$$x_{n+1}^i \in BR^i(\xi_n^{-i}). \tag{25}$$

In the game-theoretical literature, this is known as the *fictitious play assumption.*

We let $H\colon X^1 \times \cdots \times X^m \rightrightarrows X^1 \times \cdots \times X^m$ be the map given by

$$H(\xi^1, \dots, \xi^m) = (BR^1(\xi^{-1}) - \xi^1) \times \cdots \times (BR^m(\xi^{-m}) - \xi^m). \tag{26}$$

Then $H(\xi_n)$ is a nonempty, convex and compact set (hence consistent with Assumption 4.1) that contains $x_{n+1} - \xi_n$.

We have, by (25),

$$\xi_{n+1} - \xi_n \in \frac{1}{n+1} H(\xi_n),$$

and the sequence $(\xi_n, \eta_n, \epsilon_n)_{n \geqslant 1}$, with $\eta_n = 0$ and $\epsilon_n = 1/n$, satisfies (11) and Assumptions 4.4 and 4.5, and the results of Th. 4.11 hold.

A *Nash equilibrium* is a point $\xi \in X^1 \times \cdots \times X^m$ such that $U(\xi) \geqslant U(x^i, \xi^{-i})$ for all $x^i \in X^i$ and $i = 1, \dots, m$; in other words, it is a point such that $\xi^i \in BR^i(\xi^{-i})$ for all i. Observe that this is equivalent to having $0 \in H(\xi)$, so the constant solutions $\xi_0 = \xi_1 = \dots$ are precisely the Nash equilibria.

We also have:

Proposition 5.5. *The essential accumulation set of the sequence (ξ_i) is contained in the Birkhoff center $\mathrm{BC}(H)$, and satisfies the oscillation compensation property on the stable zeros $\mathrm{zer}_s(H)$.*

Remark 5.6. There are a few types of games that admit Lyapunov functions, among which we count zero-sum and potential games; see for example [22] and [6, Section 5.3], respectively. In these cases, Theorem 3.7 and Corollary 4.15 (with the second hypothesis) apply, and thus recover the results that were obtained before for example in [6, 19, 22]. Although in many of these games the orbits of fictitious play converge to the Nash equilibrium, there are some examples in which this is not the case; notably, in the generalized rock-paper-scissors game the orbits accumulate around the so-called Shapely polygons and they never converge [19].

Remark 5.7. It is possible to add a noise and to exchange the condition (25) with the assumption that the expectation of x^i falls in $BR^i(\xi^{-i})$, and still get a statement like the one in Proposition 5.5.

Appendix A. A lemma on martingales

The following lemma is a consequence of Chow's martingale theorem (see [20, Theorem 2.17]). The proof can be found in [15, Th.1.III.11], and is partly reproduced here, for completeness.

Lemma A.1. *Let $M_i = \sum_{j=1}^{i} X_i$ be a $\mathbb{R}^n$-valued $(\mathcal{F}_i)$-adapted martingale. Consider $1 \leqslant q \leqslant 2$ and $r > 0$. Define $s_i := \sum_{j \leqslant i} \mathbb{E}(\|X_{i+1}\|^q)|\mathcal{F}_i)$. Then, there exists a random variable M_∞, finite a.e., such that $M_i \to M_\infty$ a.s. on the set $\{\lim_i s_i < \infty\}$. Moreover, on the set $\{\lim_i s_i = \infty\}$,*

$$\frac{M_i}{(s_{i-1}\log^{1+r}(s_{i-1}))^{1/q}} \to 0 \ a.s.$$

Proof. The first point of the Lemma is a consequence of Chow's theorem (see [20, Theorem 2.17]). Therefore, we only establish the second point, following the arguments of [15, Th.1.III.11]. We introduce the martingale $N_i = \sum_{j \leqslant i} Y_i$ where $Y_i := (s_{i-1}\log^{1+r}(s_{i-1}))^{-1/q} X_i$. We observe that $\mathbb{E}(\|Y_{i+1}\|^q|\mathcal{F}_i) = \alpha_j (s_{i-1}\log^{1+r}(s_{i-1}))^{-1}$ is summable, because $\int_e^\infty (t\log^{1+\beta}(t))^{-1} dt = \int_1^\infty t^{-(1+r)} dt < \infty$. By the first point, there exists an almost surely finite random variable N_∞ such that $N_i \to N_\infty$. The conclusion follows from the Kronecker lemma [20, Section 2.6, pp.31]. □

Acknowledgements

The authors are very grateful for numerous enlightening discussions with Jérôme Bolte and Edouard Pauwels. This line of research grew out of inspiring conversations with Sylvain Sorin. The second author wishes to acknowledge the generous support of ANR-3IA Artificial and Natural Intelligence Toulouse Institute, and of the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-19-1-7026.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] 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.

[3] A. Barakat, P. Bianchi, W. Hachem, and S. Schechtman. Stochastic optimization with momentum: convergence, fluctuations, and traps avoidance. *Electronic Journal of Statistics*, 15(2):3892–3947, 2021.

[4] 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.

[5] M. Benaïm and M. W. Hirsch. Asymptotic pseudotrajectories and chain recurrent flows, with applications. *J. Dynam. Differential Equations*, 8(1):141–176, 1996.

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

[7] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. II. Applications. *Math. Oper. Res.*, 31(4):673–695, 2006.

[8] M. Benaïm and S. J. Schreiber. Ergodic properties of weak asymptotic pseudotrajectories for semiflows. *J. Dynam. Differential Equations*, 12(3):579–598, 2000.

[9] Patrick Bernard. Young measures, superposition and transport. *Indiana Univ. Math. J.*, 57(1):247–275, 2008.

[10] J. Bolte, E. Pauwels, and R. Rios-Zertuche. Long term dynamics of the subgradient method for lipschitz path differentiable functions. *arXiv preprint arXiv:2006.00098*, 2020.

[11] Jérôme Bolte and Edouard Pauwels. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. *Mathematical Programming*, pages 1–33, 2020.

[12] V. S. Borkar. *Stochastic approximation. A dynamical systems viewpoint*. Cambridge University Press, Cambridge; Hindustan Book Agency, New Delhi, 2008.

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

[14] D Charalambos and Border Aliprantis. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag Berlin and Heidelberg GmbH & Company KG, 2013.

[15] Marie Duflo. *Algorithmes stochastiques*. Springer, 1996.

[16] 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.

[17] M. Faure and G. Roth. Ergodic properties of weak asymptotic pseudotrajectories for set-valued dynamical systems. *Stoch. Dyn.*, 13(1):1250011, 23, 2013.

[18] S. Gadat, F. Panloup, and S. Saadane. Stochastic heavy ball. *Electronic Journal of Statistics*, 12(1):461–529, 2018.

[19] Andrea Gaunersdorfer and Josef Hofbauer. Fictitious play, shapley polygons, and the replicator equation. *Games and Economic Behavior*, 11(2):279–303, 1995.

[20] Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 2014.

[21] Josef Hofbauer. Deterministic evolutionary game dynamics. In *Evolutionary Game Dynamics*, volume 69, pages 61–79. American Mathematical Society Providence, RI, 2011.

[22] Josef Hofbauer and Sylvain Sorin. Best response dynamics for continuous zero–sum games. *Discrete & Continuous Dynamical Systems-B*, 6(1):215, 2006.

[23] 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.

[24] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE transactions on automatic control*, 22(4):551–575, 1977.

[25] W. Miller and E. Akin. Invariant measures for set-valued dynamical systems. *Trans. Amer. Math. Soc.*, 351(3):1203–1225, 1999.

[26] H. Robbins and S. Monro. A stochastic approximation method. In *Herbert Robbins Selected Papers*, pages 102–109. Springer, 1985.