A unified stochastic approximation framework for learning in games

Panayotis Mertikopoulos, Ya-Ping Hsieh, Volkan Cevher

A UNIFIED STOCHASTIC APPROXIMATION FRAMEWORK FOR LEARNING IN GAMES

PANAYOTIS MERTIKOPOULOS*, YA-PING HSIEH§, AND VOLKAN CEVHER♯

Abstract. We develop a flexible stochastic approximation framework for analyzing the long-run behavior of learning in games (both continuous and finite). The proposed analysis template incorporates a wide array of popular learning algorithms, including gradient-based methods, the exponential / multiplicative weights algorithm for learning in finite games, optimistic and bandit variants of the above, etc. In addition to providing an integrated view of these algorithms, our framework further allows us to obtain several new convergence results, both asymptotic and in finite time, in both continuous and finite games. Specifically, we provide a range of criteria for identifying classes of Nash equilibria and sets of action profiles that are attracting with high probability, and we also introduce the notion of *coherence*, a game-theoretic property that includes strict and sharp equilibria, and which leads to convergence in finite time. Importantly, our analysis applies to both oracle-based and bandit, payoff-based methods – that is, when players only observe their realized payoffs.

1. Introduction

The prototypical setting of online learning in games can be summarized as follows:

(1) At each stage of a repeated decision process, every player selects an action.

(2) The players receive a reward determined by their chosen actions and their individual payoff functions – assumed a priori unknown.

(3) Based on these payoffs and any other observed information, the players update their actions and the process repeats.

A key question that arises in this general setting is whether the players eventually settle down to a stable profile from which no player has an incentive to deviate. Put differently:

Does the players' learning process converge to a Nash equilibrium?

This question has been at the forefront of game-theoretic research ever since the field's earliest steps, and it has recently received renewed attention owing to its connection to data science, multi-agent reinforcement learning, networks, and many other applications where agents are called to make decisions under uncertainty. The first positive answer here was

* Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France.

§ Institute for Machine Learning, Universitaetstrasse 6, 8092 Zurich, Switzerland.

♯ Laboratory for Information and Inference Systems, IEL STI EPFL, 1015 Lausanne, Switzerland.

E-mail addresses: panayotis.mertikopoulos@imag.fr, yaping.hsieh@inf.ethz.ch, volkan.cevher@epfl.ch.

2020 *Mathematics Subject Classification.* Primary 91A10, 91A26; secondary 68Q32, 68T02.

Key words and phrases. Nash equilibrium; continuous games; finite games; stochastic approximation; variational stability; primal attractors; coherence.

The authors are grateful to Victor Boone, Pierre-Louis Cauvin, Angeliki Giannou, Kyriakos Lotidis, Sylvain Sorin, and Manolis Vlatakis for many fruitful discussions. Part of this work was done while P. Mertikopoulos was visiting the Simons Institute for the Theory of Computing.

given by Brown [10] and Robinson [58] who introduced the so-called fictitious play (FP) process and established its convergence in 2-player zero-sum finite games. Since then, a vast number of works have examined the convergence of a diverse array of learning procedures in different classes of games: smoothed versions of fictitious play in potential, zero-sum, supermodular and $2 \times m$ games [29, 40], gradient methods in continuous min-max games [1, 36, 54], the numerous variants of mirror descent (MD) and other regularized learning schemes in monotone [34, 44, 46, 50, 66, 67], smooth [65], and potential games [8, 12, 39], etc.

At the same time, the well-known impossibility results of Hart & Mas-Colell [24, 25] rule out the prospect of an unconditionally positive answer: there is no uncoupled learning rule – deterministic *or* stochastic – that converges to Nash equilibrium in all games. As a result, contemporary research on the subject has focused on extending the classes of games in which positive results can be obtained, relaxing the feedback requirements of the players' learning process, and understanding the convergence failures of popular learning algorithms. This has in turn revealed a very fragile convergence landscape: for example, standard gradient methods are known to converge in *strictly* monotone games [44], but they may diverge in bilinear min-max games (which are monotone but not strictly so) [14]; this failure can be overcome by means of an extra-gradient / optimistic correction term [36, 54], but it re-emerges in the presence of randomness and uncertainty [32]; and if the game is perturbed even slightly, all these methods – gradient, extra-gradient and optimistic – may end up converging to a spurious limit cycle containing no critical / equilibrium points whatsoever [33, 45].

Since these negative results are all pointwise, it is natural to turn to *sets* and instead ask:

Which sets of actions are stable and attracting under a given learning process?
Are these sets robust to the choice of method, initialization, or available information?

From a dynamic standpoint, the established notion of stability is that of an *attractor*, which characterizes outcomes that are resilient to small perturbations in the dynamics' initialization. However, the questions above call for much more: ideally, the sets under consideration should be stable in a class of learning procedures which, other than a few broad unifying features, may have radically different update structures, feedback requirements, etc. Our aim in this paper is to identify such sets and to quantify their stability and convergence properties.

Our contributions. The basis of our analysis is a flexible stochastic approximation framework which we call the *regularized Robbins–Monro* (RRM) template in reference to the seminal method of Robbins & Monro [57] and the "follow-the-regularized-leader" (FTRL) family of algorithms of Shalev-Shwartz & Singer [63]. This framework hinges on an implicit regularization mechanism in the spirit of Nesterov [50] and encompasses as special cases many popular learning algorithms: gradient-based methods for continuous games [1, 44, 50], the exponential / multiplicative weights family of algorithms for finite games [2, 41, 70], optimistic [14, 22, 36, 46, 54, 55] and bandit, payoff-based variants of the above [9, 12, 26, 66], etc. We then seek to analyze the long-run behavior of this "parent scheme" via a suitable dynamical system which captures its mean, continuous-time limit, and which is sufficiently rich to accommodate different types of feedback and update structures.

In this general context, our main results can be summarized along the following axes:

1. **Characterization of limit sets:** First, we show that the limit sets of RRM methods are internally chain transitive (ICT) in the associated mean dynamics, i.e., they are invariant and contain no smaller attractors. This property applies to all games satisfying a certain coercivity condition – which we call "*subcoercivity*" – and it allows us to deduce a series of almost sure equilibrium convergence results for min-max and potential games.

2. **Characterization of attractors:** We further show that sets that admit a local energy function (relative to the mean dynamics mentioned above) are attracting with high probability – or globally attracting with probability 1, depending on the energy function. As a corollary of this result, we readily infer convergence to Nash equilibrium in all strictly monotone games, and we likewise derive a series of high probability convergence results to equilibria that satisfy a certain variational stability requirement.

3. **Fast convergence to coherent sets:** Finally, we introduce the notion of *coherence* – an algorithm-agnostic concept which covers strict Nash equilibria in finite games, sharp equilibria in continuous games, linear programs, etc. – and we show that RRM methods converge to such sets under significantly weaker conditions for their runtime parameters (step-size, sampling radius, etc.). In addition, we show that projection-based methods (as opposed to interior-valued ones) converge to coherent sets in a *finite* number of iterations.

An appealing feature of our analysis is that it applies to both *first-order* ("oracle-based") and *zeroth-order* ("payoff-based" or "bandit") methods. More to the point, our results can be easily adapted to many other learning algorithms in the literature, reducing in this way the number of ad hoc elements required to analyze a given method. Of course, given the breadth of the relevant literature, it is impossible to include here *all* methods covered by the proposed RRM template – or that *could* be covered modulo minor modifications. Our choice of examples is only meant to illustrate different trends in the literature, and to show how some algorithms that initially seem unrelated – like the dampened gradient approximation (DGA) method of [8] – can be included in our framework.

Paper outline. In Section 2, we introduce the game-theoretic background of our work, including the various solution concepts that we use throughout our paper (critical points, Nash equilibria, variationally stable states, etc.). Subsequently, in Section 3, we introduce a range of well-known algorithms for learning in games, and we show how they can be seen as special instances of the RRM blueprint. Our analysis proper begins in Section 4, where we introduce the notion of subcoercivity and present our ICT convergence results. Subsequently, in Sections 5 and 6, we state and prove our main convergence results for stochastically attracting and coherent sets respectively.

2. Preliminaries

Notation. In what follows, $\mathcal{V}$ will denote a d-dimensional real space with norm $\|\cdot\|$. We will also write $\mathcal{Y} \coloneqq \mathcal{V}^*$ for the dual space of $\mathcal{V}$, $\langle y, x\rangle$ for the canonical pairing between $y \in \mathcal{Y}$ and $x \in \mathcal{V}$, and $\|y\|_* \coloneqq \max\{\langle y, x\rangle : \|x\| \leq 1\}$ for the induced dual norm on $\mathcal{Y}$. As is customary, if $\mathcal{V}$ is Euclidean, we will not distinguish between primal and dual vectors. Finally, if $f\colon \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ is a convex function on $\mathcal{V}$, we will write $\operatorname{dom} f \coloneqq \{x \in \mathcal{V} : f(x) < \infty\}$ for its effective domain, $\partial f(x) \coloneqq \{y \in \mathcal{Y} : f(x') \geq f(x) + \langle y, x' - x\rangle \text{ for all } x' \in \mathcal{V}\}$ for the subdifferential of f at x, and $\operatorname{dom} \partial f \coloneqq \{x \in \mathcal{V} : \partial f(x) \neq \varnothing\}$ for the domain of subdifferentiability of f.

2.1. Games in normal form.

Throughout the sequel, we will focus on games with a finite number of players $i \in \mathcal{N} = \{1, \dots, N\}$, each selecting an *action* x_i from some closed convex subset $\mathcal{X}_i$ of a d_i-dimensional normed space $\mathcal{V}_i$. Gathering all players together, we will write $\mathcal{X} = \prod_i \mathcal{X}_i$ for the space of all *action profiles* $x = (x_i)_{i\in\mathcal{N}}$ and $d = \sum_i d_i$ for the dimension of the ambient space $\mathcal{V} = \prod_i \mathcal{V}_i$. Finally, when we want to distinguish between the action of the i-th player and that of all other players, we will employ the shorthand $(x_i; x_{-i})$.

Given an action profile $x \in \mathcal{X}$, each player $i \in \mathcal{N}$ is assumed to receive a reward $u_i(x) \equiv u_i(x_i; x_{-i})$ based on an associated *payoff function* $u_i \colon \mathcal{X} \to \mathbb{R}$. In terms of regularity, we will tacitly assume that u_i is differentiable and we will write

$$v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}) \quad \text{and} \quad v(x) = (v_i(x))_{i \in \mathcal{N}} \tag{1}$$

for the players' *individual payoff gradients* and the ensemble thereof. Finally, unless explicitly mentioned otherwise, we will treat each $v_i(x)$ as an element of the corresponding dual space $\mathcal{Y}_i = \mathcal{V}_i^*$ of $\mathcal{V}_i$, and we will make the following blanket assumption:

Assumption 1. The players' payoff functions are *Lipschitz continuous and smooth*, i.e., there exist constants $G_i, L_i \geq 0$, $i \in \mathcal{N}$, such that

$$|u_i(x') - u_i(x)| \leq G_i \|x' - x\| \quad \text{and} \quad \|\nabla u_i(x') - \nabla u_i(x)\|_* \leq L_i \|x' - x\|. \tag{2}$$

for all $x, x' \in \mathcal{X}$, $i \in \mathcal{N}$. For concision, we will also write $G \coloneqq \max_i G_i$ and $L \coloneqq \max_i L_i$.

A *continuous game in normal form* is then defined as a tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ with players, actions and payoff functions as above. For concreteness, we provide some examples below:

Example 2.1 (Min-max games)**.** Consider two players $i \in \{1, 2\}$ with action spaces $\mathcal{X}_1$ and $\mathcal{X}_2$, and payoff functions $u_1 = -\mathcal{L} = -u_2$ for some smooth function $\mathcal{L} \colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$. Player 1 (the "min" player) seeks to minimize $\mathcal{L} = -u_1$ whereas Player 2 (the "max" player) seeks to maximize $\mathcal{L} = u_2$. In many applications, $\mathcal{L}$ is (strictly) convex-concave, in which case von Neumann's theorem asserts that the game $\min_{x_1 \in \mathcal{X}_1} \max_{x_2 \in \mathcal{X}_2} \mathcal{L}(x_1, x_2)$ always admits a solution if $\mathcal{X}_1 \times \mathcal{X}_2$ is compact. ◊

Example 2.2 (Cournot oligopolies)**.** Consider N firms supplying the market with a quantity $x_i \in [0, C_i]$ of some good up to each firm's capacity C_i. The good is priced as a function $P(x) = a - b \sum_{i=1}^N x_i$ of the total quantity of the good in the market, so the net utility of the i-th firm is $u_i(x) = x_i P(x) - c_i x_i$ where a, b and c_i are market-related positive constants. The resulting game $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ is known as a Cournot competition game and it plays a central role in economic theory. ◊

Example 2.3 (Finite games)**.** In a finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$, each player $i \in \mathcal{N}$ chooses an action α_i from some finite set $\mathcal{A}_i$; the players' payoffs are then determined by the action profile $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathcal{A} \coloneqq \prod_i \mathcal{A}_i$ and an ensemble of payoff functions $u_i \colon \mathcal{A} \to \mathbb{R}$, $i = 1, \dots, N$. In the *mixed extension* of Γ, a player may pick an action according to a probability distribution $x_i \in \Delta(\mathcal{A}_i)$: this is known as a *mixed strategy*, and the corresponding mixed payoff to the i-th player is $u_i(x) = \sum_{\alpha \in \mathcal{A}} x_\alpha u_i(\alpha)$ where $x_\alpha = \prod_i x_{i\alpha_i}$ is the probability of the action profile $\alpha = (\alpha_1, \dots, \alpha_N)$.

Letting $\mathcal{X}_i = \Delta(\mathcal{A}_i)$, the mixed extension of Γ is defined as the continuous game $\Delta(\Gamma) = \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$. For posterity, we note here that the "payoff gradient" of each player $i \in \mathcal{N}$ is simply their mixed payoff vector, i.e., $v_i(x) = \nabla_{x_i} u_i(x) = (u_i(\alpha_i; x_{-i}))_{\alpha_i \in \mathcal{A}_i}$. ◊

2.2. **Solution concepts.** The standard solution concept in game theory is that of a *Nash equilibrium*, i.e., an action profile that is resilient to unilateral deviations. Formally, $x^* \in \mathcal{X}$ is a Nash equilibrium of a game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ if

$$u_i(x^*) \geq u_i(x_i; x^*_{-i}) \quad \text{for all } x_i \in \mathcal{X}_i,\ i \in \mathcal{N}. \tag{NE}$$

Nash equilibria always exist if $\mathcal{X}$ is compact and each u_i is individually concave in x_i [15]. Otherwise, equilibria may not exist, in which case the following relaxations become relevant:

(1) *Local Nash equilibria*, i.e., profiles $x^* \in \mathcal{X}$ for which (NE) holds locally:

$$u_i(x^*) \geq u_i(x_i; x^*_{-i}) \quad \text{for all } x \text{ in a neighborhood } \mathcal{U} \text{ of } x^* \text{ in } \mathcal{X}. \tag{LNE}$$

(2) *Critical points*, i.e., profiles $x^* \in \mathcal{X}$ that satisfy the first-order stationarity condition:

$$\tfrac{d}{dt}\big|_{t=0^+} u_i(x_i^* + t(x_i - x_i^*); x_{-i}^*) \leq 0 \quad \text{for all } x_i \in \mathcal{X}_i,\ i \in \mathcal{N}. \tag{FOS}$$

Equivalently, (FOS) can be reformulated as a Stampacchia variational inequality of the form

$$\langle v(x^*), x - x^* \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}. \tag{SVI}$$

The solutions of (SVI) are precisely the fixed points of the "linearized" best-response correspondence $x \mapsto \arg\max_{x' \in \mathcal{X}} \langle v(x), x' \rangle$ so, by standard fixed point arguments, the set of critical points of $\mathcal{G}$ is always nonempty if $\mathcal{X}$ is compact.

Dually to the above, the *Minty variational inequality* associated to $\mathcal{G}$ is

$$\langle v(x), x - x^* \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}. \tag{MVI}$$

It is straightforward to verify that the solutions of (MVI) comprise a convex set of Nash equilibria of $\mathcal{G}$, so (MVI) can be seen as an equilibrium refinement criterion for $\mathcal{G}$. Taking this a step further, a state $x^* \in \mathcal{X}$ is said to be *variationally stable* if

$$\langle v(x), x - x^* \rangle < 0 \quad \text{for all } x \neq x^* \text{ in a neighborhood } \mathcal{U} \text{ of } x^* \text{ in } \mathcal{X} \tag{VS}$$

and x^* is called *neutrally stable* if the strict inequality "$<$" in (VS) is relaxed to "$\leq$", i.e., if

$$\langle v(x), x - x^* \rangle \leq 0 \quad \text{for all } x \text{ in a neighborhood } \mathcal{U} \text{ of } x^* \text{ in } \mathcal{X}. \tag{NS}$$

Finally, we say that x^* is *globally variationally stable* [resp. *globally neutrally stable*] if (VS) [resp. (NS)] holds with $\mathcal{U} = \mathcal{X}$ (i.e., for all $x \in \mathcal{X}$).

In general, the solution concepts discussed above are related as follows:

$$\begin{array}{ccccccc} \text{(GVS)} & \Longrightarrow & \text{(GNS)} & \equiv & \text{(MVI)} & \Longrightarrow & \text{(NE)} \\ \Downarrow & & \Downarrow & & & & \Downarrow \\ \text{(VS)} & \Longrightarrow & \text{(NS)} & \Longrightarrow & \text{(LNE)} & \Longrightarrow & \text{(FOS)} \equiv \text{(SVI)} \end{array} \tag{3}$$

Without further assumptions, the implications in (3) are all one-way; in the next section, we discuss a number of cases where some (or all) of these implications become equivalences.

Remark 1. In the optimization literature, the direction of (SVI) / (MVI) is reversed because optimization problems are usually stated in terms of cost minimization. The maximization viewpoint is more common in games, so we will maintain the "$\leq$" direction throughout. ◊

Remark 2. The notion of variational stability was introduced in [44] and echoes the seminal concept of *evolutionary stability* as introduced by Maynard Smith & Price [42] in the context of population games. Informally, "variational stability" is to games with a finite number of players and a continuum of actions what "evolutionary stability" is to games with a continuum of players and a finite number of actions; for an in-depth discussion, cf. [44]. ◊

2.3. **Special cases of interest.** We close this section with a discussion of some special cases and examples of the above definitions that will play a major role in the sequel.

► **Monotone games.** A game is *monotone* if it satisfies the monotonicity condition

$$\langle v(x') - v(x), x' - x \rangle \leq 0 \quad \text{for all } x, x' \in \mathcal{X}. \tag{Mon}$$

The strict version of this requirement (i.e., that equality holds if and only if $x = x'$) is sometimes referred to as *diagonal strict concavity* (DSC), a terminology due to Rosen [59]. In monotone games, the solutions of (MVI) and (SVI) coincide, leading to the string of equivalences

$$\text{(MVI)} \iff \text{(NE)} \iff \text{(LNE)} \iff \text{(FOS)} \equiv \text{(SVI)} \tag{4}$$

By comparison, if a game is strictly monotone, every implication in (3) becomes an equivalence, so the game admits a unique, globally variationally stable Nash equilibrium. Examples 2.1 and 2.2 are both strictly monotone (assuming $\mathcal{L}$ is strictly convex-concave in Example 2.1); other examples include socially concave games [18], Cournot oligopolies [48], Kelly auctions [35, 69], congestion control [18], and many other classes of problems.

► **Potential games.** First formalized by Monderer & Shapley [48], potential games admit a *potential function* $\Phi\colon \mathcal{X} \to \mathbb{R}$ such that

$$u_i(x_i; x_{-i}) - u_i(x_i'; x_{-i}) = \Phi(x_i; x_{-i}) - \Phi(x_i'; x_{-i}) \quad \text{for all } x, x' \in \mathcal{X} \text{ and all } i \in \mathcal{N}. \tag{Pot}$$

If $\mathcal{G}$ is a potential game, we have $v(x) = \nabla\Phi(x)$ so *a*) any local maximum of Φ is a local Nash equilibrium of $\mathcal{G}$; and *b*) any *strict* local maximum of Φ is variationally stable. The Cournot oligopoly of Example 2.2 is a textbook example of a potential game; other examples include finite congestion games [60], power allocation in wireless networks [62], etc.

► **Second-order stationarity.** Our next example concerns critical points that satisfy a condition similar to second-order sufficient conditions in optimization, namely

$$z^\top \operatorname{Jac}_v(x^*) z < 0 \quad \text{for all nonzero tangent vectors } z \text{ to } \mathcal{X} \text{ at } x^* \tag{SOS}$$

where $\operatorname{Jac}_v(x^*)$ denotes the Jacobian of v at x^*. In the context of saddle-point problems and continuous games, this condition has been studied extensively in the machine learning and control literatures, cf. [21, 31, 47, 56, 59] and references therein. Importantly, as we note below, (SOS) is a special case of (VS).

Proposition 1 (Hsieh et al., 2019, Lemma A.4)**.** *Let x^* be a critical point of $\mathcal{G}$ satisfying* (SOS)*. Then x^* is variationally stable.*

► **Finite games.** As a last example, let $\mathcal{G} = \Delta(\Gamma)$ be the mixed extension of a finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{A}, u)$. Since each player's payoff function $u_i(x_i; x_{-i})$ is linear in x_i, we readily get

$$\text{(NE)} \iff \text{(LNE)} \iff \text{(FOS)} \equiv \text{(SVI)} \tag{5}$$

In addition, we have the following characterization of stable states in finite games:

Proposition 2 (Mertikopoulos & Zhou, 2019, Prop. 5.2)**.** *A mixed strategy profile x^* is variationally stable if and only if it is a strict Nash equilibrium of Γ, i.e., if and only if* (NE) *holds as a strict inequality for all $x \neq x^*$.*

We will use all this freely in the sequel.

3. The learning framework

We now proceed to detail our online learning framework, beginning with the general model in Section 3.1 and continuing with a range of learning algorithms that can be seen as special cases thereof in Section 3.2. The reader interested only in the general theory can skip Section 3.2.

3.1. Regularized Robbins–Monro processes.

Our basic learning framework will hinge on the *regularized Robbins–Monro* template

$$Y_{n+1} = Y_n + \gamma_n \hat{v}_n \qquad X_{n+1} = Q(Y_{n+1}), \tag{RRM}$$

where:

(1) $X_n = (X_{i,n})_{i\in\mathcal{N}} \in \mathcal{X}$ denotes the players' action profile at each stage $n = 1, 2, \dots$

(2) $\hat{v}_n = (\hat{v}_{i,n})_{i\in\mathcal{N}} \in \mathcal{Y}$ is a sequence of individual "gradient-like" signals.

(3) $Y_n = (Y_{i,n})_{i\in\mathcal{N}} \in \mathcal{Y}$ is an auxiliary state variable aggregating individual gradient steps.

(4) $\gamma_n > 0$ is a step-size sequence, for which we will assume throughout that $\sum_n \gamma_n = \infty$ (typically the method is run with $\gamma_n \propto 1/n^{\ell_\gamma}$ for some $\ell_\gamma \geq 0$).

(5) $Q\colon \mathcal{Y} \to \mathcal{X}$ is a "generalized projection" map that mirrors gradient steps in $\mathcal{Y}$ to action updates in $\mathcal{X}$; we will refer to Q throughout as the players' *mirror map*.

As far as terminology is concerned, the term "*Robbins–Monro*" refers to the seminal stochastic approximation method of Robbins & Monro [57], while the adjective "regularized" alludes to the *"follow-the-regularized-leader"* (FTRL) family of algorithms of Shalev-Shwartz & Singer [63] – which, in turn, is intimately related to the *mirror descent* (MD) framework of Nemirovski & Yudin [49]. To streamline our presentation, we detail each of these elements below and defer a list of examples to Section 3.2.

▶ **The sequence of gradient signals.** To keep track of the sequence of events in (RRM), we will view X_n as a stochastic process on some complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we will write $\mathcal{F}_n \coloneqq \mathcal{F}(X_1, \dots, X_n) \subseteq \mathcal{F}$ for the history of play up to stage n (inclusive). Since we tacitly assume that $\hat{v}_n$ is generated *after* each player has selected an action at round n but *before* the $(n+1)$-th update has been triggered, we also posit that $\hat{v}_n$ is $\mathcal{F}_{n+1}$-measurable but not necessarily $\mathcal{F}_n$-measurable. In this way, we may decompose $\hat{v}_n$ as

$$\hat{v}_n = v(X_n) + U_n + b_n \tag{6}$$

where

$$U_n = \hat{v}_n - \mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] \quad \text{and} \quad b_n = \mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] - v(X_n). \tag{7}$$

Since $\mathbb{E}[U_n \,|\, \mathcal{F}_n] = 0$ by construction, U_n can be intepreted as a random, zero-mean error relative to $v(X_n)$; by contrast, b_n is $\mathcal{F}_n$-measurable, so it captures any systematic – and possibly *non-random* – offset of $\hat{v}_n$ relative to $v(X_n)$. We will quantify all this by assuming that b_n, U_n and $\hat{v}_n$ are bounded for some $q \geq 2$ as

$$\mathbb{E}[\|b_n\|_* \,|\, \mathcal{F}_n] \leq B_n \qquad \mathbb{E}[\|U_n\|_*^q \,|\, \mathcal{F}_n] \leq \sigma_n^q \qquad \text{and} \qquad \mathbb{E}[\|\hat{v}_n\|_*^q \,|\, \mathcal{F}_n] \leq M_n^q \tag{8}$$

where the sequences B_n, σ_n and M_n, $n = 1, 2, \dots$, are to be construed as deterministic upper bounds on the bias, fluctuations, and magnitude of $\hat{v}_n$ respectively (with the case $q = \infty$ taken to mean that the various quantites are bounded w.p.1). Accordingly, depending on these bounds, a gradient signal with $B_n = 0$ will be called *unbiased*, and an unbiased signal with $\sigma_n = 0$ will be called *perfect*.

Remark 1. We should stress here that $\hat{v}_n$ *should not* be interpreted narrowly as the output of a black-box oracle for $v(X_n)$, but as a "model-agnostic" surrogate thereof. In particular, the noise term U_n can model raw observational noise, but also inner randomizations of the algorithm; analogously, the bias term b_n is intended to capture situations where $\hat{v}_n$ results from actions other than X_n, the inclusion of corrective terms in a learning algorithm, etc. These modeling aspects are crucial to include in our analysis optimistic and extra-gradient methods; we explain this issue in detail in Section 3.2. ◊

Remark 2. By Assumption 1 and the inequality $(\sum_{i=1}^m a_i)^q \leq m^{q-1} \sum_{i=1}^m a_i^q$, the decomposition (6) of $\hat{v}_n$ shows that we can always pick $M_n^q = 3^{q-1}(G^q + B_n^q + \sigma_n^q)$ in (8). This makes the last part of (8) redundant, but we will maintain the explicit bound M_n for $\hat{v}_n$ to simplify the presentation. ◊

► **The players' mirror map.** The second defining element of (RRM) is the "mirror map" $Q_i\colon \mathcal{Y}_i \to \mathcal{X}_i$ of each player – or, in aggregate form, the product map $Q = (Q_i)_{i\in\mathcal{N}}\colon \mathcal{Y} \to \mathcal{X}$. This is defined by means of a "*regularizer*" on $\mathcal{X}$ as follows:[1]

Definition 1. We say that $h_i\colon \mathcal{V}_i \to \mathbb{R}\cup\{\infty\}$ is a *regularizer* on $\mathcal{X}_i$ if:

(1) h_i is *supported* on $\mathcal{X}_i$, i.e., $\operatorname{dom} h_i = \{x_i \in \mathcal{V}_i : h_i(x_i) < \infty\} = \mathcal{X}_i$.

(2) h_i is continuous and *strongly convex* on $\mathcal{X}_i$, i.e., there exists a constant $K_i > 0$ such that

$$h_i(\lambda x_i + (1-\lambda)x_i') \leq \lambda h_i(x_i) + (1-\lambda)h_i(x_i') - \tfrac{1}{2}K_i\lambda(1-\lambda)\|x_i' - x_i\|^2 \tag{9}$$

for all $x_i, x_i' \in \mathcal{X}_i$ and all $\lambda \in [0,1]$.

The *mirror map* associated to h_i is defined for all $y_i \in \mathcal{Y}_i$ as

$$Q_i(y_i) = \arg\max_{x_i\in\mathcal{X}_i}\{\langle y_i, x_i\rangle - h_i(x_i)\} \tag{10}$$

and the image $\mathcal{X}_{h_i} = \operatorname{im} Q_i$ of Q_i is called the *prox-domain* of h_i. Finally, we will also say that h_i is *steep* when $\operatorname{dom}\partial h_i = \operatorname{ri}\mathcal{X}_i$.

For concision, we will write $h(x) = \sum_i h_i(x_i)$ for the players' aggregate regularizer and $Q = (Q_i)_{i\in\mathcal{N}}$ for the induced mirror map. We provide three examples of this construction below:

Example 3.1 (Euclidean projection)**.** Consider the quadratic regularizer $h(x) = \|x\|_2^2/2$, $x \in \mathcal{X}$. Then the induced mirror map is the standard Euclidean projector

$$Q(y) = \Pi_{\mathcal{X}}(y) \equiv \arg\min_{x\in\mathcal{X}}\|y - x\|_2. \tag{11}$$

As a special case, in unconstrained settings (i.e., when $\mathcal{X} = \mathcal{V}$), we have $Q(y) = y$.

Example 3.2 (Entropic regularization on the simplex)**.** Let $\mathcal{A}_i$, $i = 1,\dots,N$, be an ensemble of pure strategies, set $\mathcal{X}_i = \Delta(\mathcal{A}_i)$, and let $h_i(x_i) = \sum_{\alpha_i\in\mathcal{A}_i} x_{i\alpha_i}\log x_{i\alpha_i}$ be the (negative) Gibbs–Shannon entropy on $\mathcal{X}_i$. By standard arguments, the resulting mirror map of each player $i \in \mathcal{N}$ is the logit choice map

$$Q_i(y_i) = \Lambda_i(y_i) \equiv \frac{(\exp(y_{i\alpha_i}))_{\alpha_i\in\mathcal{A}_i}}{\sum_{\alpha_i\in\mathcal{A}_i}\exp(y_{i\alpha_i})} \tag{12}$$

This choice map plays a central role in finite games; we will revisit it several times in Section 3.2.

Example 3.3 (Regularization on the orthant)**.** Let $\mathcal{X}_i = [0,\infty)$ and set $h_i(x_i) = x_i\log x_i - x_i$ for all $x_i \in \mathcal{X}_i$, $i \in \mathcal{N}$. By a straightforward calculation, the induced mirror map is $Q_i(y_i) = \exp(y_i)$. As we discuss in Section 3.2, this provides the relevant setup for games with half-space constraints.[2]

3.2. **Specific algorithms.** We now proceed to describe a representative range of learning algorithms that can be incorporated as specific instances of the general framework (RRM). Depending on the information available to the players, we classify the algorithms under study as *oracle-based* or *payoff-based*.

[1] The authors thank S. Sorin for proposing this definition.

[2] Strictly speaking, the regularizer $x\log x - x$ is not strongly convex over $\mathbb{R}_+$ but it *is* strongly convex over any bounded subset of $\mathbb{R}_+$ – and it can be made strongly convex over all of $\mathbb{R}_+$ by adding a small quadratic penalty of the form $\varepsilon x^2/2$. This issue does not change the essence of our results, so we sidestep the details.

► **Oracle-based methods.** In the first batch of methods under consideration, the players are assumed to have access to a *stochastic first-order oracle* (SFO), that is, a "black-box" feedback mechanism that returns an estimate of their individual payoff gradients at the chosen action profile. Formally, when queried at $x \in \mathcal{X}$, an SFO outputs a random vector of the form

$$V(x;\theta) = v(x) + \mathrm{err}(x;\theta), \tag{SFO}$$

where θ is a random variable taking values in some measurable space Θ and $\mathrm{err}(x;\theta)$ is an umbrella error term capturing all sources of uncertainty in the model. In practice, (SFO) is queried repeatedly at a sequence of action profiles $X_n \in \mathcal{X}$, $n = 1, 2, \dots$, possibly with a different random seed θ_n each time.[3] For concreteness, we will assume that the noise in (SFO) is zero-mean and bounded in L^q for some $q \geq 2$, i.e.,

$$\mathbb{E}[\mathrm{err}(x;\theta_n) \,|\, \mathcal{F}_n] = 0 \quad \text{and} \quad \mathbb{E}[\|\mathrm{err}(x;\theta_n)\|_*^q \,|\, \mathcal{F}_n] \leq \sigma^q \tag{13}$$

for some $\sigma \geq 0$ and all $x \in \mathcal{X}$ (with $q = \infty$ taken to mean that $\mathrm{err}(x;\theta)$ is bounded w.p.1).

We are now in a position to introduce the array of *oracle-based* methods under study; to lighten notation, we present some of these policies in an unconstrained setting.

Algorithm 1 (Stochastic gradient ascent)**.** Perhaps the most basic iterative policy for multi-agent online learning is the standard (individual) gradient ascent method

$$X_{n+1} = X_n + \gamma_n V(X_n;\theta_n) \tag{SGA}$$

with θ_n drawn i.i.d. from Θ. From a loss minimization viewpoint, (SGA) is a multi-agent analogue of the standard stochastic gradient descent algorithm; in min-max games, (SGA) is sometimes referred to as the Arrow–Hurwicz method [1]. Clearly, (SGA) is immediately recovered from (RRM) if the latter is run with the sequence of gradient signals $\hat{v}_n \leftarrow V(X_n;\theta_n)$ and the trivial mirror map $Q(y) = y$. ◊

Algorithm 2 (Extra-gradient)**.** Going a step further from (SGA), the (stochastic) *extra-gradient* (EG) algorithm of Korpelevich [36] is based on the following principle: starting at some "base" state X_n, the players first take a gradient step to an interim, "leading" state $X_{n+1/2}$; subsequently, to anticipate their payoff landscape, they update the base state X_n with gradient information from $X_{n+1/2}$ instead of X_n, and the process repeats. Formally, this leads to the policy

$$\begin{aligned} X_{n+1/2} &= X_n + \gamma_n V(X_n;\theta_n) \\ X_{n+1} &= X_n + \gamma_n V(X_{n+1/2};\theta_{n+1/2}) \end{aligned} \tag{EG}$$

with $\theta_n, \theta_{n+1/2}$ drawn i.i.d. from Θ. Accordingly, (EG) is readily recovered from (RRM) by taking $\hat{v}_n \leftarrow V(X_{n+1/2};\theta_{n+1/2})$. ◊

Algorithm 3 (Optimistic gradient)**.** A computational drawback of (EG) is that it requires two oracle queries per update – and hence, more overhead per iteration. One way to overcome this hurdle is to reuse past gradient information in the hope that it provides a good enough approximation of the present; this leads to the *optimistic gradient* policy

$$\begin{aligned} X_{n+1/2} &= X_n + \gamma_n V(X_{n-1/2};\theta_{n-1}) \\ X_{n+1} &= X_n + \gamma_n V(X_{n+1/2};\theta_n) \end{aligned} \tag{OG}$$

Similarly to (EG), (OG) is recovered from (RRM) by setting $\hat{v}_n \leftarrow V(X_{n+1/2};\theta_n)$. This "gradient reuse" idea goes back at least to Popov [54], and it has resurfaced several times in the literature since then, cf. [14, 22, 31, 55] and references therein. To simplify our

[3] In some cases, the index set may be enlarged to include all positive half-integers ($n = 1/2, 1, 3/2, \dots$).

presentation, we will assume in the sequel that (OG) is run with an SFO satisfying (13) with $q = \infty$. ◊

The next method concerns learning in mixed extensions of finite games.

Algorithm 4 (Exponential / multiplicative weights)**.** Let $\mathcal{G} = \Delta(\Gamma)$ be the mixed extension of a finite game $\Gamma(\mathcal{N}, \mathcal{A}, u)$ as per Example 2.3. In this setting, the players' learning process typically unfolds as follows: at each stage $n = 1, 2, \dots$, every player selects a mixed strategy $X_{i,n} \in \Delta(\mathcal{A}_i)$ and draws a pure strategy $\alpha_{i,n} \in \mathcal{A}_i$ according to $X_{i,n}$. Then, depending on the amount of information available to the players, we have the following oracle models:

(1) *Full information feedback:* in this case, players observe their *mixed* payoff vectors, i.e.,

$$V_i(X_n; \alpha_n) = v_i(X_{i,n}; X_{-i,n}). \tag{14a}$$

(2) *Realization-based feedback:* here, players instead observe their *pure* payoff vectors, i.e.,

$$V_i(X_n; \alpha_n) = v_i(\alpha_{i,n}; \alpha_{-i,n}). \tag{14b}$$

Both models can be seen as SFOs with seed α_n, i.e., the (pure) action profile chosen by the players at stage n; the oracle (14a) is deterministic, while the oracle (14b) is stochastic and satisfies (13) with $q = \infty$.

In this context, one of the most widely used learning methods is the so-called *exponential / multiplicative weights* algorithm – or HEDGE – which unfolds iteratively as

$$\begin{aligned} Y_{i,n+1} &= Y_{i,n} + \gamma_n V_i(X_n; \alpha_n) \\ X_{i,n+1} &= \Lambda_i(Y_{i,n+1}) \end{aligned} \tag{HEDGE}$$

with Λ_i denoting the logit choice map of (12) and V_i given by (14a) or (14b) depending on the information available to the players. In both cases, (HEDGE) is recovered immediately from (RRM) by letting $\hat{v}_n \leftarrow V(X_n; \alpha_n)$ and $Q_i = \Lambda_i$. For an overview of the method's history and its applications, see [11, 38] and references therein. ◊

► **Payoff-based methods.** Moving forward, it is important to recall that Algorithms 1–4 all assume that players have access to a black-box oracle mechanism, but do not specify how this could be achieved in practice. Albeit commonplace, this assumption is not realistic in many applications where players may only be able to observe their realized payoffs and have no information about the strategies of other players or actions they did not play. To bridge this disconnect, we describe below a range of *payoff-based* policies where players estimate their individual payoff gradients indirectly, from their realized, "in-game" payoffs.

Algorithm 5 (Single-point stochastic approximation)**.** A straightforward way of reconstructing gradients from zeroth-order feedback is via the *single-point stochastic approximation* framework of Spall [64]. In the unconstrained case ($\mathcal{X} = \mathcal{V}$), the relevant update step is:

$$\begin{aligned} \hat{X}_{i,n} &= X_{i,n} + \delta_n W_{i,n}, \\ X_{i,n+1} &= X_{i,n} + \gamma_n (u_i(\hat{X}_n)/\delta_n)\, W_{i,n}. \end{aligned} \tag{SPSA}$$

In (SPSA), each player's "query state" $\hat{X}_{i,n}$, $i = 1, \dots, N$, is a perturbation of the "base state" $X_{i,n}$ by a step of magnitude $\delta_n > 0$ along a random direction $W_{i,n}$ drawn from the ensemble of signed basis vectors $\mathcal{E}_i \coloneqq \{(\pm e_1, \dots, \pm e_{d_i})\}$. In this manner, (SPSA) can be seen as a special case of (RRM) with $\hat{v}_{i,n} \leftarrow (u_i(\hat{X}_n)/\delta_n)\, W_{i,n}$ for all $i \in \mathcal{N}$.[4] ◊

[4]This formulation of (SPSA) is tailored to unconstrained problems. In this case, to ensure that the resulting gradient estimator remains bounded, it is customary to include an indicator of the form $\mathbb{1}(\|\hat{X}_n\| \leq R_n)$ for some suitably chosen sequence $R_n \to \infty$ [64]. This would lead to the same analysis but at the cost of heavier

Algorithm 6 (Dampened gradient approximation)**.** An alternative approach to (SPSA) is the two-point, "explore-then-update" approach of Bervoets et al. [8] who focused on games with $\mathcal{X}_i = [0, \infty)$ for all $i \in \mathcal{N}$ and introduced the *dampened gradient approximation* policy

$$\begin{aligned} X_{i,n+1/2} &= X_{i,n} + (1/n)W_{i,n} \\ X_{i,n+1} &= X_{i,n}[1 + (u_i(X_{n+1/2}) - u_i(X_n))W_{i,n}] \end{aligned} \tag{DGA}$$

In the above, the "exploration direction" $W_{i,n}$ is sampled uniformly at random from $\{\pm 1\}$ at each n. In other words, (DGA) is a two-stage process where players first "explore" their individual payoff functions at a nearby state, and then use this information to estimate their individual payoff gradients and update their base state.

To include (DGA) in the framework of (RRM), take $Q_i(y_i) = \exp(y_i)$ as per Example 3.3. Then, letting $Y_n = \log X_n$, we get

$$Y_{n+1} = Y_n + \log(1 + (u_i(X_{n+1/2}) - u_i(X_n))W_{i,n}). \tag{15}$$

We may therefore view (DGA) as an instance of (RRM) with $\gamma_n = 1/n$ and gradient signals given by $\hat{v}_{i,n} \leftarrow n \cdot \log(1 + (u_i(X_{n+1/2}) - u_i(X_n))W_{i,n})$. ◊

Algorithm 7 (The EXP3 algorithm)**.** In our final example, we return to finite games, and we focus on the "bandit" case where players only observe the payoffs of the pure strategies that they played. In this setting, it is common to employ the *importance-weighted estimator*

$$V_{i\alpha_i}(\hat{X}_n; \hat{\alpha}_n) = \frac{\mathbb{1}(\hat{\alpha}_{i,n} = \alpha_i)}{\hat{X}_{i\alpha_i,n}} u_i(\hat{\alpha}_{i,n}; \hat{\alpha}_{-i,n}) \qquad \text{for all } \alpha_i \in \mathcal{A}_i,\ i \in \mathcal{N}, \tag{IWE}$$

where each player $i \in \mathcal{N}$ draws an action $\hat{\alpha}_{i,n}$ from $\mathcal{A}_i$ according to a mixed strategy $\hat{X}_{i,n} \in \Delta(\mathcal{A}_i)$. Then, plugging (IWE) into (Hedge), we obtain the method known as *exponential weights for exploration and exploitation* (EXP3), viz.

$$\begin{aligned} Y_{i,n+1} &= Y_{i,n} + \gamma_n V_i(\hat{X}_n; \hat{\alpha}_n), \\ X_{i,n+1} &= \Lambda_i(Y_{i,n+1}), \end{aligned} \tag{EXP3}$$

where the sampling strategy of the i-th player at stage n is given by

$$\hat{X}_{i,n} = (1 - \delta_n)X_{i,n} + \delta_n \operatorname{unif}(\mathcal{A}_i). \tag{16}$$

In the above, $\delta_n \geq 0$ is an "explicit exploration" parameter that determines the mixing between $X_{i,n}$ and the uniform distribution $\operatorname{unif}(\mathcal{A}_i)$ on $\mathcal{A}_i$. Accordingly, (EXP3) can be seen as an instance of (RRM) with $Q_i = \Lambda_i$ and $\hat{v}_n$ given by (IWE) with pure strategies $\hat{\alpha}_n$ drawn according to $\hat{X}_n$. ◊

► **Runtime parameters.** The above justifies the characterization of (RRM) as a "parent scheme" for Algorithms 1–7. In particular, thanks to the explicit expressions for $\hat{v}_n$ derived in each case, we can likewise estimate the error bounds B_n, σ_n and M_n of each method.

Proposition 3. *Suppose that Algorithms 1–7 are run with step-size $\gamma_n \propto 1/n^{\ell_\gamma}$, $\ell_\gamma \in [0, 1]$, and, where applicable, a sampling parameter $\delta_n \propto 1/n^{\ell_\delta}$, $\ell_\delta \in (0, 1/2)$. Then the corresponding sequence of gradient signals $\hat{v}_n$ in (RRM) enjoys the bounds:*

- *For Algorithms 1 and 4: $B_n = 0$, $\sigma_n = \mathcal{O}(1)$, and $M_n = \mathcal{O}(1)$.*
- *For Algorithms 2 and 3: $B_n = \mathcal{O}(1/n^{\ell_\gamma})$, $\sigma_n = \mathcal{O}(1)$, and $M_n = \mathcal{O}(1)$.*
- *For Algorithms 5 and 7: $B_n = \mathcal{O}(1/n^{\ell_\delta})$, $\sigma_n = \mathcal{O}(n^{\ell_\delta})$, and $M_n = \mathcal{O}(n^{\ell_\delta})$.*

notation so, instead, we will assume that the players' payoff functions are bounded when discussing (SPSA). For a detailed discussion of how to adapt (SPSA) in the presence of constraints, we refer the reader to Bravo et al. [9] who show that the relevant entries of Table 1 apply verbatim when $\mathcal{X}$ is compact.

Algorithm	Actions ($\mathcal{X}_i$)	Mirror Map (Q)	Feedback	Bias (B_n)	Magnitude (M_n)
(SGA)	$\mathbb{R}^{d_i}$	y	oracle	0	$\mathcal{O}(1)$
(EG) / (OG)	$\mathbb{R}^{d_i}$	y	oracle	$\mathcal{O}(1/n^{\ell_\gamma})$	$\mathcal{O}(1)$
(Hedge)	$\Delta(\mathcal{A}_i)$	$\Lambda(y)$	oracle	0	$\mathcal{O}(1)$
(SPSA)	$\mathbb{R}^{d_i}$	y	payoff	$\mathcal{O}(1/n^{\ell_\delta})$	$\mathcal{O}(n^{\ell_\delta})$
(DGA)	$[0,\infty)$	$\exp(y)$	payoff	$\mathcal{O}(1/n)$	$\mathcal{O}(1)$
(EXP3)	$\Delta(\mathcal{A}_i)$	$\Lambda(y)$	payoff	$\mathcal{O}(1/n^{\ell_\delta})$	$\mathcal{O}(n^{\ell_\delta})$

Table 1: The algorithms of Section 3.2 as instances of (RRM). Where applicable, the methods' step-size and sampling parameters are taken to be of the form $\gamma_n \propto 1/n^{\ell_\gamma}$ and $\delta_n \propto 1/n^{\ell_\delta}$ for some $\ell_\gamma \in [0,1]$ and $\ell_\delta \in (0,1/2)$ respectively.

- *For Algorithm 6:* $B_n = \mathcal{O}(1/n)$, $\sigma_n = \mathcal{O}(1)$, *and* $M_n = \mathcal{O}(1)$.

For ease of reference, we summarize the above in Table 1 (for the proof of Proposition 3, see Appendix B). Of course, we can also "mix'n'match" different methods to include other algorithms considered in the literature: for instance, coupling (SPSA) with a general mirror map leads to the bandit mirror descent algorithm of Bravo et al. [9]; incorporating the gradient reuse step of (OG) in the setup of (Hedge) yields the optimistic multiplicative weights (OMW) method of Daskalakis & Panageas [13]; etc. In the sections to come, we will exploit the expressive power of (RRM) to provide a synthetic analysis for all these policies.

4. Stochastic approximation and first results

4.1. **Mean dynamics and stochastic approximation.** In this section, we derive a series of convergence results for (RRM) by treating it as a "noisy" discretization of the *mean dynamics*

$$\dot{y} = v(x) \qquad x = Q(y). \tag{MD}$$

In this continuous-time interpretation, $\dot{y}$ represents the limit of the finite difference quotient $(Y_{n+1} - Y_n)/\gamma_n$. As such, if γ_n is "sufficiently small" and the gradient signal $\hat{v}_n$ is a "good enough" approximation of $v(X_n)$, it is plausible to expect that the iterates of (RRM) and the solutions of (MD) will eventually come together.

Following [6, 7], this heuristic can be made precise as follows: First, let $\psi\colon \mathbb{R} \times \mathcal{Y} \to \mathcal{Y}$ denote the *flow* associated to (MD), i.e., the map which sends an initial condition $y \in \mathcal{Y}$ to the point $\psi_t(y) \in \mathcal{Y}$ obtained by following the orbit of (MD) starting at y for time $t \in \mathbb{R}$. Then, to compare the sequence of iterates Y_n generated by (RRM) with the solution orbits of (MD), define the *effective time* $\tau_n = \sum_{k=1}^{n} \gamma_k$ and the associated *affine interpolation* $Y(t)$ of Y_n as

$$Y(t) = Y_n + \frac{t - \tau_n}{\tau_{n+1} - \tau_n}(Y_{n+1} - Y_n) \quad \text{for all } t \in [\tau_n, \tau_{n+1}],\ n = 1, 2, \dots \tag{17}$$

We then have the following notion of "asymptotic closeness" between (RRM) and (MD):

Definition 2 (Benaïm, 1999)**.** The sequence Y_n is an *asymptotic pseudotrajectory* (APT) of (MD) if

$$\lim_{t\to\infty} \sup_{0\le s\le T} \|Y(t+s) - \psi_s(Y(t))\|_* = 0 \quad \text{for all } T > 0. \tag{APT}$$

In words, Definition 2 posits that Y_n asymptotically tracks the orbits $y(t)$ of (MD) with arbitrary precision over windows of arbitrary length. In our setting, the following proposition can be used as an explicit criterion guaranteeing this property:

Proposition 4. *Suppose that* (RRM) *is run with step-size and gradient signal sequences such that a)* $\gamma_n \to 0$*; b)* $B_n \to 0$*; and c)* $\sum_n \gamma_n^{1+q/2} M_n^q < \infty$*. Then, with probability* 1*, the sequence* $X_n = Q(Y_n)$ *is an APT of* (MD)*.*

Proposition 4 shadows a basic result of Benaïm [6, Prop. 4.1, cf. Eq. (13) and onwards] so we omit its proof. For our purposes, it is more important to note that a tandem application of Propositions 3 and 4 immediately yields the following concrete conditions for Algorithms 1–7:

Corollary 1. *Suppose that Algorithms 1–7 are run with parameters as in Proposition 3. Then the sequence* Y_n *comprises an APT of* (MD) *provided that:*

- *For Algorithms 1–3:* $\ell_\gamma > 2/(2+q)$
- *For Algorithm 4:* $\ell_\gamma > 0$
- *For Algorithms 5 and 7:* $\ell_\gamma > 2\ell_\delta > 0$

Corollary 1 provides a minimal set of hypotheses under which (MD) is a faithful representation of Algorithms 1–7. For some of these algorithms, this property is already known in the literature, see e.g., [6] for (SGA) and [8] for (DGA). For others, the link with (MD) appears to be new: especially in the case of (EG) / (OG), Corollary 1 settles a standing question in the literature concerning the mean dynamics of optimistic gradient methods.

4.2. **The primal-dual dichotomy.** To proceed, we will need some basic definitions from the theory of dynamical systems. Specifically, given a flow $\phi\colon \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ on some metric space $\mathcal{M}$ and a nonempty compact subset $\mathcal{S}$ of $\mathcal{M}$, we say that:

(1) $\mathcal{S}$ is *invariant* under ϕ if $\phi_t(\mathcal{S}) = \mathcal{S}$ for all $t \in \mathbb{R}$.

(2) $\mathcal{S}$ is an *attractor* of ϕ if it admits a neighborhood $\mathcal{W} \subseteq \mathcal{Y}$ such that $\operatorname{dist}(\phi_t(y), \mathcal{S}) \to 0$ uniformly in $y \in \mathcal{W}$ as $t \to \infty$.

(3) $\mathcal{S}$ is *internally chain transitive* (ICT) if it is invariant and $\phi|_{\mathcal{S}}$ has no attractors except $\mathcal{S}$.

With all this in hand, the general theory of Benaïm & Hirsch [7] yields the following convergence result when applied to (MD):

Theorem 1 (Benaïm & Hirsch, 1996)**.** *Let* Y_n*,* $n = 1, 2, \dots$*, be an APT of* (MD) *with* $\sup_n \|Y_n\|_* < \infty$*. Then* Y_n *converges to an ICT set of* (MD)*.*

Proof. Lemma A.1 in Appendix A shows that Q is Lipschitz continuous. Since v is also Lipschitz continuous by Assumption 1, our assertion follows from Benaïm & Hirsch [7, Theorem 0.1]. ∎

Taken together, Theorem 1 and Corollary 1 suggest that the behavior of the various algorithms presented in Section 3.2 (and many more) can be understood by looking at the ICT sets of the *same* mean dynamics. However, from a practical viewpoint, this conclusion carries two important limitations: First, the boundedness caveat for Y_n cannot be readily checked against the game's primitives, so it is not clear when Theorem 1 applies – and, in much of the literature, this assumption has persisted as a condition that needs to be enforced "by hand" [6, 37]. Second – and perhaps more importantly – this reasoning ignores the fact that X_n evolves in $\mathcal{X}$, the game's *action space*, whereas the orbits of (MD) live in $\mathcal{Y}$, the game's *dual space*. In turn, this leads to a fundamental mismatch: a dual orbit $y(t)$ may *diverge* in $\mathcal{Y}$, even though the induced primal orbit $x(t) = Q(y(t))$ *converges* in $\mathcal{X}$.

Example. Consider the single-player game with $u(x) = -x$, $x \geq 0$. Then the dynamics (MD) give $\dot{y} = -1$, so $y(t) \to -\infty$ as $t \to \infty$, i.e., $y(t)$ *diverges*; however, under the exponential mirror map of Example 3.3, the player's trajectory of actions evolves as $x(t) = \exp(y(t))$, i.e., $x(t)$ *converges* (to 0). In this case, even if we were to ignore the boundedness issue, Theorem 1 becomes vacuous (and, in a sense, misleading): the dynamics (MD) do not have any ICT sets and are divergent, even though the induced trajectory of actions converges in $\mathcal{X}$. ◊

The above creates a relatively awkward situation in which *dynamical* notions of stationarity and stability are defined on $\mathcal{Y}$, whereas the corresponding *game-theoretic* notions reside in $\mathcal{X}$. To reconcile this incompatibility, it is instead natural to focus directly on $\mathcal{X}$ and ask whether the notion of an ICT set can be transposed there. However, this is only meaningful if the ensemble of trajectories $x(t) = Q(y(t))$ constitute a flow on $\mathcal{X}$; that is, formally, we must posit the existence of a flow χ on $\mathcal{X}$ (or a subset thereof) that is *conjugate* to ψ in the sense that $Q \circ \psi_t = \chi_t \circ Q$ for all $t \in \mathbb{R}$.

In general, this may fail to hold: for example, consider the single-player game $u(x) = x$, $x \in [0,1]$, and consider the Euclidean projector $Q(y) = [y]_0^1 \equiv \min\{\max\{y, 0\}, 1\}$ induced by the quadratic regularizer $h(x) = x^2/2$ on $\mathcal{X} = [0,1]$. Clearly, (MD) gives $\psi_t(y) = y + t$ for all $t \in \mathbb{R}$ and all $y \in \mathbb{R}$. Nevertheless, even though the orbits $\psi_t(0)$ and $\psi_t(-1)$ both have the same starting point $Q(0) = Q(-1) = 0$ in $\mathcal{X}$, the induced primal trajectories evolve differently: for all $t \in [0,1]$, we have $Q(\psi_t(0)) = t$ and $Q(\psi_t(-1)) = 0$, implying in turn that there can be no flow χ on $\mathcal{X}$ whose orbits are the images of the orbits of (MD).

Because this discrepancy arises at the boundary of $\mathcal{X}$, Theorem 1 is more relevant for cases where all orbits are contained in $\operatorname{ri}\mathcal{X}$. In this case, we have the following result.

Proposition 5. *Let $\mathcal{D} \subseteq \mathcal{Y}$ be a forward-invariant set of ψ such that $\mathcal{B} = Q(\mathcal{D})$ is contained in $\operatorname{ri}\mathcal{X}$. Then there exists a flow χ on $\mathcal{B}$ such that $\chi_t(Q(y)) = Q(\psi_t(y))$ for all $y \in \mathcal{D}$ and all $t \geq 0$; in particular, χ can be defined on all of $\operatorname{ri}\mathcal{X}$ if $\operatorname{im} Q = \operatorname{ri}\mathcal{X}$.*

Proof. Clearly, the only candidate for χ is to set $\chi_t(x) = Q(\psi_t(y))$ whenever $x = Q(y)$ for $y \in \mathcal{D}$. To see that this construction is well-defined, suppose that $Q(y') = Q(y) = x$ for some $y, y' \in \mathcal{D}$, and let $w = y' - y$; then, it suffices to show that $Q(\psi_t(y)) = Q(\psi_t(y + w))$ for all $t \geq 0$.

As a first step, we claim that $Q(\psi_t(y) + w) = Q(\psi_t(y))$ for all $t \geq 0$. Indeed, since $Q(y) \in \operatorname{ri}\mathcal{X}$, Lemma A.1 in Appendix A shows that w annihilates all tangent directions to $\mathcal{X}$ at x, i.e., $\langle w, x' - x\rangle = 0$ for all $x' \in \mathcal{X}$. However, since $\psi_t(\mathcal{D}) \subseteq \mathcal{D}$ and $Q(\mathcal{D}) \subseteq \operatorname{ri}\mathcal{X}$, the point $x_t = Q(\psi_t(y))$ will also be interior; hence, with $\psi_t(y) \in \partial h(x_t)$ and $\langle w, x' - x_t\rangle = \langle w, x' - x\rangle + \langle w, x - x_t\rangle = 0$ for all $x' \in \mathcal{X}$, invoking Lemma A.1 in the converse direction gives $\psi_t(y) + w \in \partial h(x_t)$, i.e., $Q(\psi_t(y) + w) = Q(\psi_t(y))$.

In view of the above, a simple differentiation yields

$$\frac{d}{dt}[\psi_t(y) + w] = \frac{d}{dt}\psi_t(y) = v(Q(\psi_t(y))) = v(Q(\psi_t(y) + w)) \tag{18}$$

which, by the Picard-Lindelöf theorem, shows that $\psi_t(y) + w$ is the (necessarily unique) solution orbit of (MD) starting at $y + w$ at time $t = 0$. We thus conclude that $Q(\psi_t(y+w)) = Q(\psi_t(y) + w) = Q(\psi_t(y))$, i.e., χ is well-defined. Our second assertion follows from the fact that $\operatorname{im} Q = \operatorname{dom}\partial h = \operatorname{ri}\mathcal{X}$, so we can apply our first claim to $\mathcal{D} = \mathcal{Y}$. ∎

Proposition 5 indicates that, if Q is interior-valued, we can map the flow on $\mathcal{Y}$ to an induced primal flow on $\mathcal{X}$. For concreteness, before discussing the precise connection between (RRM) and the induced dynamics on $\mathcal{X}$, we illustrate the latter in the case of Examples 3.1–3.3:

Example 4.1. Take $\mathcal{X} = \mathcal{V}$ and $Q(y) = y$ as in Example 3.1. Then (MD) trivially gives the (individual) *gradient dynamics*

$$\dot{x} = v(x) \tag{GD}$$

Example 4.2. Let $\mathcal{X}_i = \Delta(\mathcal{A}_i)$ and take $Q_i = \Lambda_i$ as in Example 3.2. Then, by a standard calculation, (MD) boils down to the *replicator dynamics* of Taylor & Jonker [68]

$$\dot{x}_{i\alpha_i} = x_{i\alpha_i}[u_i(\alpha_i; x_{-i}) - u_i(x_i; x_{-i})]. \tag{RD}$$

Example 4.3. Let $\mathcal{X}_i = [0, \infty)$ and take $Q_i(y_i) = \exp(y_i)$ as in Example 3.3. Then, by differentiating $x_i = e^{y_i}$ we obtain the *dampened gradient dynamics* of Bervoets et al. [8], viz.

$$\dot{x}_i = x_i v_i(x). \tag{DGD}$$

4.3. **Subcoercivity and convergence.** Other than the primal-dual dichotomy described above, the other important caveat in Theorem 1 is the boundedness of the generated sequence Y_n. In the optimization literature, a standard way to establish this type of control on an algorithm designed to find a zero of a vector field v on $\mathbb{R}^d$ is to assume that it is *coercive*, i.e., $\lim_{\|x\|\to\infty} \langle v(x), x\rangle/\|x\| = -\infty$. Intuitively, coercivity means that $v(x)$ is strongly "inward pointing" for large values of x, so it acts as a natural barrier against escape phenomena; at the same time, it also imposes that $\|v(x)\|_*$ grows superlinearly in x, and is only relevant for unconstrained problems, so it is not particularly well-suited for our purposes. Instead, we will consider the following, weaker requirement:

Definition 3. We say that $\mathcal{G}$ is *subcoercive* if there exists a compact set $\mathcal{K} \in \operatorname{ri}\mathcal{X}$ and a reference point $p \in \mathcal{K}$ such that

$$\langle v(x), x - p\rangle \leq 0 \quad \text{for all } x \in \mathcal{X} \setminus \mathcal{K}. \tag{SC}$$

Geometrically, subcoercivity simply posits that the Nash field $v(x)$ of the game points weakly towards p outside $\mathcal{K}$, so any "attracting" behavior in $\mathcal{G}$ must be contained in $\mathcal{K}$: for example, it is straightforward to verify that any variationally stable state of $\mathcal{G}$ must lie within $\mathcal{K}$ if (SC) holds. Beyond this, it is important to note that v may vanish at infinity, and $\mathcal{K}$ can be arbitrarily large (relative to $\mathcal{X}$). We provide some examples below:

Example 4.4 (Potential games)**.** Suppose that $\mathcal{G}$ admits a quasiconcave potential Φ with $\arg\max\Phi \subseteq \operatorname{ri}\mathcal{X}$. If we fix a maximizer p of Φ, we have $\langle v(x), x - p\rangle = \langle \nabla\Phi(x), x - p\rangle \leq 0$ for all $x \in \mathcal{X}$, so $\mathcal{G}$ is subcoercive. More generally, $\mathcal{G}$ is subcoercive whenever Φ is "eventually quasiconcave", i.e., the upper level sets $L_c^+(\Phi) = \{x \in \mathcal{X} : \Phi(x) \geq c\}$ of Φ are convex for sufficiently small $c > \inf\Phi$ and at least one such set is contained in $\operatorname{ri}\mathcal{X}$.[5] ◊

Example 4.5 (Min-max games)**.** Consider the toy game $\min_{x_1\in[-1,1]}\max_{x_2\in[-1,1]} x_1x_2$. Since $\langle v(x), x\rangle = -x_2x_1 + x_1x_2 = 0$ for all $x \in [-1,1]\times[-1,1]$, the game is trivially subcoercive. More generally, it is easy to check that any two-player, quasi-convex / quasi-concave game with an interior equilibrium is subcoercive. ◊

By itself, subcoercivity ensures that there is no consistent drift pointing away from $\mathcal{K}$, so it is reasonable to expect that Y_n does not escape to infinity either. To control the inherent stochasticity in Y_n and make this intuition precise, we will require the following summability conditions on the bias, variance, and magnitude of the gradient signal process $\hat{v}_n$:

$$\textstyle\sum_n \gamma_n B_n < \infty \qquad \sum_n \gamma_n^2\sigma_n^2 < \infty \qquad \text{and} \qquad \sum_n \gamma_n^2 M_n^2 < \infty. \tag{Sum}$$

[5] To see this, let $\mathcal{K} = L_{c_0}^+(\Phi)$ be a convex upper level set of Φ in $\operatorname{ri}\mathcal{X}$. Then, for all $c \leq c_0$ and all x with $\Phi(x) = c$, the segment $x + \tau(p - x)$, $\tau \in [0,1]$, is contained in $L_c^+(\Phi) \supseteq L_{c_0}^+(\Phi)$, so the function $\phi(\tau) = \Phi(x + \tau(p - x))$ cannot have $\phi'(0) < 0$. This implies that $0 \leq \langle \nabla\Phi(x), p - x\rangle = \langle v(x), p - x\rangle$ for all $x \in \mathcal{X} \setminus \mathcal{K}$, i.e., $\mathcal{G}$ is subcoercive.

Under these conditions, we have the following stability result.

Proposition 6. *Suppose that* (RRM) *is run with step-size and gradient signal sequences satisfying* (Sum)*. If $\mathcal{G}$ is subcoercive, the sequence of iterates Y_n generated by* (RRM) *is bounded w.p.*1*.*

Before proving Proposition 6, it is important to note that subcoercivity only concerns the primitives of the game under study, and it is otherwise "algorithm-agnostic". In this regard, given the primal-dual nature of the underlying dynamics (MD), Proposition 6 plays a major role in enabling the use of stochastic approximation tools and techniques (otherwise, the boundedness of $\mathcal{X}$ by itself does not suffice).

Moreover, from an operational viewpoint, Proposition 3 makes the verification of (Sum) a trivial affair for the algorithms under study. In particular, a joint application of Corollary 1, Propositions 3 and 6, and Theorem 1 readily yields the general convergence result below:

Theorem 2. *Suppose that Algorithms 1–7 are run with step-size $\gamma_n \propto 1/n^{\ell_\gamma}$, $\ell_\gamma \in (1/2, 1]$, and, where applicable, a sampling parameter $\delta_n \propto 1/n^{\ell_\delta}$ with $1 - \ell_\gamma < \ell_\delta < \ell_\gamma - 1/2$. If $\mathcal{G}$ is subcoercive, then: a) Y_n converges to an ICT set of* (MD) *with probability* 1*; and, in addition, b) if the players' mirror map Q is interior-valued, the induced sequence of play $X_n = Q(Y_n)$ converges with probability* 1 *to an ICT set of the primal flow χ on $\mathcal{X}$.*

We then have the following consequences for potential and zero-sum games (both stated for simplicity under the assumption that Q is interior-valued):

Corollary 2. *If $\mathcal{G}$ admits a subcoercive potential, X_n converges to a component of critical points of $\mathcal{G}$ w.p.*1*. In particular, if the potential is concave, X_n converges to the set of Nash equilibria of $\mathcal{G}$.*

Corollary 3. *Suppose that $\mathcal{G}$ is a strictly convex-concave min-max game with an interior equilibrium $x^* \in \operatorname{ri}\mathcal{X}$. Then X_n converges to x^* w.p.*1*.*

Corollaries 2 and 3 follow respectively from the fact that the only ICT sets of potential games and strictly convex-concave games are their sets of critical points, see e.g., [6, 44] and references therein. As for Theorem 2 (which we prove below), it should not be viewed as an equilibrium convergence guarantee, but as a characterization of what types of behaviors may arise in the limit of a game-theoretic learning process – equilibrium and non-equilibrium alike. However, because of the subcoercivity requirement, this characterization only extends to limit sets that are contained in the relative interior of the players' action spaces; games with boundary solutions require a different treatment, which we undertake in the next two sections.

4.4. **Technical proofs.** We conclude this section with the proof of Proposition 6 and Theorem 2; we begin with the latter, which is more conceptual and less technical.

Proof of Theorem 2. By Propositions 3 and 6, Y_n is a bounded APT of (MD), so the first part of the theorem follows directly from Theorem 1. As for the second, since Q is Lipschitz continuous (cf. Lemma A.1 in Appendix A), the sequence $X_n = Q(Y_n)$ is also bounded and, in addition, we have:

$$\begin{aligned}\|X(t+s) - \chi_s(X(t))\| &= \|Q(Y(t+s)) - \chi_s(Q(Y(t)))\| \\ &= \|Q(Y(t+s)) - Q(\psi_s(Y(t)))\| \leq (1/K)\|Y(t+s) - \psi_s(Y(t))\|_*\end{aligned} \tag{19}$$

where K is the strong convexity modulus of h and we used the fact that $\chi_t \circ Q = Q \circ \psi_t$ for all t (by Proposition 5). This implies that X_n is an APT of χ,[6] so our claim follows from the limit set theorem of Benaïm & Hirsch [7, Theorem 0.1]. ∎

We are thus left to prove the boundedness guarantee of Proposition 6.

Proof of Proposition 6. Our proof hinges on the construction of a suitable "energy function" $E\colon \mathcal{Y} \to \mathbb{R}_+$ for (RRM). To define it, we will assume for simplicity – and without loss of generality – that $\mathcal{X}$ has nonempty topological interior in $\mathcal{V}$ (which can be achieved by redefining $\mathcal{V}$ to be the affine hull of $\mathcal{X}$), that the reference point p in Definition 3 is the origin $0 \in \mathcal{V}$, and that $h(p) = 0$ (which can be achieved by a simple translation).

With this in mind, let $h^*(y) = \max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ denote the convex conjugate of h. Then, by Lemma A.2 in Appendix A, we have

$$(K/2)\|Q(y)\|^2 \leq h^*(y) \leq -\min h + \langle y, Q(y)\rangle + (2/K)\|y\|_*^2 \quad \text{for all } y \in \mathcal{Y} \tag{20}$$

where we note that $\min h \leq h(p) = 0$ by assumption. Since h is lower-semicontinuous, we have $h = h^{**}$ by the Fenchel–Moreau theorem. In addition, the Moreau–Rockafellar theorem [4, Theorem 4.17] implies that h^* is coercive because it can be written as $h^*(y) = h^*(y) - \langle y, p\rangle$ and $0 = p \in \operatorname{ri}\mathcal{X} = \operatorname{ri}\operatorname{dom} h^{**}$ by subcoercivity. Finally, since $\mathcal{X}$ has nonempty interior, it follows that the polar cone $\operatorname{PC}(x)$ is trivially 0 for all $x \in \operatorname{ri}\mathcal{X}$, so the subdifferential ∂h of h is compact-valued on $\mathcal{K} \subseteq \operatorname{ri}\mathcal{X}$. Thus, by the upper hemicontinuity of the subdifferential and the compactness of $\mathcal{K}$, we deduce that the image $\mathcal{D} = \partial h(\mathcal{K})$ of $\mathcal{K}$ under ∂h is compact, cf. [28, p. 201]. Hence, by the coercivity of h^* and the fact that $Q(y) = x$ if and only if $\partial h(x) \ni y$ (cf. Lemma A.1 in Appendix A), there exists some $c > 0$ such that $h^*(y) \leq c$ whenever $Q(y) \in \mathcal{K}$, i.e., $Q^{-1}(\mathcal{K})$ is contained in the c-sublevel set $L_c^-(h^*)$ of h^*.

With all this said and done, fix some $c' > c$ and let

$$E(y) = \varphi(h^*(y)) \quad \text{for all } y \in \mathcal{Y} \tag{21}$$

where $\varphi\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a C^2-smooth "gauge function" with the following properties: *i)* $\varphi(u) = 0$ for $u \leq c$; *ii)* $\varphi(u) = \sqrt{u}$ for $u \geq c'$; *iii)* $\varphi'(u) \geq 0$ and $\varphi''(u) \leq 1$ for all $u \in \mathbb{R}_+$.[7] Then, setting $x = Q(y)$ and differentiating, we readily obtain

$$\nabla E(y) = \varphi'(h^*(y)) \cdot \nabla h^*(y) = \varphi'(h^*(y)) \cdot x \quad \text{for all } y \in \mathcal{Y} \tag{22}$$

and hence, by the smoothness properties of φ and h^*, there exists some constant $C_2 \geq 0$ such that

$$E(y + w) = E(y) + \varphi'(h^*(y)) \cdot \langle w, x\rangle + C_2\|w\|_*^2 \quad \text{for all } y, w \in \mathcal{Y}. \tag{23}$$

Therefore, combining Eqs. (22) and (23) and letting $E_n = E(Y_n)$, we obtain

$$E_{n+1} \leq E_n + \varphi'(h^*(Y_n)) \cdot \langle \hat{v}_n, X_n\rangle + C_2\|\hat{v}_n\|_*^2 \leq E_n + \varphi_n\langle b_n + U_n, X_n\rangle + C_2\|\hat{v}_n\|_*^2 \tag{24}$$

where we set $\varphi_n = \varphi'(h^*(Y_n))$ and we used the fact that $\varphi(h^*(y)) \cdot \langle v(x), x\rangle \leq 0$ for all $y \in \mathcal{Y}$ (the latter being a consequence of subcoercivity and the defining properties of φ). Accordingly, conditioning on $\mathcal{F}_n$ and taking expectations, we finally get

$$\mathbb{E}[E_{n+1} \,|\, \mathcal{F}_n] \leq E_n + \gamma_n\varphi_n B_n\|X_n\| + \gamma_n^2 M_n^2, \tag{25}$$

where we used the Cauchy-Schwarz inequality to bound $\langle b_n, X_n\rangle$ from above by $B_n\|X_n\|$ (recall also that $\mathbb{E}[U_n \,|\, \mathcal{F}_n] = 0$ by definition).

Now, let $\varepsilon_n = \gamma_n\varphi_n B_n\|X_n\| + M_n^2$ denote the "residual" term in (25), and consider the auxiliary process $E_n = E_{n+1} + \sum_{k=n+1}^{\infty}\varepsilon_k$. By (25), we have $\mathbb{E}[E_n \,|\, \mathcal{F}_n] \leq E_n + \sum_{k=n}^{\infty}\varepsilon_n =$

[6] We are grateful to V. Boone for pointing out this simple argument.

[7] That such a function exists is an exercise in the construction of aproximate identities, which we omit.

E_{n-1}, i.e., E_n is a supermartingale relative to $\mathcal{F}_n$. Moreover, by (20) and the definition of φ, we further have

$$\varphi_n = \frac{1}{2\sqrt{h^*(Y_n)}} \leq \frac{1}{\sqrt{2K}\|X_n\|} \quad \text{whenever } h^*(Y_n) \geq c' \tag{26}$$

so there exists some (deterministic) positive constant C_1 such that $\sup_n \varphi_n\|X_n\| \leq C_1$. We thus get

$$\sum_{n=1}^{\infty} \varepsilon_n \leq C_1 \sum_{n=1}^{\infty} \gamma_n B_n + C_2 \sum_{n=1}^{\infty} \gamma_n^2 M_n^2 < \infty \tag{27}$$

by the summability condition (Sum). This shows that $\mathbb{E}[\sum_n \varepsilon_n] < \infty$ and, in turn, that $\mathbb{E}[E_n] \leq \mathbb{E}[E_1] < \infty$, i.e., E_n is uniformly bounded in L^1. Accordingly, by Doob's submartingale convergence theorem [23, Theorem 2.5], it follows that E_n converges with probability 1 to some finite random limit E_∞. Since $\sum_n \varepsilon_n < \infty$, this implies that $E_n = E_{n-1} - \sum_{k=n}^{\infty} \varepsilon_n$ also converges to some (random) finite limit (a.s.). Therefore, by the coercivity of E, we deduce that $\limsup_n \|Y_n\|_* < \infty$ w.p.1, as claimed. ∎

5. Robust convergence and stable limit sets

Even though Theorem 2 provides a universal characterization of the long-run behavior of any algorithm of the general form (RRM), there are several issues that remain open, namely:

(1) How is the long-run behavior of an algorithm affected by a small perturbation in its initialization?

(2) Does the update structure of the gradient-like signals $\hat{v}_n$ affect the algorithm's end-state?

(3) Are some limit sets independent of the amount of information available to the players?

In view of all this, the rest of this section will focus on whether we can identify a class of "robust" limit sets that satisfy the above desiderata. We present and discuss our main results in Section 5.2 after some necessary definitions and prerequisites in Section 5.1.

5.1. Stochastically attracting sets and energy functions.

In the theory of dynamical systems, the established way of analyzing such questions is via the notion of an attractor (cf. the relevant discussion in Section 4). Following Nevel'son & Khasminskii [51], this notion can be adapted to our stochastic setting as follows:

Definition 4. Let $\mathcal{S}$ be a nonempty closed subset of $\mathcal{X}$. We say that $\mathcal{S}$ is *stochastically attracting* under (RRM) for a given tolerance level $\eta > 0$ if there exists a neighborhood $\mathcal{U}$ of $\mathcal{S}$ in $\mathcal{X}$ such that

$$\mathbb{P}(X_n \text{ converges to } \mathcal{S} \mid X_1 \in \mathcal{U}) \geq 1 - \eta. \tag{28}$$

In the context of stochastic approximation algorithms, the requirement (28) is reminiscent of results guaranteeing convergence with positive probability toward an attractor. Such guarantees are usually conditioned on the notion of *attainability*, as pioneered by Benaïm [5] and Duflo [16];[8] however, in the present setting, there are two salient difficulties with this approach, both having to do with the primal-dual nature of the mean dynamics (MD). On the one hand, if the players' mirror map $Q\colon \mathcal{Y} \to \mathcal{X}$ is surjective on the boundary of $\mathcal{X}$ (e.g., as in the case of Euclidean projections), it is in general impossible to define a conjugate primal flow on $\mathcal{X}$ – and hence, it is not possible to treat $\mathcal{S}$ as an attractor. On the other hand, if Q is interior-valued (that is, $\operatorname{im} Q = \operatorname{ri} \mathcal{X}$), the defining Robbins–Monro process Y_n

[8] A point $y \in \mathcal{Y}$ is said to be *attainable* by Y_n if, for every neighborhood $\mathcal{W}$ of y in $\mathcal{Y}$ and for all $n \geq 1$, we have $\mathbb{P}(Y_k \in \mathcal{W} \text{ for some } k \geq n) > 0$.

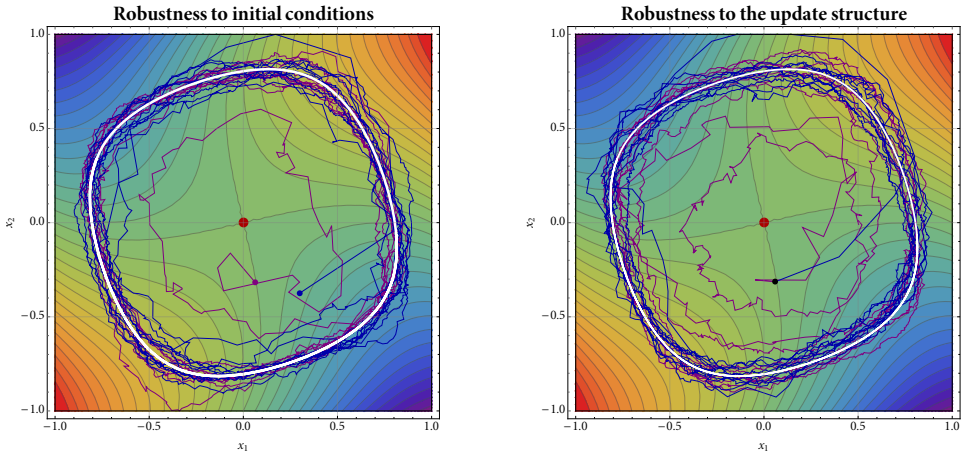


Figure 1: The long-run behavior of different online learning algorithms in the two-player min-max game defined by the loss function $\mathcal{L}(x_1, x_2) = x_1 x_2 + \varepsilon[\phi(x_2) - \phi(x_1)]$ with $\phi(z) = 2z^2 - 4z^4$, $\varepsilon \geq 0$, and $x_1, x_2 \in [-1, 1]$. The left plot shows two different initializations of (SGA), while the right plot displays the trajectories of (SGA) and (EG) from the same initialization. In both cases, the existence of an attractor allows for robust predictions that are largely independent of the initialization or exact update structure of (RRM).

may escape to infinity if $Q^{-1}(\mathcal{S}) = \varnothing$, and establishing attainability in this case can be as difficult as the original problem of proving (28) directly.

On account of all this, we will instead seek to establish (28) via a primal-dual variant of Lyapunov's direct method. In particular, since we are interested in the attracting properties of subsets of the *primal space* $\mathcal{X} \subseteq \mathcal{V}$, but the dynamics evolve in the *dual space* $\mathcal{Y} = \mathcal{V}^*$ of $\mathcal{V}$, our analysis will hinge on the following construction:

Definition 5. Let $\mathcal{S}$ be a nonempty closed subset of $\mathcal{X}$. We will say that $E\colon \mathcal{Y} \to [0, \infty)$ is a *local energy function for* $\mathcal{S}$ *under* (MD) *if* *a*) E is Lipschitz continuous and smooth; *b*) $Q(y) \to \mathcal{S}$ if and only if $E(y) \to 0$; and *c*) $\sup\{\dot{E}(y) : E_- < E(y) < E_+\} < 0$ for all sufficiently small $E_+ > E_- > 0$. In particular, if the last requirement holds for all $E_+ \leq \sup E$, we will refer to E as a *global* energy function for $\mathcal{S}$.

Informally, Definition 5 posits that E is smooth, positive-definite, and strictly decreasing along all nearby primal orbits $x(t) = Q(y(t))$ that do not lie in $\mathcal{S}$. For concreteness, we provide below a series of representative examples that will play an essential part in the sequel.

Example 5.1 (Variational stability)**.** Suppose that x^* satisfies (VS), i.e., $\langle v(x), x - x^* \rangle < 0$ for all $x \neq x^*$ in some neighborhood $\mathcal{U}$ of x^* in $\mathcal{X}$. Then a suitable primal-dual measure of distance from x^* is provided by the so-called "Fenchel coupling" [43]

$$F(y) = h(x^*) + h^*(y) - \langle y, x^* \rangle. \tag{29}$$

The key property of this coupling is that, under (MD), we have

$$\dot{F}(y) = \langle \dot{y}, \nabla h^*(y) \rangle - \langle \dot{y}, x^* \rangle = \langle v(x), x - x^* \rangle < 0 \quad \text{whenever } x \in \mathcal{U} \backslash \{x^*\} \tag{30}$$

where, in the penultimate step, we set $x = Q(y)$ and we invoked Lemma A.1 in Appendix A to write $Q(y) = \nabla h^*(y)$. By the Fenchel-Young inequality, we also have $F(y) \geq 0$ with

equality if and only if $Q(y) = x^*$ (cf. Lemma A.2), so $F(y)$ is a prime candidate for a local energy function.

To meet the entire range of requirements of Definition 5, we will need two further technical ingredients. The first is a regularity assumption on h, namely that

$$h(x_n) + \langle y_n, x - x_n \rangle \to h(x) \tag{R}$$

for all $x \in \mathcal{X}$ and all sequences of primal points $x_n \to x$ and subgradients $y_n \in \partial h(x_n)$. This condition simply posits that the first-order approximation of $h(x)$ from $h(x_n)$ is always accurate when $x_n \to x$, a property which is satisfied by all examples of regularizers that we have considered so far; for an in-depth discussion, cf. Azizian et al. [3] and references therein.

The second technicality is that F must grow at most linearly in $\|y\|_*$ in order to ensure the global Lipschitz continuity requirement of Definition 5. To achieve this, it suffices to rescale F for large values of x by means of the gauge function

$$\varphi(z) = \begin{cases} z & \text{if } 0 \leq z \leq 1 \\ 2\sqrt{z} - 1 & \text{if } z \geq 1 \end{cases} \tag{31}$$

which ensures that $\varphi \circ F$ behaves like F for small values of F, and like $\sqrt{F}$ for large values of F. We then have the following result:

Lemma 1. *Suppose that x^* satisfies (VS) and the players' regularizers satisfy (R). Then, with notation as above, the function $E(y) = \varphi(F(y))$ is a local energy function for x^* under (MD); moreover, if x^* is globally stable, E is a global energy function for x^* under (MD).*

To streamline our presentation, we defer the proof of Lemma 1 to Appendix A. We only note here that, by the relevant discussion in Section 2, the above yields an energy function for a wide class of games, including *a*) strictly monotone games (a global one in this case); *b*) games with second-order stationary equilibria as per (SOS); and *c*) all finite games admitting a strict Nash equilibrium. ◊

Example 5.2 (Convex optimization)**.** Consider the convex minimization problem $\min_{x \in \mathcal{X}} f(x)$ where $f\colon \mathcal{X} \to \mathbb{R}$ is a smooth convex function with a nonempty, compact set of minimizers $\mathcal{S} = \arg\min f$. To get a primal-dual measure of distance from $\mathcal{S}$, we may extend the definition of the coupling (29) to the current setting as

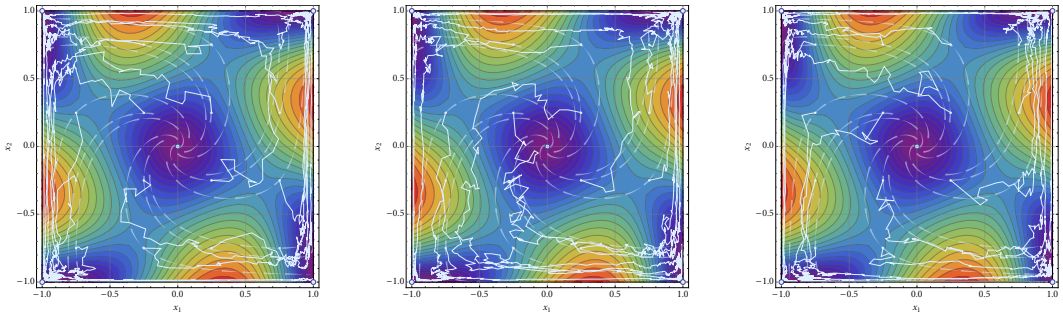
$$F(y) = h^*(y) - h^*_{\mathcal{S}}(y) \tag{32}$$

where $h^*_{\mathcal{S}}(y) = \max_{x \in \mathcal{S}} \{\langle y, x \rangle - h(y)\}$ denotes the convex conjugate of h relative to $\mathcal{S}$. As we show below, rescaling F by the gauge function (31) yields a global energy function for $\mathcal{S}$ under (MD):

Lemma 2. *Suppose that (R) holds. Then, with notation as above, the function $E(y) = \varphi(F(y))$ is a global energy function for $\mathcal{S} = \arg\min f$.*

As before, to keep the discussion going, we defer the proof of Lemma 2 to Appendix A. ◊

Example 5.3 (Discoordination games)**.** As a last example, consider a two-player discoordination game with payoff functions $u_1(x_1, x_2) = (x_1 - x_2)^2/2$ and $u_2(x_1, x_2) = (x_1 + x_2)^2/2$ for $x_1, x_2 \in [-1, 1]$. This game admits five critical points, the origin $(0, 0)$ and the four vertices $\{\pm 1, \pm 1\}$ of $\mathcal{X} = [-1, 1]^2$. None of these critical points is an equilibrium: the origin is unstable to deviations by both players, whereas the vertices are unstable to deviations by one of the players (but not the other). Given the lack of an equilibrium in pure strategies (a standard feature of discoordination games), the players' limiting behavior is quite difficult to predict; however, since the critical point at $(0, 0)$ is unstable for both players, it is reasonable to expect that it should be selected against.



(a) Gradient ascent **(b)** Extra-gradient **(c)** Optimistic gradient

Figure 2: Learning in a 2-player discoordination game. All algorithms under study converge to a corner critical point which resists deviations by one of the players (but not the other). The critical point at $(0, 0)$ is unstable to deviations by *both* players, and no trajectories converge there, even though it is the only interior ICT of (MD).

To examine this issue in the context of (MD), consider for concreteness the mirror map $Q_i(y_i) = \tanh(y_i/2)$ that is induced by the entropic regularizer $h_i(x_i) = (1 - x_i)\log(1 - x_i) + (1 + x_i)\log(1 + x_i)$.[9] In this case, it is straightforward to check that $E(y_1, y_2) = 2\operatorname{sech}(y_1/2)\operatorname{sech}(y_2/2)$ is an (almost global) local energy function for the four-corner set $\mathcal{S} = \{-1, 1\} \times \{-1, 1\}$. As a result, the sequence of play generated by (RRM) is expected to spend most time near one of these points, cf. Fig. 2. ◊

In Section 6, we present an additional range of examples that cover such cases as (stochastic) linear programming, the set of undominated strategies of a game, etc.

5.2. **Main results, implications, and applications.** We are now in a position to state our main results on the stable limit sets of (RRM). To do so, we will assume for concreteness that (RRM) is run with step-size and gradient signal sequences such that

$$\gamma_n = \gamma/n^{\ell_\gamma} \qquad B_n = \mathcal{O}(1/n^{\ell_b}) \qquad \text{and} \qquad M_n = \mathcal{O}(n^{\ell_\sigma}) \tag{33}$$

for some $\ell_\gamma \in [0, 1]$, $\ell_b > 0$ and $\ell_\sigma < 1/2$. Since the schedule (33) involves B_n and M_n (which, depending on the algorithm, may be beyond the players' control), this requirement may seem unverifiable at first glance. However, in view of Proposition 3, the exponents ℓ_b and ℓ_σ can be directly expressed in terms of the parameters of the specific algorithm under study, so this is not an issue.

Without further ado, we have the following general result:

Theorem 3. *Fix a tolerance level $\eta > 0$, and let $X_n = Q(Y_n)$ be the sequence of play generated by (RRM) with step-size and gradient signal sequences such that $\ell_\gamma + \ell_b > 1$ and $\ell_\gamma - \ell_\sigma > 1/2$ in (33). If $\mathcal{S}$ admits a local energy function and γ is sufficiently small, $\mathcal{S}$ is stochastically attracting; specifically, there exists a neighborhood $\mathcal{U}$ of $\mathcal{S}$, independent of the tolerance level η, such that $\mathbb{P}(X_n \text{ converges to } \mathcal{S} \mid X_1 \in \mathcal{U}) \geq 1 - \eta$. In addition, if the energy function on $\mathcal{S}$ is global, then, with probability 1, X_n converges to $\mathcal{S}$ from any initialization.*

Corollary 4. *Suppose that Algorithms 1–7 are run with step-size $\gamma_n \propto 1/n^{\ell_\gamma}$, $\ell_\gamma \in (1/2, 1]$, and, where applicable, a sampling parameter $\delta_n = \delta/n^{\ell_\delta}$ such that $1 - \ell_\gamma < \ell_\delta < \ell_\gamma - 1/2$. Then the conclusions of Theorem 3 hold.*

[9] For the general case, take $E(y) = [h^*(y) - \inf h^*]^{-1}$.

Remark 3. In the baseline case $B_n = 0$, $\sup_n \sigma_n =: \sigma < \infty$, the proof of Theorem 3 shows that it suffices to take $\gamma = \mathcal{O}(\min\{E_+/(\sigma H), M\sqrt{E_+/\beta}\}) \cdot \sqrt{\eta}$.

Theorem 3 and Corollary 4 are our main results concerning the stable limit sets of (RRM) so, before discussing their proof, we present a series of corollaries and applications thereof.

Corollary 5. *Suppose that Algorithms 1–7 are run with parameters as in Corollary 4. If x^* is globally variationally stable, then X_n converges to x^* w.p.1.*

Corollary 6. *Suppose that Algorithms 1–7 are run with parameters as in Corollary 4. If $\mathcal{G}$ is strictly monotone, then X_n converges to the game's unique Nash equilibrium w.p.1.*

The guarantees of Corollary 4 are particularly important from an equilibrium convergence standpoint, because, as we mentioned in Section 2, strictly monotone games account for a very wide range of applications – socially concave games [18], Cournot oligopolies [48], Kelly auctions [35], etc.

We should also stress here that neither of the above results can be inferred by the ICT convergence analysis of Section 4. In particular, if x^* lies at the boundary of $\mathcal{X}$, it may fail to be accessible unless the dual process Y_n escapes to infinity, in which case Theorem 2 no longer applies. This illustrates the flexibility of Definition 5, as it allows us to tackle at the same time both boundary *and* interior solutions, in both bounded and unbounded domains.

To the best of our knowledge, the only comparable global convergence results in the literature for oracle-based methods concern the convergence of the standard mirror descent algorithm ($B_n = 0$, $\sup_n \sigma_n^2 < \infty$) in strictly monotone games with compact domains [44]. For payoff-based algorithms, the closest results we are aware of are by Bravo et al. [9] and Tatarenko & Kamgarpour [66, 67] for a constrained variant of (SPSA) in strictly monotone games with compact domains (the latter actually showing convergence in probability, but without requiring strict monotonicity).

Finally, in terms of local results, Theorem 3 further yields the following corollaries:

Corollary 7. *Suppose that Algorithms 1–7 are initialized and run as per Corollary 4. If x^* is variationally stable – or, more narrowly, if it satisfies (SOS) – then X_n converges locally to x^* with arbitrarily high probability.*

Corollary 8. *Let x^* be a strict Nash equilibrium of a finite game. If Algorithms 4 and 7 are initialized and run as per Corollary 4, X_n converges locally to x^* with arbitrarily high probability.*

Of the above results, a special case of Corollary 7 was proven in [44] for unbiased signal sequences with finite unconditional variance (i.e., $B_n = 0$ and $\sup_n \mathbb{E}[\|U_n\|_*^2] < \infty$); at the time of writing, this seems to be the closest antecedent of our results in the literature. In particular, the convergence of Algorithms 2, 3 and 6 to variationally stable states and LNE satisfying (SOS) seems to be new.

Importantly, points satisfying (SOS) are the game-theoretic analogue of minimizers with a positive-definite Hessian in non-convex minimization problems [56]. In this regard, Corollary 7 is particularly important as it shows that such equilibria are attracting under the entire class of algorithms under study. Likewise, Corollary 8 is a key result because, generically – i.e., except on a set of games which is meager in the sense of Baire – pure Nash equilibria in finite games are always strict. Thus, coupled with the inherent instability of mixed equilibria in finite games [19], Corollary 8 goes a long way toward establishing a learning analogue of the "folk theorem" of evolutionary game theory which states that a Nash equilibrium is stable and attracting if and only if it is strict [30].

5.3. **Technical proofs.** We conclude this section with the proof of Theorem 3. The main ingredient of our analysis is a "template inequality" for (RRM) when the set under study admits an energy function (local or global).

To state it, note first that if E is an energy function for $\mathcal{S}$ under (MD), there exists some $E_+ > 0$ (possibly equal to $\sup E$) such that the sublevel set

$$\mathcal{D} = \{y \in \mathcal{Y} : E(y) \leq E_+\} \tag{34}$$

is forward invariant under (MD) and $\sup\{\dot{E}(y) : E_+ \geq E(y) > E_-\} < 0$ for all $E_- \in (0, E_+)$. Moreover, by assumption, there exist positive constants $\beta, H > 0$ such that $\|\nabla E(y)\| \leq H$ and

$$E(y') \leq E(y) + \langle \nabla E(y), y' - y\rangle + \tfrac{1}{2}\beta\|y' - y\|_*^2 \tag{35}$$

for all $y, y' \in \mathcal{Y}$. With all this in hand, we have the following template inequality:

Lemma 3. *Let $E_n \coloneqq E(Y_n)$. Then, for all $n = 1, 2, \dots$, we have*

$$E_{n+1} \leq E_n + \gamma_n\langle v(X_n), \nabla E(Y_n)\rangle + \gamma_n\xi_n + \gamma_n\chi_n + \gamma_n^2\psi_n^2, \tag{36}$$

where the error terms ξ_n, χ_n, and ψ_n are given by

$$\xi_n = \langle U_n, \nabla E(Y_n)\rangle, \quad \chi_n = H\|b_n\|_* \quad \text{and} \quad \psi_n^2 = \tfrac{1}{2}\beta\|\hat{v}_n\|_*^2. \tag{37}$$

Proof. Simply unroll (35) after substituting $y \leftarrow Y_n$ and $y' \leftarrow Y_{n+1} = Y_n + \gamma_n\hat{v}_n$ with $\hat{v}_n$ as in (6). ∎

Now, by the definition of E, we have $\dot{E}(y) = \langle v(Q(y)), \nabla E(y)\rangle < 0$ whenever $y \in \mathcal{D} \setminus Q^{-1}(\mathcal{S})$. Hence, for $X_n \in Q(\mathcal{D})$, (36) becomes

$$E_{n+1} \leq E_n + \gamma_n\xi_n + \gamma_n\chi_n + \gamma_n^2\psi_n^2. \tag{38}$$

Of course, each of these error terms can be positive, so E_n may fail to be decreasing, even when $X_n \in Q(\mathcal{D})$. On that account, it will be convenient to introduce the error processes

$$\mathrm{I}_n = \textstyle\sum_{k=1}^n \gamma_k\xi_k \qquad \mathrm{II}_n = \sum_{k=1}^n \gamma_k\chi_k \qquad \text{and} \qquad \mathrm{III}_n = \sum_{k=1}^n \gamma_k^2\psi_k^2 \tag{39}$$

which measure directly the aggregate effect of each error term in (36). As it turns out, under (Sum), these errors can be compensated by the negative drift of (36), leading to the following global result:

Proposition 7. *Suppose that $\mathcal{S}$ admits a global energy function, and let $X_n = Q(Y_n)$ be the sequence of play generated by (RRM). If (Sum) holds, then, with probability 1, X_n converges to $\mathcal{S}$.*

To streamline our discussion, before proving Proposition 7, we present a similar convergence result for sets that only admit *local* energy functions. In this case, even if the algorithm begins play close to $\mathcal{S}$, a single "bad" realization of the noise could force the process to exit the basin of attraction of $\mathcal{S}$, possibly never to return. With a fair degree of hindsight, we will control the probability with which this "bad event" occurs via the stability requirement

$$\mathbb{P}(Z_n > E_+/4 \text{ for some } n) < \eta/3 \tag{Stab}$$

where $\eta > 0$ is the target tolerance level, and $Z_n \leftarrow \mathrm{I}_n$, II_n or III_n, depending on the error term that we wish to control. Modulo this requirement, we obtain the following local analogue of Proposition 7:

Proposition 8. *Suppose that $\mathcal{S}$ admits a local energy function, and let $X_n = Q(Y_n)$ be the sequence of play generated by (RRM). Assume further that the algorithm begins play at a neighborhood $\mathcal{U}$ of $\mathcal{S}$ such that $E(Y_1) \leq E_+/4$. If (Sum) and (Stab) hold, then*

$$\mathbb{P}(E(Y_n) < E_+ \text{ for all } n \text{ and } \lim_{n\to\infty}\operatorname{dist}(X_n, \mathcal{S}) = 0) \geq 1 - \eta. \tag{40}$$

Of course, Propositions 7 and 8 can be difficult to employ in practice because of their reliance on the conditions (Sum) and (Stab). Because of this, we defer the proof of Propositions 7 and 8 to the end of this section, and we proceed below to complete the proof of Theorem 3 by showing that (Sum) and (Stab) both hold under the stated step-size and gradient signal requirements.

Proof of Theorem 3. We begin by noting that (Sum) holds trivially under the stated conditions for $\gamma_n \propto 1/n^{\ell_\gamma}$, $B_n = \mathcal{O}(1/n^{\ell_b})$ and $M_n = \mathcal{O}(n^{\ell_\sigma})$. As a result, the first part of the theorem follows immediately from Proposition 7.

Likewise, for the second part, it will suffice to establish the stability condition (Stab). To that end, proceeding term-by-term, we have:

(1) Since I_n is a martingale, Kolmogorov's inequality [23, Corollary 2.1] gives

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq k\leq n} \mathrm{I}_k \geq E_+/4\right) &\leq \mathbb{P}\left(\max_{1\leq k\leq n} |\mathrm{I}_k| \geq E_+/4\right)\\ &\leq \frac{16\,\mathbb{E}[\mathrm{I}_n^2]}{E_+^2} = \frac{16\,\mathbb{E}\big[\big(\sum_{k=1}^n \gamma_k\xi_k\big)^2\big]}{E_+^2}\\ &\leq \frac{16H^2\sum_{k=1}^n \gamma_k^2\sigma_k^2}{E_+^2} =: C_{\mathrm{I}} \end{aligned} \tag{41}$$

where we used the variance bound

$$\mathbb{E}[\xi_k^2] = \mathbb{E}[\mathbb{E}[|\langle U_k, \nabla E(Y_k)\rangle|^2 \,|\, \mathcal{F}_k]] \leq H^2\sigma_k^2 \tag{42}$$

and the fact that $\mathbb{E}[\xi_k\xi_m] = \mathbb{E}[\xi_k\xi_m \,|\, \mathcal{F}_{k\vee m}] = 0$ whenever $k \neq m$. Thus, given that $\{X_n \geq E_+/4 \text{ for some } n\} = \bigcup_n \{\mathrm{I}_n^* \geq E_+/4\}$ is a union of nested events, we conclude that (Stab) holds for $Z_n \leftarrow \mathrm{I}_n$ whenever $C_{\mathrm{I}} \leq \eta/3$.

(2) For the second term, we have $\mathrm{II}_n \leq H\sum_{k=1}^\infty \gamma_k B_k$ for all n with probability 1, so (Stab) holds for $Z_n \leftarrow \mathrm{II}_n$ as long as $C_{\mathrm{II}} := (4H/E_+)\sum_n \gamma_n B_n \leq 1$.

(3) Finally, for the last term, Markov's inequality yields

$$\mathbb{P}(\mathrm{III}_n \geq E_+/4) \leq \frac{4\,\mathbb{E}[\mathrm{III}_n]}{E_+} = \frac{2\beta\sum_{k=1}^n \gamma_k^2 M_k^2}{E_+} =: C_{\mathrm{III}}. \tag{43}$$

We thus see that the event $\{\mathrm{III}_n \geq E_+/4 \text{ for some } n\} = \bigcup_n \{\mathrm{III}_n \geq E_+/4\}$ occurs with probability no more than C_{III}, which implies in turn that the requirement (Stab) for $Z_n \leftarrow \mathrm{III}_n$ holds whenever $C_{\mathrm{III}} \leq \eta/3$.

Since C_{I}, C_{II} and C_{III} are all $\mathcal{O}(\gamma^2)$, we can choose γ sufficiently small so that $C_{\mathrm{I}} \leq \eta/3$, $C_{\mathrm{II}} \leq 1$ and $C_{\mathrm{III}} \leq \eta/3$. In this case, (Stab) holds by construction, and our claim follows from Proposition 8. ∎

We are thus left to prove Propositions 7 and 8. To that end, we begin with a technical lemma showing that the aggregate error processes $\mathrm{I}_n, \mathrm{II}_n$ and III_n of (39) are subleading relative to the long-run drift of (36).

Lemma 4. *Under (Sum), the aggregate error processes of (39) are sublinear in τ_n, i.e., we have*

$$Z_n/\tau_n \to 0 \quad \textit{with probability } 1, \tag{Sub}$$

where $Z_n \leftarrow \mathrm{I}_n$, II_n or III_n, depending on the error term under study.

Proof. We treat each case $Z_n \leftarrow \mathrm{I}_n$, II_n or III_n separately.

(1) For I_n, (Sum) readily gives

$$\sum_{n=1}^{\infty} \mathbb{E}[\gamma_n^2 \xi_n^2 \,|\, \mathcal{F}_n] \leq \sum_{n=1}^{\infty} \gamma_n^2 \,\mathbb{E}[\|\nabla E(Y_n)\|^2 \|U_n\|_*^2 \,|\, \mathcal{F}_n] \leq H^2 \sum_{n=1}^{\infty} \gamma_n^2 \sigma_n^2 < \infty. \tag{44}$$

Thus, by the strong law of large numbers for martingale difference sequences [23, Theorem 2.18], we conclude that I_n/τ_n converges to 0 with probability 1.

(2) For II_n, the conclusion is immediate by the fact that $\sum_n \gamma_n B_n < \infty$ under (Sum).

(3) Finally, for the submartingale term III_n, we have

$$\mathbb{E}[\mathrm{III}_n] = \sum_{k=1}^{n} \gamma_k^2 \,\mathbb{E}[\psi_k^2] \leq \frac{\beta}{2} \sum_{k=1}^{n} \gamma_k^2 \,\mathbb{E}[\|\hat{v}_k\|_*^2], \leq \frac{\beta}{2} \sum_{k=1}^{n} \gamma_k^2 M_k^2, \tag{45}$$

so, by (Sum), it follows that III_n is bounded in L^1. Therefore, by Doob's submartingale convergence theorem [23, Theorem 2.5], we further deduce that III_n converges (a.s.) to some (finite) random variable III_∞, implying in turn that $\mathrm{III}_n/\tau_n \to 0$ with probability 1. ∎

Moving forward, we present two lemmas that will allow us to deduce the convergence of the energy iterates $E_n \coloneqq E(Y_n)$ modulo the occurrence of the favorable event

$$\mathcal{E} = \{Y_n \in \mathcal{D} \text{ for all } n\} \tag{46}$$

where $\mathcal{D} = \{y \in \mathcal{Y} : E(y) \leq E_+\}$ is defined as in (34). In particular, we have the following results:

Lemma 5. *Suppose that $\mathbb{P}(\mathcal{E}) > 0$. If (Sub) holds, then $\mathbb{P}(\liminf_{n\to\infty} E_n = 0 \,|\, \mathcal{E}) = 1$.*

Lemma 6. *Suppose that $\mathbb{P}(\mathcal{E}) > 0$. If (Sum) holds, there exists some finite random variable E_∞ such that $\mathbb{P}(\lim_{n\to\infty} E_n = E_\infty \,|\, \mathcal{E}) = 1$.*

Proposition 9. *Suppose that E is a local energy function for $\mathcal{S}$. If $\mathbb{P}(\mathcal{E}) > 0$ and (Sum) holds, then $\mathbb{P}(X_n$ converges to $\mathcal{S} \,|\, \mathcal{E}) = 1$.*

Proof of Lemma 5. Since $\mathbb{P}(\mathcal{E}) > 0$, it suffices to show that the hitting time $N_a = \inf\{n \in \mathbb{N} : E_n \leq a\}$ is finite with probability 1 on $\mathcal{E}$ for all sufficiently small $a > 0$. More precisely, building on an argument of Duvocelle et al. [17], we will show that the event $\mathcal{N}_a = \mathcal{E} \cap \{N_a = \infty\}$ has $\mathbb{P}(\mathcal{N}_a) = 0$ whenever $0 < a \leq E_+$: indeed, if this is the case and $a_k \in (0, E_+)$, $k = 1, 2, \dots$, is a sequence converging monotonically to 0, we will have $\mathbb{P}(\mathcal{N}_{a_k}) = 0$ for all $k \in \mathbb{N}$. Thus, with only a countable number of $\mathcal{N}_{a_k}$ in play, we will have

$$\begin{aligned}
\mathbb{P}(\liminf_{n\to\infty} E_n = 0 \,|\, \mathcal{E}) &= \mathbb{P}(N_{a_k} < \infty \text{ for all } k \,|\, \mathcal{E}) \\
&= \mathbb{P}(\textstyle\bigcap_{k=1}^{\infty} \{N_{a_k} < \infty\} \,|\, \mathcal{E}) = 1 - \mathbb{P}(\textstyle\bigcup_{k=1}^{\infty} \{N_{a_k} < \infty\} \,|\, \mathcal{E}) \\
&= 1 - \frac{\mathbb{P}(\mathcal{E} \cap (\bigcup_{k=1}^{\infty} \{N_{a_k} < \infty\}))}{\mathbb{P}(\mathcal{E})} = 1 - \frac{\mathbb{P}(\bigcup_{k=1}^{\infty} \mathcal{N}_{a_k})}{\mathbb{P}(\mathcal{E})} = 1,
\end{aligned} \tag{47}$$

as per our original assertion.

Now, to establish our claim for $\mathcal{N}_a$, assume to the contrary that $\mathbb{P}(\mathcal{N}_a) > 0$ for some sufficiently small $a > 0$, and let $c_a = -\sup\{\dot{E}(y) : a \leq E(y) \leq E_+\}$, so $c_a > 0$ by Definition 5. Then, by telescoping (36), we get

$$\begin{aligned}
E_{n+1} &\leq E_1 + \sum_{k=1}^{n} \gamma_k \dot{E}(Y_k) + \sum_{k=1}^{n} \gamma_k \xi_k + \sum_{k=1}^{n} \gamma_k \chi_k + \sum_{k=1}^{n} \gamma_k \psi_k^2 \\
&\leq E_1 - \left[c_a - \frac{\mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n}{\tau_n} \right] \cdot \tau_n \qquad \text{for all } n = 1, 2, \dots
\end{aligned} \tag{48}$$

with probability 1 on $\mathcal{N}_a$. Since $\mathbb{P}(\mathcal{N}_a) > 0$ by assumption and $(\mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n)/\tau_n \to 0$ with probability 1 by (Sub), the above gives $\mathbb{P}(\lim_{n\to\infty} E_n = -\infty \,|\, \mathcal{N}_a) = 1$. However, with $\inf_n E_n \geq a > 0$ on $\mathcal{N}_a$ by construction, we get a contradiction, and our proof is complete. ∎

Proof of Lemma 6. Consider the nested sequence of events

$$\mathcal{E}_n = \{\dot{E}(Y_k) \leq 0 \text{ for all } k = 1, 2, \dots, n\} \tag{49}$$

so $\mathcal{E} = \bigcap_{n=1}^{\infty} \mathcal{E}_n$. Then, letting $\tilde{E}_n = \mathbb{1}_{\mathcal{E}_n} E_n$, Eq. (36) readily gives

$$\begin{aligned}
\tilde{E}_{n+1} = \mathbb{1}_{\mathcal{E}_{n+1}} E_{n+1} &\leq \mathbb{1}_{\mathcal{E}_n} E_{n+1} \\
&\leq \mathbb{1}_{\mathcal{E}_n} E_n + \big(\gamma_n \dot{E}(Y_n) + \gamma_n \xi_n + \gamma_n \chi_n + \gamma_n^2 \psi_n^2\big)\, \mathbb{1}_{\mathcal{E}_n} \\
&\leq \tilde{E}_n + \gamma_n\, \mathbb{1}_{\mathcal{E}_n}\, \xi_n + \big(\gamma_n \chi_n + \gamma_n^2 \psi_n^2\big)\, \mathbb{1}_{\mathcal{E}_n},
\end{aligned} \tag{50}$$

where we used the fact that $\dot{E}(Y_k) = \langle v(X_k), \nabla E(Y_k)\rangle \leq 0$ for all $k = 1, 2, \dots, n$ if $\mathcal{E}_n$ occurs. Since $\mathcal{E}_n$ is $\mathcal{F}_n$-measurable, conditioning on $\mathcal{F}_n$ and taking expectations then yields

$$\begin{aligned}
\mathbb{E}[\tilde{E}_{n+1} \,|\, \mathcal{F}_n] &\leq \tilde{E}_n + \gamma_n\, \mathbb{1}_{\mathcal{E}_n}\, \mathbb{E}[\xi_n \,|\, \mathcal{F}_n] + \mathbb{1}_{\mathcal{E}_n}\, \mathbb{E}[\gamma_n \chi_n + \gamma_n^2 \psi_n^2 \,|\, \mathcal{F}_n] \\
&\leq \tilde{E}_n + \mathbb{E}[\gamma_n \chi_n + \gamma_n^2 \psi_n^2 \,|\, \mathcal{F}_n] \\
&\leq \tilde{E}_n + \gamma_n H B_n + \tfrac{1}{2} \beta \gamma_n^2 M_n^2.
\end{aligned} \tag{51}$$

Now, given that $\sum_n \gamma_n B_n$ and $\sum_n \gamma_n^2 M_n^2$ are both finite by (Sum), $\tilde{E}_n$ is an almost supermartingale with summable increments, i.e., $\sum_n \Big[\mathbb{E}[\tilde{E}_{n+1} \,|\, \mathcal{F}_n] - \tilde{E}_n\Big] < \infty$ w.p.1. Therefore, by Gladyshev's lemma [53, p. 49], we conclude that $\tilde{E}_n$ converges almost surely to some (finite) random variable. Since $\mathbb{P}(\mathcal{E}) > 0$ and $\mathbb{1}_{\mathcal{E}_n} = 1$ for all n if and only if $\mathcal{E}$ occurs, we further deduce that $\mathbb{P}(E_n \text{ converges} \,|\, \mathcal{E}) = \mathbb{P}(\tilde{E}_n \text{ converges} \,|\, \mathcal{E}) = 1$, and our claim follows. ∎

Proof of Proposition 9. By Lemma 4, (Sub) is satisfied whenever (Sum) is. Thus, by a tandem application of Lemmas 5 and 6, we conclude that $\lim_{n\to\infty} E_n = 0$ with probability 1 on $\mathcal{E}$. Finally, since $Q(y) \to \mathcal{S}$ whenever $E(y) \to 0$, our claim follows. ∎

We are now in a position to prove Propositions 7 and 8.

Proof of Proposition 7. By the definition of a global attractor, we have $E_+ = \sup E$, so $\mathbb{P}(\mathcal{E}) = 1$. Our claim is then an immediate consequence of Proposition 9. ∎

Proof of Proposition 8. Suppose that $E(Y_1) \leq E_+/4$. We then claim that the event $\mathcal{E}$ always occurs on the intersection of the events $\mathcal{E}_{\mathrm{I}}$, $\mathcal{E}_{\mathrm{II}}$, and $\mathcal{E}_{\mathrm{III}}$, where $\mathcal{E}_Z = \{Z_n \leq E_+/4 \text{ for all } n\}$. Indeed, this being trivially the case for $n = 1$, assume that $Y_k \in \mathcal{D}$ for all $k = 1, 2, \dots, n$ for some $n \geq 1$. Then, telescoping (36) yields

$$E_{n+1} \leq E_1 + \sum_{k=1}^{n} \gamma_k \dot{E}(Y_k) + \mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n \leq E_+/4 + 0 + E_+/4 + E_+/4 + E_+/4 = E_+ \tag{52}$$

by the inductive hypothesis and our other assumptions. This shows that $E_{n+1} \in \mathcal{D}$, so the induction argument is complete, and we conclude that $\mathcal{E} \supseteq \mathcal{E}_{\mathrm{I}} \cap \mathcal{E}_{\mathrm{II}} \cap \mathcal{E}_{\mathrm{III}}$. Now, by (Stab), we have $\mathbb{P}(\mathcal{E}_{\mathrm{I}}) \leq \eta/3$ and likewise for the rest, so we get

$$\mathbb{P}(\mathcal{E}) \geq \mathbb{P}(\mathcal{E}_{\mathrm{I}} \cap \mathcal{E}_{\mathrm{II}} \cap \mathcal{E}_{\mathrm{III}}) = 1 - \mathbb{P}(\mathcal{E}_{\mathrm{I}} \cup \mathcal{E}_{\mathrm{II}} \cup \mathcal{E}_{\mathrm{III}}) \geq 1 - \mathbb{P}(\mathcal{E}_{\mathrm{I}}) - \mathbb{P}(\mathcal{E}_{\mathrm{II}}) - \mathbb{P}(\mathcal{E}_{\mathrm{III}}) \geq 1 - \eta. \tag{53}$$

Our claim then follows directly from Proposition 9. ∎

6. Fast convergence to coherent sets

6.1. **The notion of coherence: definition and examples.** In this section, we will show that the analysis of the previous section can be strengthened considerably under a "structural alignment" notion, which we dub *coherence*. We begin with the definition and a series of motivating examples.

Definition 6. A nonempty compact subset $\mathcal{S}$ of $\mathcal{X}$ will be called *coherent* if it admits a (finite) set of *deviation directions* $\mathcal{Z} = \{z_1, \dots, z_m\} \subseteq \mathcal{V}$ such that

$$a) \quad \langle v(x), z\rangle < 0 \quad \text{for all } x \in \mathcal{S} \text{ and all } z \in \mathcal{Z}. \tag{54a}$$

$$b) \quad Q(y) \to \mathcal{S} \quad \text{whenever } \max_{z\in\mathcal{Z}} \langle y, z\rangle \to -\infty. \tag{54b}$$

In particular, if (54a) holds for all $x \in \mathcal{X}$, we will say that $\mathcal{S}$ is *globally coherent*; and if we want to stress that $\mathcal{S}$ is coherent but not globally so, we will say that $\mathcal{S}$ is *locally coherent*.

The motivation behind Definition 6 is as follows. First, the geometric condition (54a) posits that any deviation from $\mathcal{S}$ along a vector $z \in \mathcal{Z}$ is actively disincentivized by the players' individual gradient field v so, in a certain sense, v points locally "toward" $\mathcal{S}$. The second condition is game-independent and asks that the elements of $\mathcal{Z}$ are sufficient to identify $\mathcal{S}$ by acting as primal-dual "support vectors" for $\mathcal{S}$ under Q. The terminology "coherence" has been chosen precisely to indicate that these two properties dovetail to create a favorable convergence landscape under (RRM).

To illustrate this notion, we proceed below with a series of examples. The first two concern finite games; the last two concern continuous ones.

Example 6.1 (Strict equilibria in finite games)**.** Recall that a strict Nash equilibrium of a finite game $\Gamma = \Gamma(\mathcal{N}, \mathcal{A}, u)$ is a strategy profile x^* such that (NE) holds as a strict inequality for all $x \neq x^*$. An immediate consequence of this definition is that *a*) x^* is *pure*, i.e., it is supported on a single pure strategy profile $\alpha^* \in \mathcal{A}$; and that *b*) unilateral deviations from α^* lead to *strictly* inferior payoffs, i.e., $u_\alpha(\alpha_i^*; \alpha_{-i}^*) > u_i(\alpha_i; \alpha_{-i}^*)$ for all $\alpha_i \in \mathcal{A}_i \backslash \{\alpha_i^*\}$, $i \in \mathcal{N}$.

With this in mind, consider the set of unilateral deviations

$$\mathcal{Z} = \{e_{i\alpha_i} - e_{i\alpha_i^*} : \alpha_i \in \mathcal{A}_i \backslash \{\alpha_i^*\}, i \in \mathcal{N}\}. \tag{55}$$

Since $\langle v(x^*), e_{i\alpha_i} - e_{i\alpha_i^*}\rangle = u_i(\alpha_i; \alpha_{-i}^*) - u_i(\alpha_i^*; \alpha_{-i}^*) < 0$ for all $\alpha_i \in \mathcal{A}_i \backslash \{\alpha_i^*\}$, $i \in \mathcal{N}$, condition (54a) is satisfied. Lemma A.4 further shows that $Q_{i\alpha_i}(y) \to 0$ whenever $y_{i\alpha_i} - y_{i\alpha_i^*} \to -\infty$, so the requirement $Q(y) \to x^*$ of (54b) is also satisfied. In other words, *strict equilibria are coherent.* ◊

Example 6.2 (Undominated strategies)**.** Recall that a pure strategy $\alpha_i \in \mathcal{A}_i$ is *dominated* by $\beta_i \in \mathcal{A}_i$ if $u_i(\alpha_i; x_{-i}) < u_i(\beta_i; x_{-i})$ for all $x \in \mathcal{X}$. We then say that α_i is *eliminated* in a mixed strategy profile $x \in \mathcal{X}$ if α_i is not supported in x_i, i.e., if $x_{i\alpha_i} = 0$. A fundamental requirement for game-theoretic learning is that dominated strategies become extinct over time, i.e., that the trajectory of play converges to the set $\mathcal{X}^*$ of action profiles that eliminate all dominated strategies.[10]

This set is globally coherent. To see this, consider the set of dominating deviations

$$\mathcal{Z} = \{e_{i\alpha_i} - e_{i\beta_i} : \alpha_i \text{ is dominated by } \beta_i\}. \tag{56}$$

By definition, $\langle v(x), e_{i\alpha_i} - e_{i\beta_i}\rangle = u_i(\alpha_i; x_{-i}) - u_i(\beta_i; x_{-i}) < 0$ for all $x \in \mathcal{X}$, so (54a) holds globally. Moreover, for any finite game, $\mathcal{X}^*$ is a face of $\mathcal{X}$ [61] and hence compact. Finally, Lemma A.4 shows that $Q_{i\alpha_i}(y) \to 0$ if $y_{i\alpha_i} - y_{i\beta_i} \to -\infty$, so the requirement $Q(y) \to \mathcal{X}^*$

[10] The case of mixed strategies dominated by mixed strategies requires heavier notation, so we do not treat it.

of (54b) is also satisfied, and we conclude that *the set of undominated strategies is globally coherent.* ◊

Example 6.3 (Sharp equilibria in concave games)**.** Following Polyak [53], a Nash equilibrium of a concave game is *sharp* if the stationarity condition (FOS) holds as a strict inequality for all $x \neq x^*$, i.e.,

$$\langle v(x^*), x - x^* \rangle < 0 \quad \text{for all } x \neq x^*. \tag{Sharp}$$

Examples of sharp equilibria include deterministic Nash policies in generic stochastic games [71], power control and resource allocation games [62], etc.

Geometrically, sharp equilibria can be characterized by the condition that $v(x^*)$ lies in the (topological) interior of the polar cone $\mathrm{PC}(x^*)$ to $\mathcal{X}$ at x^*. This means in particular that there exists a polyhedral cone $\mathcal{C}$ that is spanned by a finite set of vectors $\mathcal{Z} = \{z_1, \dots, z_m\} \subseteq \mathcal{V}$ such that *a*) the tangent cone $\mathrm{TC}(x^*)$ to $\mathcal{X}$ at x^* is contained in the interior of $\mathcal{C}$; and *b*) $\langle v(x^*), z \rangle < 0$ for all $z \in \mathcal{Z}$. Lemma A.5 in Appendix A shows that $Q(y) \to \mathcal{S}$ if $\max_{z \in \mathcal{Z}} \langle y, z \rangle \to -\infty$, so we conclude that *sharp equilibria are coherent.* ◊

Example 6.4 (Stochastic linear programming)**.** To borrow an example from optimization (viewed here as a single-player game), let $\mathcal{X}$ be a convex polytope and consider the stochastic linear program

$$\begin{aligned} \text{maximize} \quad & u(x) = \mathbb{E}_\theta[\langle V(\theta), x \rangle] \\ \text{subject to} \quad & x \in \mathcal{X} \end{aligned} \tag{SLP}$$

where $V(\theta)$ is a random payoff vector drawn from some complete probability space $(\Theta, \mathbb{P}_\theta)$. By linearity, the set of solutions $\mathcal{X}^* = \arg\max u$ of (SLP) is a face of $\mathcal{X}$; moreover, if we let $v = \mathbb{E}_\theta[V(\theta)] = \nabla u(x)$, we have $\langle v, x - x^* \rangle < 0$ whenever $x^* \in \mathcal{X}^*$ and $x \in \mathcal{X} \setminus \mathcal{X}^*$. Finally, since $\mathcal{X}$ is a convex polytope, there exists a finite set of vectors $\mathcal{Z} = \{z_1, \dots, z_m\}$ such that *a*) $x^* + z \in \mathcal{X} \setminus \mathcal{X}^*$ for all $x^* \in \mathcal{X}^*$, $z \in \mathcal{Z}$; and *b*) every point $x \in \mathcal{X} \setminus \mathcal{X}^*$ can be decomposed as $x = x^* + \lambda z$ for some $x^* \in \mathcal{X}^*$, $z \in \mathcal{Z}$ and $\lambda > 0$. Lemma A.5 in Appendix A shows that $Q(y) \to \mathcal{X}^*$ whenever $\langle y, z \rangle \to -\infty$ for all $z \in \mathcal{Z}$, so (54b) is satisfied and we conclude that *the solution set $\mathcal{X}^*$ of (SLP) is globally coherent.* ◊

The above examples illustrate that the notion of coherence underlies a diverse range of game-theoretic settings and problems. In light of this, we devote the rest of this section to analyzing the convergence properties of coherent sets under (RRM).

6.2. **Convergence analysis and results.** The first thing to note is that, if $\mathcal{S}$ is coherent, it admits the local energy function

$$E(y) = \log\Big(1 + \sum\nolimits_{z \in \mathcal{Z}} \exp\langle y, z \rangle\Big) \tag{57}$$

Indeed, if $E(y) \to \inf E = 0$, we must have $\langle y, z \rangle \to -\infty$ for all $z \in \mathcal{Z}$, and hence $Q(y) \to \mathcal{S}$ by Definition 6. Moreover, for all y such that $x = Q(y)$ is sufficiently close to $\mathcal{S}$, we have $\nabla E(y) = \sum_{z \in \mathcal{Z}} \langle v(x), z \rangle e^{\langle y, z \rangle} / (1 + \sum_{z \in \mathcal{Z}} e^{\langle y, z \rangle}) < 0$ by the continuity of v. This shows that the requirements of Definition 5 are all satisfied, leading to the following corollary of Theorem 3:

Corollary 9. *Suppose that $\mathcal{S}$ is coherent, and let X_n be the sequence of play of (RRM) with step-size and gradient signal assumptions as in Theorem 3. Then the conclusions of Theorem 3 hold, namely (i) if $\mathcal{S}$ is globally coherent, X_n converges to $\mathcal{S}$ with probability 1; and (ii) if $\mathcal{S}$ is locally coherent, X_n converges locally to $\mathcal{S}$ with probability at least $1 - \eta$ if γ is small enough.*

Corollary 9 is a strong convergence guarantee in itself, but it does not exploit the sharper structural properties of coherent sets. As we show below, the assumptions of Theorem 3 on the method's step-size and gradient signals can be relaxed considerably, allowing in many cases the use of even *constant* step-sizes. To simplify the presentation, we will assume throughout that (RRM) adheres to the general parameter schedule (33). In this general setting, we have:

Theorem 4. *Let $X_n = Q(Y_n)$ be the sequence of play generated by (RRM) with step-size and gradient signal sequences as per (33). Then:*

Case 1: *If $\mathcal{S}$ is globally coherent, then, from any initialization, X_n converges to $\mathcal{S}$ w.p.1.*

Case 2: *If $\mathcal{S}$ is locally coherent and, in addition, (i) $\ell_\gamma - \ell_\sigma > 1/2$; or (ii) $0 \leq \ell_\gamma < q/(2+q)$ and $\ell_\sigma < 1/2 - 1/q$, there exists an open initialization domain $\mathcal{W} \subseteq \mathcal{Y}$ such that, for any $\eta > 0$*

$$\mathbb{P}(X_n \text{ converges to } \mathcal{S} \mid Y_1 \in \mathcal{W}) \geq 1 - \eta \tag{58}$$

provided that $\gamma > 0$ is small enough.

Before discussing the proof of Theorem 4, we present a series of explicit convergence guarantees for specific algorithms:

Corollary 10. *Suppose that Algorithms 1–7 are run with step-size $\gamma_n \propto 1/n^{\ell_\gamma}$, $\ell_\gamma \in [0,1]$, and, where applicable, a sampling parameter $\delta_n \propto 1/n^{\ell_\delta}$, $\ell_\delta \in (0,1/2)$. If $\mathcal{S}$ is globally coherent, X_n converges to $\mathcal{S}$ with probability 1 provided the following conditions are met:*

- *For Algorithms 1 and 4–7: no other requirements needed.*
- *For Algorithms 2 and 3: $\ell_\gamma > 0$.*

Corollary 11. *Suppose that Algorithms 1–7 are run with step-size $\gamma_n \propto 1/n^{\ell_\gamma}$, $\ell_\gamma \in [0,1]$, and, where applicable, a sampling parameter $\delta_n \propto 1/n^{\ell_\delta}$, $\ell_\delta \in (0,1/2)$. Then the conclusions of Theorem 4 for locally coherent sets continue to hold provided the following conditions are met:*

- *For Algorithm 1: $\ell_\gamma > 1/2$ if $q = 2$; no such requirement needed if $q > 2$.*
- *For Algorithms 2 and 3: $\ell_\gamma > 1/2$ if $q = 2$; $\ell_\gamma > 0$ otherwise.*
- *For Algorithms 4–7: no other requirements needed.*

We should stress here that, depending on the statistical properties of the players' feedback mechanism, the above results imply convergence even with a *constant* step-size, a feature which is quite unique in the context of stochastic approximation. To the best of our knowledge, the only comparable result in the literature in terms of step-size assumptions is the recent work of Giannou et al. [20] for local convergence to strict Nash equilibria: since strict equilibria are locally coherent, the analysis of Giannou et al. [20] corresponds to the last item of Corollary 11.

Perhaps surprisingly, the principal reason for this relaxation in terms of step-size requirements is *not* the boundedness of the q-th moments of the players' oracle: the step-size requirements of Section 5 cannot be relaxed for non-coherent attractors even if $q = \infty$; at the same time, the convergence guarantees of Theorem 4 for globally coherent sets yield convergence with a constant step-size even when $q = 2$. Instead, as we hinted at before, these sharper convergence properties are due to the fact that the quadratic error term $\mathrm{III}_n = \sum_{k=1}^{n} \gamma_k^2 \psi_k^2$ is not present in the case of coherent sets: it is precisely this simplification that leads to convergence with significantly faster step-size schedules.

Our last result builds on this observation to show that convergence occurs at a finite number of iterations if the mirror map of the process is surjective (e.g., if it is a Euclidean projection):[11]

Theorem 5. *Suppose that the mirror map $Q\colon \mathcal{Y} \to \mathcal{X}$ of (RRM) is surjective. If $\mathcal{S}$ is coherent, then, with probability 1, every trajectory $X_n = Q(Y_n)$ that converges to $\mathcal{S}$ does so in a finite number of iterations, i.e., there exists some n_0 such that $X_n \in \mathcal{S}$ for all $n \geq n_0$.*

Corollary 12. *Suppose that (RRM) is run with Euclidean projections and step-size and gradient signal sequences as per (33). If $\mathcal{S}$ is globally coherent and $\mathcal{X}$ is compact, the induced sequence of play $X_n = Q(Y_n)$ converges to $\mathcal{S}$ in a finite number of iterations (a.s.).*

In view of the above, coherent sets comprise perhaps the most well-behaved class of rational outcomes under (RRM): the agents' sequence of play converges to such sets in a finite number of iterations, even with bandit, payoff-based feedback. In turn, this means that the algorithms' long-run behavior remains robust in the face of uncertainty, a property with important implications for the general theory of learning in games.

6.3. **Technical proofs.** We conclude this section with the proof of Theorem 4. The key step to achieve this is the following refinement of Lemma 3 for coherent sets.

Lemma 7. *Suppose that $\mathcal{S} \subseteq \mathcal{X}$ is coherent, and let $E_z(y) = \langle y, z\rangle$ for $y \in \mathcal{Y}$, $z \in \mathcal{Z}$. Then the iterates $E_n = E_z(Y_n)$ of E_z satisfy the template inequality*

$$E_{n+1} \leq E_n + \gamma_n \langle v(X_n), z\rangle + \gamma_n \xi_n + \gamma_n \chi_n. \tag{59}$$

where the error terms ξ_n and χ_n are now given by

$$\xi_n = \langle U_n, z\rangle \qquad \text{and} \qquad \chi_n = \max_{z\in\mathcal{Z}} \|z\| \cdot B_n. \tag{60}$$

Proof. Simply set $y \leftarrow Y_{n+1}$ in $E_z(y)$ and invoke the definition of (RRM). ∎

Compared to Lemma 3, the template inequality (59) *does not* have a second-order term, so the second moment of $\hat{v}_n$ plays a much more minor role when dealing with coherent sets. This can be seen very clearly in the following coherent analogue of Proposition 7:

Proposition 10. *Suppose that $\mathcal{S}$ is globally coherent, and let $X_n = Q(Y_n)$ be the sequence of play generated by (RRM). If (Sub) holds, then X_n converges to $\mathcal{S}$ with probability 1.*

The crucial difference between Propositions 7 and 10 is that the former requires the summability condition (Sum), while the latter requires *only* the subleading growth requirement (Sub). The latter assumption grants much more flexibility to the players because they can employ practically *any* step-size of the form $\gamma_n \propto 1/n^{\ell_\gamma}$ for some $\ell_\gamma \in [0,1]$. A similar situation arises for locally coherent sets, in which case the stability requirement (Stab) can be replaced by the "dominance" condition

$$\mathbb{P}(\mathrm{I}_n \leq C\tau_n^\mu/2 \text{ for all } n) \geq 1-\eta \tag{Dom.I}$$

$$\mathbb{P}(\mathrm{II}_n \leq C\tau_n^\mu/2 \text{ for all } n) \geq 1-\eta \tag{Dom.II}$$

for some $C > 0$ and $\mu \in [0,1)$. Under this milder condition, we have:

Proposition 11. *Suppose that $\mathcal{S}$ is locally coherent, fix some confidence level $\eta > 0$, and let $X_n = Q(Y_n)$ be the sequence of play generated by (RRM). If (Sub) and (Dom) hold, there exists an unbounded open initialization domain $\mathcal{W} \subseteq \mathcal{Y}$ such that*

$$\mathbb{P}(X_n \text{ converges to } \mathcal{S} \mid Y_1 \in \mathcal{W}) \geq 1 - (m+1)\eta. \tag{62}$$

[11] As we explain in Appendix A, the image $\operatorname{im} Q$ of Q coincides with the prox-domain $\mathcal{X}_h = \operatorname{dom} \partial h$ of h. As such, a sufficient condition for Q to be surjective is for h to be Lipschitz continuous on $\mathcal{X}$.

To prove Propositions 10 and 11 – and, through them Theorems 4 and 5 – it will be convenient to introduce the family of sets

$$\mathcal{W}(a) = \{y \in \mathcal{Y} : \max_{z\in\mathcal{Z}} \langle y, z\rangle < -a\}. \tag{63}$$

By Definition 6, these sets are mapped to neighborhoods of $\mathcal{S}$ under Q, so they are particularly well-suited to serve as initialization domains for (RRM). In particular, by the requirements of Definition 6 and the continuity of v, there exists some a such that $c \coloneqq -\sup\{\langle v(Q(y)), z\rangle : y \in \mathcal{W}(a), z \in \mathcal{Z}\} < 0$. With all this in hand, the proofs of Propositions 10 and 11 are fairly straightforward.

Proof of Proposition 10. Since $\mathcal{S}$ is globally coherent, we can take $a = -\infty$ in the definition of $\mathcal{W}(a)$ above. Then, telescoping (59) readily yields

$$E_z(Y_{n+1}) \leq E_z(Y_1) - c\tau_n + \mathrm{I}_n + \mathrm{II}_n \quad \text{for all } z \in \mathcal{Z}. \tag{64}$$

Thus, if (Sub) holds, we get $E_z(Y_n) \to -\infty$ for all $z \in \mathcal{Z}$, i.e., $X_n = Q(Y_n) \to \mathcal{S}$. ∎

Proof of Proposition 11. Let $\mu \in [0, 1)$ be such that (Dom) holds for every $z \in \mathcal{Z}$ (recall that ξ_n depends on z), and let $\Delta a = \max_n\{C\tau_n^\mu - c\tau_n\}$. Then, if Y_1 is initialized in $\mathcal{W} \coloneqq \mathcal{W}(a + \Delta a)$, we claim that $Y_n \in \mathcal{W}(a)$ for all n. Indeed, this being trivially true for $n = 1$, assume it to be the case for all $k = 1, 2, \dots, n$. Then, by (59) and our inductive hypothesis, we get

$$\begin{aligned} E_z(Y_{n+1}) &\leq E_z(Y_1) - \sum_{k=1}^{n} \gamma_k \langle v(X_k), z\rangle + \mathrm{I}_n + \mathrm{II}_n \\ &\leq E_z(Y_1) - c\tau_n + C\tau_n^\mu/2 + C\tau_n^\mu/2 \leq -a - \Delta a + \Delta a \leq -a \end{aligned} \tag{65}$$

i.e., $Y_{n+1} \in \mathcal{W}(a)$, as claimed. Since $Y_n \in \mathcal{W}(a)$ for all n, we conclude that (64) holds with probability 1 on the event that (Dom.I) and (Dom.II) both hold for all $z \in \mathcal{Z}$. Since (Dom.I) involves $|\mathcal{Z}| = m$ separate events (one for each $z \in \mathcal{Z}$) and II_n does not depend on z, it follows that $E_z(Y_n) \to -\infty$ for all $z \in \mathcal{Z}$ with probability at least $1 - (m + 1)\eta$. Our claim then follows from Definition 6. ∎

We are now in a position to prove Theorem 4.

Proof of Theorem 4. As in the case of Theorem 3, our proof will hinge on showing that (Sub) and (Dom) hold under the stated step-size and sampling parameter schedules. Our claim will then follow by a direct application of Propositions 10 and 11.

First, regarding (Sub), the law of large numbers for martingale difference sequences [23, Theorem 2.18] shows that $\mathrm{I}_n/\tau_n \to 0$ w.p.1 on the event $\{\sum_n \gamma_n^2 \mathbb{E}[\xi_n^2 \,|\, \mathcal{F}_n]/\tau_n^2 < \infty\}$. However

$$\mathbb{E}[\xi_n^2 \,|\, \mathcal{F}_n] \leq \|z\|^2 \,\mathbb{E}[\|U_n\|_*^2 \,|\, \mathcal{F}_n] \leq \|z\|^2 \sigma_n^2 = \mathcal{O}(n^{2\ell_\sigma}) \tag{66}$$

so, in turn, given that $\ell_\sigma < 1/2$, we get

$$\sum_n \frac{\gamma_n^2 \,\mathbb{E}[\xi_n^2 \,|\, \mathcal{F}_n]}{\tau_n^2} = \mathcal{O}\left(\sum_n \frac{\gamma_n^2 \sigma_n^2}{\tau_n^2}\right) = \mathcal{O}\left(\sum_n \frac{n^{-2\ell_\gamma} n^{2\ell_\sigma}}{n^{2(1-\ell_\gamma)}}\right) = \mathcal{O}\left(\sum_n \frac{1}{n^{2-2\ell_\sigma}}\right) < \infty. \tag{67}$$

This establishes (Sub) for $Z_n \leftarrow \mathrm{I}_n$; as for the case of II_n, our claim follows by noting that $\sum_{k=1}^n \gamma_k B_k / \sum_{k=1}^n \gamma_k \to 0$ if and only if $B_n \to 0$, which is immediate from (33). This shows that (Sub) holds, so the first case of the theorem follows from Proposition 10.

Now, for the second case of the theorem, since B_n is deterministic and $B_n = \mathcal{O}(1/n^{\ell_b})$ for some $\ell_b > 0$, it is always possible to find $C > 0$ and $\mu \in (0, 1)$ so that (Dom.II)

holds. We are thus left to establish (Dom.I). To that end, let $\mathrm{I}_n^* = \sup_{1\leq k\leq n}|\mathrm{I}_n|$ and set $P_n \coloneqq \mathbb{P}(\mathrm{I}_n^* > C\tau_n^\mu/2)$ so

$$P_n \leq \frac{\mathbb{E}[|\mathrm{I}_n|^q]}{(C/2)^q\tau_n^{\mu q}} \leq c_q \frac{\mathbb{E}[\big(\sum_{k=1}^n \gamma_k^2\|U_k\|_*^2\big)^{q/2}]}{\tau_n^{\mu q}} \tag{68}$$

where c_q is a positive constant depending only on C and q, and we used Kolmogorov's inequality [23, Corollary 2.1] in the first step and the Burkholder–Davis–Gundy inequality [23, Theorem 2.10] in the second. To proceed, we will require the following variant of Hölder's inequality [6, p. 15]:

$$\left(\sum_{k=1}^n a_k b_k\right)^\rho \leq \left(\sum_{k=1}^n a_k^{\frac{\lambda\rho}{\rho-1}}\right)^{\rho-1} \sum_{k=1}^n a_k^{(1-\lambda)\rho} b_k^\rho \tag{69}$$

valid for all $a_k, b_k \geq 0$ and all $\rho > 1$, $\lambda \in [0,1)$. Then, substituting $a_k \leftarrow \gamma_k^2$, $b_k \leftarrow \|U_k\|_*^2$, $\rho \leftarrow q/2$ and $\lambda \leftarrow 1/2 - 1/q$, (68) gives

$$P_n \leq c_q \frac{\big(\sum_{k=1}^n \gamma_k\big)^{q/2-1} \sum_{k=1}^n \gamma_k^{1+q/2}\,\mathbb{E}[\|U_k\|_*^q]}{\tau_n^{\mu q}} \leq c_q \frac{\sum_{k=1}^n \gamma_k^{1+q/2}\sigma_k^q}{\tau_n^{1+(\mu-1/2)q}} \tag{70}$$

We now consider two cases, depending on whether the numerator of (70) is summable or not.

Case 1: $\ell_\gamma(1+q/2) \geq 1 + q\ell_\sigma$. In this case, the numerator of (70) is summable under (33), so the fraction in (70) behaves as $\mathcal{O}(1/n^{(1-\ell_\gamma)(1+(\mu-1/2)q)})$.

Case 2: $\ell_\gamma(1+q/2) < 1 + q\ell_\sigma$. In this case, the numerator of (70) is not summable under (33), so the fraction in (70) behaves as $\mathcal{O}\big(n^{1-\ell_\gamma(1+q/2)+q\ell_\sigma}\big/n^{(1-\ell_\gamma)(1+(\mu-1/2)q)}\big)$.

Thus, working out the various exponents, a straightforward – if tedious – calculation shows that there exists some $\mu \in (0,1)$ such that P_n is summable as long as $\ell_\sigma < 1/2 - 1/q$ and $0 \leq \ell_\gamma < q/(2+q)$. Hence, if γ is sufficiently small relative to η, we conclude that

$$\mathbb{P}(\mathrm{I}_n \leq C\tau_n^\mu/2 \text{ for all } n) \geq 1 - \textstyle\sum_n P_n \geq 1 - \eta/2. \tag{71}$$

Finally, if $\ell_\gamma > 1/2 + \ell_\sigma$, (Dom.I) is a straightforward consequence of (Stab) with $Z_n \leftarrow \mathrm{I}_n$. Our assertion then follows by putting everything together and invoking Proposition 11. ∎

We conclude this section with the proof of our finite-time convergence result.

Proof of Theorem 5. Since Q is surjective, Lemma A.1 shows that $Q^{-1}(\mathcal{S})$ contains a shifted copy of $\bigcup_{x\in\mathcal{S}} \mathrm{PC}(x)$. Thus, given that $\max_{z\in\mathcal{Z}}\langle Y_n, z\rangle \to -\infty$ by the proof of Theorem 4, it follows that, for every $a \in \mathbb{R}$, there exists some (possibly random) $n_0 \equiv n_0(a)$ such that $\max_{z\in\mathcal{Z}}\langle Y_n, z\rangle < -a$ for all $n \geq n_0$. This shows that Y_n converges to $Q^{-1}(\mathcal{S})$ within a finite number of iterations, as claimed. ∎

7. Concluding remarks

The proposed regularized Robbins–Monro (RRM) stochastic approximation framework captures a wide range of existing algorithms, both first- and zeroth-order, and it allows us to derive a series of convergence results in a unified way. Conceptually speaking, an appealing feature of this framework lies in the fact that it provides an analysis bluepring that can be used in several other settings and algorithms of interest. The associated workflow for this is as follows: it suffices to simply estimate the bounds B_n and M_n for the method under study (in the sense of Table 1); the long-run behavior of the method may then be harvested from Theorems 2–4. We leave the inclusion of even more general frameworks – such as algorithms

with adaptive step-sizes, asyncrhonous and/or delayed feedback [27, 52], etc. – to future work.

Another fruitful direction for future research concerns the wandering behavior of RRM methods. By Conley's decomposition theorem (the "fundamental theorem of dynamical systems"), a flow decomposes into a chain recurrent part and an attracting part; of these, the former is fragile because of the Kupka-Smale theorem, and is actually absent in all but a meager set of flows (in the Baire category sense of the term). This means that, generically, one would expect a learning process to wander about different basins of attraction, until, by the attainability theory of Benaïm [5] and Duflo [16], it is captured by one of them. Making this statement precise would complement our results in an essential way, and would give us a better understanding of the complex phenomena that arise in game-theoretic learning.

Appendix A. Regularizers and mirror maps

In this appendix we present some basic properties of the mirror map Q. To state them, recall first that the subdifferential of a h at $x \in \mathcal{X}$ is defined as $\partial h(x) \coloneqq \{y \in \mathcal{Y} : h(x') \geq h(x) + \langle y, x' - x\rangle \text{ for all } x' \in \mathcal{V}\}$, the *domain of subdifferentiability* of h is $\operatorname{dom} \partial h \coloneqq \{x \in \operatorname{dom} h : \partial h \neq \varnothing\}$, and the convex conjugate of h is defined as $h^*(y) = \max_{x\in\mathcal{X}}\{\langle y, x\rangle - h(x)\}$ for all $y \in \mathcal{Y}$. We then have the following basic results.

Lemma A.1. *Let h be a regularizer on $\mathcal{X}$, and let $Q\colon \mathcal{Y} \to \mathcal{X}$ be its induced mirror map. Then:*

1. *Q is single-valued on $\mathcal{Y}$: in particular, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, we have $x = Q(y) \iff y \in \partial h(x)$.*
2. *The prox-domain $\mathcal{X}_h \coloneqq \operatorname{im} Q$ of h satisfies $\mathcal{X}_h = \operatorname{dom} \partial h$ and, hence, $\operatorname{ri}\mathcal{X} \subseteq \mathcal{X}_h \subseteq \mathcal{X}$.*
3. *Q is $(1/K)$-Lipschitz continuous and $Q = \nabla h^*$.*
4. *For all $x \in \operatorname{ri}\mathcal{X}$, we have $y, y' \in \partial h(x)$ if and only if $\langle y' - y, x' - x\rangle = 0$ for all $x' \in \mathcal{X}$.*

Our second basic result concerns the Fenchel coupling

$$F(p, y) = h(p) + h^*(y) - \langle y, p\rangle \quad \text{for } p \in \mathcal{X},\ y \in \mathcal{Y}. \tag{A.1}$$

For our purposes, the most relevant properties of F are as follows:

Lemma A.2. *For all $p \in \mathcal{X}$ and all $y, y' \in \mathcal{Y}$, we have:*

$$a)\quad F(p, y) \geq 0 \quad \text{with equality if and only if } p = Q(y). \tag{A.2a}$$

$$b)\quad F(p, y) \geq \tfrac{1}{2}K\,\|Q(y) - p\|^2. \tag{A.2b}$$

$$c)\quad F(p, y') \leq F(p, y) + \langle y' - y, Q(y) - p\rangle + \tfrac{1}{2K}\|y' - y\|_*^2. \tag{A.2c}$$

In particular, if $h(0) = 0$, we have

$$(K/2)\|Q(y)\|^2 \leq h^*(y) \leq -\min h + \langle y, Q(y)\rangle + (2/K)\|y\|_*^2 \quad \text{for all } y \in \mathcal{Y} \tag{A.3}$$

Variants of Lemmas A.1 and A.2 already exist in the literature (see e.g., [44] and references therein), so we do not provide a proof. Instead, we proceed below to show how the above extends to the *setwise* Fenchel coupling

$$F_{\mathcal{S}}(y) \coloneqq h^*(y) - h^*_{\mathcal{S}}(y) = \min_{p\in\mathcal{S}}\{h(p) + h^*(y) - \langle y, p\rangle\} = \min_{p\in\mathcal{S}} F(p, y) \tag{A.4}$$

where $\mathcal{S}$ is a nonempty compact convex subset of $\mathcal{X}$ and $h^*_{\mathcal{S}}(y) = \max_{x\in\mathcal{S}}\{\langle y, x\rangle - h(y)\}$ denotes the convex conjugate of h relative to $\mathcal{S}$. The most important properties of $F_{\mathcal{S}}$ are encoded in the following lemma.

Lemma A.3. *With notation as above, we have:*

(1) $F_{\mathcal{S}}(y) \geq 0$ *with equality if and only if* $Q(y) \in \mathcal{S}$. *Moreover, under the reciprocity condition* (R), *we have* $F_{\mathcal{S}}(y) \to 0$ *if and only if* $Q(y) \to \mathcal{S}$.

(2) $F_{\mathcal{S}}$ *is differentiable and* $\nabla F_{\mathcal{S}}(y) = Q(y) - Q_{\mathcal{S}}(y)$, *where* $Q_{\mathcal{S}}(y) = \arg\max_{x\in\mathcal{S}}\{\langle y, x\rangle - h(x)\}$.

(3) *For all* $y, y' \in \mathcal{Y}$ *we have* $\|\nabla F_{\mathcal{S}}(y') - \nabla F_{\mathcal{S}}(y)\| \leq (2/K)\|y' - y\|_*$.

Proof. Since $\mathcal{S} \subseteq \mathcal{X}$, we have $h^*_{\mathcal{S}} \leq h^*$ by definition, and hence $F_{\mathcal{S}} \geq 0$. Moreover, since the minimum in (A.4) must be attained in $\mathcal{S}$, we get $F_{\mathcal{S}}(y) = 0$ if and only if $h(p) - h^*(y) - \langle y, p\rangle = 0$ for some $p \in \mathcal{S}$; by Lemma A.2, this occurs if and only if $Q(y) = p \in \mathcal{S}$, so our first claim follows.

Moving forward, to show that $F_{\mathcal{S}}(y) \to 0$ if and only if $Q(y) \to \mathcal{S}$, let y_n be a sequence in $\mathcal{Y}$, and let $x_n = Q(y_n)$. For the "if" part, since $\mathcal{S}$ is compact, we may assume without loss of generality (and by descending to a subsequence if necessary) that x_n converges to some $x \in \mathcal{S}$. Observe now that *a)* $0 \leq F_{\mathcal{S}}(y_n) \leq F(x, y_n)$ by the minimum (A.4); and *b)* $F(x, y_n) \to 0$ by (R). Thus, by sandwiching, we conclude that $F_{\mathcal{S}}(y_n) \to 0$. Conversely, if $F_{\mathcal{S}}(y_n) \to 0$, we may again assume by compactness (and by descending to a subsequence if necessary) that $x_n = Q(y_n)$ converges to some $\hat{x} \in \mathcal{X}$. If $\hat{x} \notin \mathcal{S}$, then, by (R), we have $\lim_{n\to\infty} F(x, y_n) > 0$ for all $x \in \mathcal{S}$. Since $\mathcal{S}$ is compact and $F(x, y)$ is continuous in x, we conclude that $\liminf_{n\to\infty} F_{\mathcal{S}}(y_n) > 0$, a contradiction which establishes our claim.

Our last two claims follow by applying Lemma A.1 to h and $h + \delta_{\mathcal{S}}$ where $\delta_{\mathcal{S}}$ denotes the convex indicator of $\mathcal{S}$. ∎

The next properties we discuss concern the way that different regions of $\mathcal{Y}$ are mapped to $\mathcal{X}$ under Q.

Lemma A.4 (Mertikopoulos & Sandholm, 2016, Prop. A.1)**.** *Let* h *be a regularizer on the simplex* $\Delta(\mathcal{A}) \subseteq \mathbb{R}^{\mathcal{A}}$. *If* $y_\alpha - y_\beta \to -\infty$, *then* $Q_\alpha(y) \to 0$.

Lemma A.5. *Let* h *be a regularizer on* $\mathcal{X}$, *let* y_n, $n = 1, 2, \dots$ *be a sequence in* $\mathcal{Y}$, *and fix some* $x \in \mathcal{X}$. *If* $\langle y_n, z\rangle \to -\infty$ *for every nonzero* $z \in \mathrm{TC}(x)$, *we have* $Q(y_n) \to x$.

Proof. Assume that $\limsup_n \|x_n - x\| > 0$. Then, given that $y_n \in \partial h(x_n)$, we get $h(x) \geq h(x_n) + \langle y_n, x - x_n\rangle \geq h(x_n) - \langle y_n, z_n\rangle\|x_n - x\|$, where we set $z_n = (x_n - x)/\|x_n - x\|$. If we further assume (by descending to a subsequence if needed) that z_n converges in the unit sphere of $\|\cdot\|$, there exists some $z \in \mathrm{TC}(x)$ with $\|z\| = 1$ and such that $\langle y_n, z_n\rangle \leq (1 + \varepsilon)\langle y_n, z\rangle$ for some $\varepsilon > 0$. Thus, taking the lim sup of the above estimate gives $h(x) \geq \infty$, a contradiction which proves our claim. ∎

Lemma A.6. *Let* h *be a regularizer on a convex polytope* $\mathcal{P}$ *of* $\mathcal{V}$, *let* $\mathcal{S}$ *be a face of* $\mathcal{P}$, *and let* $\mathcal{Z} = \{z_1, \dots, z_m\}$ *be a set of unit vectors of* $\mathcal{V}$ *such that every point* $x \in \mathcal{P} \setminus \mathcal{S}$ *can be written as* $x = p + \lambda z$ *for some* $p \in \mathcal{S}$, $z \in \mathcal{Z}$ *and* $\lambda > 0$. *If* $\max_{z\in\mathcal{Z}}\langle y, z\rangle \to -\infty$, *then* $Q(y) \to \mathcal{S}$.

Proof. By the compactness of $\mathcal{P}$ (and descending to a subsequence if necessary), we may assume that $x_n = Q(y_n)$ converges to some $x \in \mathcal{P}$. If $x \notin \mathcal{S}$, there exist $p \in \mathcal{S}$, $z \in \mathcal{Z}$ and $\lambda > 0$ such that $x = p + \lambda z$. In turn, this gives $h(p) \geq h(x_n) + \langle y_n, p - x_n\rangle = h(x_n) - \langle y_n, z_n\rangle\|x_n - p\|$ where we set $z_n = (x_n - p)/\|x_n - p\|$. Since $z_n \to z$, taking $n \to \infty$ yields $h(p) \geq \infty$, a contradiction which shows that $x = \lim x_n \in \mathcal{S}$, as claimed. ∎

We conclude this appendix with the dynamical properties of the Fenchel coupling under (MD). The heavy lifting will be provided by the following simple lemma:

Lemma A.7. *Let* $x(t) = Q(y(t))$ *be an orbit of* (MD). *Then, for every nonempty closed convex subset* $\mathcal{S}$ *of* $\mathcal{X}$, *we have*

$$\dot{F}_{\mathcal{S}}(y) = \langle v(x), x - x_{\mathcal{S}}\rangle \tag{A.5}$$

where $x_{\mathcal{S}} = Q_{\mathcal{S}}(y)$ denotes the mirror image of y on $\mathcal{S}$. In particular, if $\mathcal{S} = \{p\}$, we have $\dot{F}(p, y) = \langle v(x), x - p\rangle$.

Proof. Simply note that $\dot{y} = -v(x)$ and apply Lemma A.3. ∎

Appendix B. Omitted proofs and calculations

B.1. **Error bounds for specific algorithms.** Our aim in this appendix is to prove the bounds on the bias and magnitude of $\hat{v}_n$ reported in Proposition 3 and Table 1.

Proof of Proposition 3. We proceed in a method-by-method basis starting with the oracle-based methods of Section 3.2, that is, Algorithms 1–4. For this, we will make free use of the fact that we can take $M_n^q = 3^{q-1}(G^q + B_n^q + \sigma_n^q)$ in (8), cf. the discussion after (6).

► **Algorithm 1: Stochastic gradient ascent.** For (SGA), we have $U_n = \operatorname{err}(X_n; \theta_n)$ and $b_n = 0$, so our claim follows immediately from the stated assumptions for (SFO).

► **Algorithm 2: Extra-gradient.** For (EG), we have $\hat{v}_n = V(X_{n+1/2}; \theta_{n+1/2})$ so $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = \mathbb{E}[v(X_{n+1/2}) \,|\, \mathcal{F}_n]$. We thus get

$$\begin{aligned}\|b_n\|_* = \|\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] - v(X_n)\|_* &\leq \mathbb{E}[\|v(X_{n+1/2}) - v(X_n)\|_* \,|\, \mathcal{F}_n]\\ &\leq L\,\mathbb{E}[\|X_{n+1/2} - X_n\| \,|\, \mathcal{F}_n]\\ &\leq \gamma_n L\,\mathbb{E}[\|V(X_n; \theta_n)\|_* \,|\, \mathcal{F}_n]\\ &= \gamma_n L\,\mathbb{E}[\|v(X_n) + \operatorname{err}(X_n; \theta_n)\|_* \,|\, \mathcal{F}_n]\\ &\leq \gamma_n L(G + \sigma) = \mathcal{O}(\gamma_n) = \mathcal{O}(1/n^{\ell_\gamma}) \qquad \text{(B.1)}\end{aligned}$$

and, analogously

$$\|U_n\|_* = \|\hat{v}_n - \mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n]\|_* = \|v(X_{n+1/2}) - \mathbb{E}[v(X_{n+1/2}) \,|\, \mathcal{F}_n] + \operatorname{err}(X_{n+1/2}; \theta_{n+1/2})\|_* \quad \text{(B.2)}$$

so $\mathbb{E}[\|U_n\|_*^q \,|\, \mathcal{F}_n] = \mathcal{O}(G^q + \sigma^q) = \mathcal{O}(1)$ under (13), as claimed.

► **Algorithm 3: Optimistic gradient.** For (OG), we have again $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = \mathbb{E}[v(X_{n+1/2}) \,|\, \mathcal{F}_n]$, so the same series of arguments as above gives

$$\begin{aligned}\|b_n\|_* &= \|\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] - v(X_n)\|_*\\ &\leq L\,\mathbb{E}[\|X_{n+1/2} - X_n\| \,|\, \mathcal{F}_n]\\ &\leq \gamma_n L\,\mathbb{E}[\|V(X_{n-1/2}; \theta_{n-1})\|_* \,|\, \mathcal{F}_n]\\ &= \gamma_n L\,\mathbb{E}[\|v(X_{n-1/2}) + \operatorname{err}(X_{n-1/2}; \theta_{n-1})\|_* \,|\, \mathcal{F}_n]\\ &\leq \gamma_n L(G + \sigma) = \mathcal{O}(\gamma_n) = \mathcal{O}(1/n^{\ell_\gamma}) \qquad \text{(B.3)}\end{aligned}$$

under (13) with $q = \infty$. The noise term U_n can be bounded in exactly the same way, so we omit the calculations.

► **Algorithm 4: Exponential / multiplicative weights.** We consider two cases, based on the information available to the players. For the full information oracle (14a), we have $\hat{v}_n = v(X_n)$ so $b_n = U_n = 0$ by definition (i.e., the oracle is perfect). Otherwise, under the realization-based oracle (14b), we have $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = \mathbb{E}[v(\alpha_n) \,|\, \mathcal{F}_n] = v(X_n)$ because α_n is sampled according to X_n. We thus get $b_n = 0$ and $U_n = \mathcal{O}(1)$, which proves our assertion.

We now proceed with the payoff-based methods of Section 3.2, namely Algorithms 5–7.

► **Algorithm 5: Single-point stochastic approximation.** Since u_i is assumed bounded in the context of (SPSA), the bound for M_n follows trivially. As for the bias of (SPSA), it will be convenient to set $V_i^\delta(x;w) = (d_i/\delta)\, u_i(x+\delta w)\, w_i$ so, in obvious notation, $\hat{v}_{i,n} = V_i^{\delta_n}(X_n; W_n)$. Thus, if we fix a pivot point $x \in \mathcal{X}$ and a query point $\hat{x} = x + \delta w$ for some $w \in \mathcal{E} = \prod_i \mathcal{E}_i$, a first-order Taylor expansion of u_i with integral remainder gives

$$V_i^\delta(x;w) = \frac{d_i}{\delta} u_i(\hat{x}) \cdot w_i = \frac{d_i}{\delta} u_i(x) \cdot w_i + \frac{d_i}{\delta} \langle \nabla u_i(x), z \rangle \cdot w_i \tag{B.4a}$$

$$+ \int_0^1 \langle \nabla u_i(x+\tau z) - \nabla u_i(x), z \rangle \, d\tau \cdot w_i \tag{B.4b}$$

where we set $z = \hat{x} - x = \delta w$. Hence, if w is drawn uniformly at random from $\mathcal{E}$, taking expectations yields

$$\begin{aligned}
\mathbb{E}[\text{(B.4a)}] &= \frac{d_i}{\delta} \mathbb{E}[\langle v_i(x), z_i \rangle \, w_i] + \frac{d_i}{\delta} \sum_{j \neq i} \langle \nabla_{x_j} u_i(x), \mathbb{E}[z_j] \rangle \, \mathbb{E}[w_i] \\
&= d_i \, \mathbb{E}[\langle v_i(x), w_i \rangle \, w_i] = d_i \cdot \frac{1}{2d_i} \sum_{\ell=1}^{d_i} [v_{i\ell}(x) e_{i\ell} - v_{i\ell}(x)(-e_{i\ell})] = v_i(x)
\end{aligned} \tag{B.5}$$

where we used the fact that $\mathbb{E}[w_i] = 0$ for all $i \in \mathcal{N}$ and that w_i and w_j are independent for all $i, j \in \mathcal{N}$, $i \neq j$. As for the second term, Assumption 1 readily yields

$$\|\mathbb{E}[\text{(B.4b)}]\| \leq \frac{d_i}{\delta} \int_0^1 L_i \delta^2 \|w\|^2 \tau \, d\tau = \mathcal{O}(L\delta). \tag{B.6}$$

Thus, by combining (B.5) and (B.6), we conclude that $b_{i,n} = \mathbb{E}[V_i^{\delta_n}(X_n; W_n) \,|\, \mathcal{F}_n] - v_i(X_n) = \mathcal{O}(\delta_n)$, which immediately yields the desired bound $B_n = \mathcal{O}(\delta_n) = \mathcal{O}(1/n^{\ell_\delta})$ for (SPSA).

► **Algorithm 6: Dampened gradient approximation.** Recall that $\hat{v}_{i,n} = n \cdot \log(1 + (u_i(X_{n+1/2}) - u_i(X_n)) W_{i,n})$. Since $u_i(X_{n+1/2}) - u_i(X_n) = (1/n) v_i(X_n) W_{i,n} + \mathcal{O}(1/n^2)$ by the definition of $X_{n+1/2}$, expanding the logairthm readily yiels $B_n = \mathcal{O}(1/n)$ and $M_n = \mathcal{O}(1)$. Our claim then follows as above.

► **Algorithm 7: Exponential weights for exploration and exploitation.** Since $\hat{\alpha}_n$ is sampled according to $\hat{X}_n$, we readily get $\mathbb{E}[\hat{v}_{i,n} \,|\, \mathcal{F}_n] = v_i(\hat{X}_n)$, so $B_n = \mathcal{O}(\|\hat{X}_n - X_n\|) = \mathcal{O}(\delta_n) = \mathcal{O}(1/n^{\ell_\delta})$. Moreover, since $\hat{X}_{i\alpha_i,n} \geq \delta_n / A_i$, it follows that $\|\hat{v}_n\|_* = \mathcal{O}(1/\delta_n) = \mathcal{O}(n^{\ell_\delta})$, and our proof is complete. ■

B.2. **Energy function derivations.** Our aim in this last appendix is to prove the energy properties of the Fenchel coupling as stated in Lemmas 1 and 2. For concision, we will prove both as a special case of the following general result:

Proposition B.1. *Let $\mathcal{S}$ be a nonempty compact convex subset of $\mathcal{X}$, and assume that there exists a neighborhood $\mathcal{U}$ of $\mathcal{S}$ such that*

$$\langle v(x), x - p \rangle \leq 0 \quad \text{for all } x \in \mathcal{U},\ p \in \mathcal{S}, \tag{B.7}$$

with equality if and only if $x \in \mathcal{S}$. If (R) holds and φ is defined as in (31), the function $E\colon \mathcal{Y} \to \mathbb{R}$ given by

$$E(y) = \varphi(F_{\mathcal{S}}(y)) \quad \text{for all } y \in \mathcal{Y} \tag{B.8}$$

is a local energy function for $\mathcal{S}$ under (MD). In addition, if $\mathcal{U} = \mathcal{X}$, E is a global energy function for $\mathcal{S}$.

Proof. We will verify the requirements of Definition 5 in order.

(1) For the first (Lipschitz continuity and smoothness), note that $\nabla E(y) = \varphi'(F_{\mathcal{S}}(y))\nabla F_{\mathcal{S}}(y)$, so, lettting $x = Q(y)$ and $x_{\mathcal{S}} = Q_{\mathcal{S}}(y)$, Lemma A.3 yields

$$\nabla E(y) = (x - x_{\mathcal{S}}) \cdot \begin{cases} 1 & \text{if } F_{\mathcal{S}}(y) \leq 1, \\ 1/\sqrt{F_{\mathcal{S}}(y)} & \text{otherwise.} \end{cases} \tag{B.9}$$

Furthermore, by Lemma A.2, we also have $F(p, y) \geq (K/2)\|x - p\|^2$ for all $p \in \mathcal{S}$, so, by minimizing over $p \in \mathcal{S}$, we get

$$F_{\mathcal{S}}(y) \geq (K/2)\operatorname{dist}(x, \mathcal{S})^2 = (K/2)\|x - \operatorname{pr}_{\mathcal{S}}(x)\|^2 \tag{B.10}$$

where $\operatorname{pr}_{\mathcal{S}}(x) \coloneqq \arg\min_{p\in\mathcal{S}}\|x - p\|$. In turn, this gives $\|x - \operatorname{pr}_{\mathcal{S}}(x)\| \leq \sqrt{2/K}$ whenever $F_{\mathcal{S}}(y) \leq 1$, so we get

$$\|\nabla E(y)\| = \|x - x_{\mathcal{S}}\| \leq \|x - \operatorname{pr}_{\mathcal{S}}(x)\| + \|\operatorname{pr}_{\mathcal{S}}(x) - x_{\mathcal{S}}\| \leq \sqrt{2/K} + \operatorname{diam}(\mathcal{S}) \tag{B.11}$$

whenever $F_{\mathcal{S}}(y) \leq 1$. On the other hand, if $F_{\mathcal{S}}(y) \geq 1$, we have

$$\begin{aligned} \|\nabla E(y)\| = \frac{\|\nabla F_{\mathcal{S}}(y)\|}{\sqrt{F_{\mathcal{S}}(y)}} &\leq \sqrt{\frac{2}{K}}\frac{\|x - x_{\mathcal{S}}\|}{\|x - \operatorname{pr}_{\mathcal{S}}(x)\|} && \text{\# by (B.9) and (B.10)} \\ &\leq \sqrt{\frac{2}{K}}\left(1 + \frac{\|\operatorname{pr}_{\mathcal{S}}(x) - x_{\mathcal{S}}\|}{\|x - \operatorname{pr}_{\mathcal{S}}(x)\|}\right) && \text{\# by the triangle inequality} \\ &\leq \sqrt{\frac{2}{K}}\left(1 + \frac{\operatorname{diam}(\mathcal{S})}{\|x - \operatorname{pr}_{\mathcal{S}}(x)\|}\right) && \text{(B.12)} \end{aligned}$$

By the reciprocity condition (R), it follows that the set $\{x = Q(y) : F_{\mathcal{S}}(y) \geq 1\}$ is well-separated from $\mathcal{S}$, so $\|x - \operatorname{pr}_{\mathcal{S}}(x)\|$ is bounded away from zero if $F_{\mathcal{S}}(y) \geq 1$. Thus, by combining Eqs. (B.11) and (B.12), we conclude that $\|\nabla E(y)\|$ is bounded. Finally, again by Lemma A.2, $F(y)$ is $(1/K)$-Lipschitz smooth, so Eq. (35) – which is equivalent to the Lipschitz smoothness of E – follows immediately from the concavity of φ.

(2) For the positive-definiteness requirement of Definition 5, note that Lemma A.3 and the reciprocity condition (R) yield $Q(y) \to \mathcal{S}$ if and only if $F_{\mathcal{S}}(y) \to 0$. Thus, given that $\varphi(z) = z$ for small z, the same will hold for $E = \varphi \circ F_{\mathcal{S}}$, and our claim follows.

(3) Finally, for the Lyapunov properties of E under (MD), recall that Lemma A.7 gives $\dot{F}_{\mathcal{S}}(y) = \langle v(x), x - x_{\mathcal{S}}\rangle$, so

$$\dot{E}(y) = \langle \dot{y}, \nabla E(y)\rangle = \varphi'(F(y))\langle v(x), x - x_{\mathcal{S}}\rangle < 0 \quad \text{whenever } x \in \mathcal{U} \setminus \mathcal{S} \tag{B.13}$$

where we used the defining property (B.7) of $\mathcal{S}$ (recall that $x_{\mathcal{S}} \in \mathcal{S}$ by construction). Moving forward, by Lemma A.3, there exists some $E_+ > 0$ such that the sublevel set $\mathcal{D} = \{y \in \mathcal{Y} : F_{\mathcal{S}}(y) \leq E_+\}$ is mapped to $\mathcal{U}$ under Q, i.e., $Q(y) \in \mathcal{U}$ whenever $F_{\mathcal{S}}(y) \leq E_+$. Thus, putting everything together, we conclude that $\dot{E}(y) \to 0$ if and only if $F_{\mathcal{S}}(y) \to 0$, which implies that $\sup\{\dot{E}(y) : E_- < E(y) < E_+\} < 0$ for all $E_- \in (0, E_+)$, and our proof is complete. ∎

Acknowledgments

P. Mertikopoulos is grateful for financial support by the French National Research Agency (ANR) in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02), the LabEx PERSYVAL (ANR-11-LABX-0025-01), MIAI@Grenoble Alpes (ANR-19-P3IA-0003), and the bilateral ANR-NRF grant ALIAS (ANR-19-CE48-0018-01). This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and

innovation programme (grant agreement № 725594-TIME-DATA) and from the Swiss National Science Foundation (SNSF) under grant number 200021-205011.

References

[1] Arrow, K. J., Hurwicz, L., and Uzawa, H. *Studies in linear and non-linear programming*. Stanford University Press, 1958.

[2] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 1995.

[3] Azizian, W., Iutzeler, F., Malick, J., and Mertikopoulos, P. On the rate of convergence of Bregman proximal methods in constrained variational inequalities. http://arxiv.org/abs/2211.08043, 2022.

[4] Bauschke, H. H. and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, NY, USA, 2 edition, 2017.

[5] Benaïm, M. Vertex reinforced random walks and a conjecture of Pemantle. *Annals of Probability*, 25: 361–392, 1997.

[6] Benaïm, M. Dynamics of stochastic approximation algorithms. In Azéma, J., Émery, M., Ledoux, M., and Yor, M. (eds.), *Séminaire de Probabilités XXXIII*, volume 1709 of *Lecture Notes in Mathematics*, pp. 1–68. Springer Berlin Heidelberg, 1999.

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

[8] Bervoets, S., Bravo, M., and Faure, M. Learning with minimal information in continuous games. *Theoretical Economics*, 15:1471–1508, 2020.

[9] Bravo, M., Leslie, D. S., and Mertikopoulos, P. Bandit learning in concave N-person games. In *NeurIPS '18: Proceedings of the 32nd International Conference of Neural Information Processing Systems*, 2018.

[10] Brown, G. W. Iterative solutions of games by fictitious play. In Coopmans, T. C. (ed.), *Activity Analysis of Productions and Allocation*, 374-376. Wiley, 1951.

[11] Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[12] Coucheney, P., Gaujal, B., and Mertikopoulos, P. Penalty-regulated dynamics and robust learning procedures in games. *Mathematics of Operations Research*, 40(3):611–633, August 2015.

[13] Daskalakis, C. and Panageas, I. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *ITCS '19: Proceedings of the 10th Conference on Innovations in Theoretical Computer Science*, 2019.

[14] Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. Training GANs with optimism. In *ICLR '18: Proceedings of the 2018 International Conference on Learning Representations*, 2018.

[15] Debreu, G. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences of the USA*, 38(10):886–893, October 1952.

[16] Duflo, M. Cibles atteignables avec une probabilité positive d'après M. Benaïm. mimeo, 1997.

[17] Duvocelle, B., Mertikopoulos, P., Staudigl, M., and Vermeulen, D. Multi-agent online learning in time-varying games. *Mathematics of Operations Research*, 48(2):914–941, May 2023.

[18] Even-dar, E., Mansour, Y., and Nadav, U. On the convergence of regret minimization dynamics in concave games. In *STOC '09: Proceedings of the 41st annual ACM symposium on the Theory of Computing*, pp. 523–532, New York, NY, 2009. ACM.

[19] Flokas, L., Vlatakis-Gkaragkounis, E. V., Lianeas, T., Mertikopoulos, P., and Piliouras, G. No-regret learning and mixed Nash equilibria: They do not mix. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

[20] Giannou, A., Vlatakis-Gkaragkounis, E. V., and Mertikopoulos, P. The convergence rate of regularized learning in games: From bandits and uncertainty to optimism and beyond. In *NeurIPS '21: Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021.

[21] Giannou, A., Lotidis, K., Mertikopoulos, P., and Vlatakis-Gkaragkounis, E. V. On the convergence of policy gradient methods to Nash equilibria in general stochastic games. In *NeurIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022.

[22] Gidel, G., Berard, H., Vignoud, G., Vincent, P., and Lacoste-Julien, S. A variational inequality perspective on generative adversarial networks. In *ICLR '19: Proceedings of the 2019 International Conference on Learning Representations*, 2019.

[23] Hall, P. and Heyde, C. C. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York, 1980.

[24] Hart, S. and Mas-Colell, A. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93(5):1830–1836, 2003.

[25] Hart, S. and Mas-Colell, A. Stochastic uncoupled dynamics and Nash equilibrium. *Games and Economic Behavior*, 57:286–303, 2006.

[26] Héliou, A., Cohen, J., and Mertikopoulos, P. Learning with bandit feedback in potential games. In *NIPS '17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

[27] Héliou, A., Mertikopoulos, P., and Zhou, Z. Gradient-free online learning in continuous games with delayed rewards. In *ICML '20: Proceedings of the 37th International Conference on Machine Learning*, 2020.

[28] Hiriart-Urruty, J.-B. and Lemaréchal, C. *Fundamentals of Convex Analysis*. Springer, Berlin, 2001.

[29] Hofbauer, J. and Sandholm, W. H. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, November 2002.

[30] Hofbauer, J. and Sigmund, K. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479–519, July 2003.

[31] Hsieh, Y.-G., Iutzeler, F., Malick, J., and Mertikopoulos, P. On the convergence of single-call stochastic extra-gradient methods. In *NeurIPS '19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 6936–6946, 2019.

[32] Hsieh, Y.-G., Iutzeler, F., Malick, J., and Mertikopoulos, P. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

[33] Hsieh, Y.-P., Mertikopoulos, P., and Cevher, V. The limits of min-max optimization algorithms: Convergence to spurious non-critical sets. In *ICML '21: Proceedings of the 38th International Conference on Machine Learning*, 2021.

[34] Juditsky, A., Nemirovski, A. S., and Tauvel, C. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[35] Kelly, F. P., Maulloo, A. K., and Tan, D. K. H. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, March 1998.

[36] Korpelevich, G. M. The extragradient method for finding saddle points and other problems. *Èkonom. i Mat. Metody*, 12:747–756, 1976.

[37] Kushner, H. J. and Yin, G. G. *Stochastic approximation algorithms and applications*. Springer-Verlag, New York, NY, 1997.

[38] Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, Cambridge, UK, 2020.

[39] Leslie, D. S. and Collins, E. J. Individual Q-learning in normal form games. *SIAM Journal on Control and Optimization*, 44(2):495–514, 2005.

[40] Leslie, D. S. and Collins, E. J. Generalised weakened fictitious play. *Games and Economic Behavior*, 56 (2):285–298, August 2006.

[41] Littlestone, N. and Warmuth, M. K. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[42] Maynard Smith, J. and Price, G. R. The logic of animal conflict. *Nature*, 246:15–18, November 1973.

[43] Mertikopoulos, P. and Sandholm, W. H. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4):1297–1324, November 2016.

[44] Mertikopoulos, P. and Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, January 2019.

[45] Mertikopoulos, P., Papadimitriou, C. H., and Piliouras, G. Cycles in adversarial regularized learning. In *SODA '18: Proceedings of the 29th annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.

[46] Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C.-S., Chandrasekhar, V., and Piliouras, G. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *ICLR '19: Proceedings of the 2019 International Conference on Learning Representations*, 2019.

[47] Mertikopoulos, P., Hallak, N., Kavis, A., and Cevher, V. On the almost sure convergence of stochastic gradient descent in non-convex problems. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

[48] Monderer, D. and Shapley, L. S. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996.

[49] Nemirovski, A. S. and Yudin, D. B. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York, NY, 1983.

[50] Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120 (1):221–259, 2009.

[51] Nevel'son, M. B. and Khasminskii, R. Z. *Stochastic Approximation and Recursive Estimation*. American Mathematical Society, Providence, RI, 1976.

[52] Oliveira, T. R., Rodrigues, V. H. P., Kristić, M., and Başar, T. Nash equilibrium seeking with arbitrarily delayed player actions. In *CDC '20: Proceedings of the 59th IEEE Annual Conference on Decision and Control*, 2019.

[53] Polyak, B. T. *Introduction to Optimization*. Optimization Software, New York, NY, USA, 1987.

[54] Popov, L. D. A modification of the Arrow–Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

[55] Rakhlin, A. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *NIPS '13: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2013.

[56] Ratliff, L. J., Burden, S. A., and Sastry, S. S. On the characterization of local Nash equilibria in continuous games. *IEEE Trans. Autom. Control*, 61(8):2301–2307, August 2016.

[57] Robbins, H. and Monro, S. A stochastic approximation method. *Annals of Mathematical Statistics*, 22: 400–407, 1951.

[58] Robinson, J. An iterative method for solving a game. *Annals of Mathematics*, 54:296–301, 1951.

[59] Rosen, J. B. Existence and uniqueness of equilibrium points for concave N-person games. *Econometrica*, 33(3):520–534, 1965.

[60] Rosenthal, R. W. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

[61] Samuelson, L. and Zhang, J. Evolutionary stability in asymmetric games. *Journal of Economic Theory*, 57:363–391, 1992.

[62] Scutari, G., Facchinei, F., Palomar, D. P., and Pang, J.-S. Convex optimization, game theory, and variational inequality theory in multiuser communication systems. *IEEE Signal Process. Mag.*, 27(3): 35–49, May 2010.

[63] Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *NIPS' 06: Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 1265–1272. MIT Press, 2006.

[64] Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control*, 37(3):332–341, March 1992.

[65] Syrgkanis, V., Agarwal, A., Luo, H., and Schapire, R. E. Fast convergence of regularized learning in games. In *NIPS '15: Proceedings of the 29th International Conference on Neural Information Processing Systems*, pp. 2989–2997, 2015.

[66] Tatarenko, T. and Kamgarpour, M. Learning generalized Nash equilibria in a class of convex games. *IEEE Trans. Autom. Control*, 64(4):1426–1439, 2019.

[67] Tatarenko, T. and Kamgarpour, M. Learning Nash equilibria in monotone games. In *CDC '19: Proceedings of the 58th IEEE Annual Conference on Decision and Control*, 2019. doi: 10.1109/CDC40024.2019.9029659.

[68] Taylor, P. D. and Jonker, L. B. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1-2):145–156, 1978.

[69] Tullock, G. Efficient rent seeking. In Tollison, J. M. B. R. D. and Tullock, G. (eds.), *Toward a theory of the rent-seeking society*. Texas A&M University Press, 1980.

[70] Vovk, V. G. Aggregating strategies. In *COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory*, pp. 371–383, 1990.

[71] Zhang, R., Ren, Z., and Li, N. Gradient play in multi-agent Markov stochastic games: Stationary points and convergence. https://arxiv.org/abs/2106.00198, 2021.