# LEARNING IN CONCAVE GAMES WITH IMPERFECT INFORMATION

PANAYOTIS MERTIKOPOULOS

Abstract. This paper examines the convergence properties of a class of learning schemes for concave $N$-person games – that is, games with convex action spaces and individually concave payoff functions. Specifically, we focus on a family of learning methods where players adjust their actions by taking small steps along their individual payoff gradients and then "mirror" the output back to their feasible action spaces. Assuming players only have access to gradient information that is accurate up to a zero-mean error with bounded variance, we show that when the process converges, its limit is a Nash equilibrium. We also introduce an equilibrium stability notion which we call *variational stability* (VS), and we show that stable equilibria are locally attracting with high probability whereas globally stable states are globally attracting with probability 1. Additionally, in finite games, we find that dominated strategies become extinct, strict equilibria are locally attracting with high probability, and the long-term average of the process converges to equilibrium in 2-player zero-sum games. Finally, we examine the scheme's convergence speed and we show that if the game admits a strict equilibrium and the players' mirror maps are surjective, then, with high probability, the process converges to equilibrium in a finite number of steps, no matter the level of uncertainty.

## Contents



## 1. Introduction

In the standard framework of online sequential optimization, an optimizing agent selects at each instance $n = 0, 1, \dots$ an action $x_n$ from some set $\mathcal{X}$ and obtains a reward $u_n(x_n)$ based on an a priori unknown payoff function $u_n \colon \mathcal{X} \to \mathbb{R}$. The agent then receives some problem-dependent feedback (e.g. a noisy estimate of the

2010 *Mathematics Subject Classification.* Primary 91A26, 90C15; secondary 90C33, 68Q32.

*Key words and phrases.* Concave games; learning; variational stability; mirror descent; imperfect feedback; Fenchel coupling; Nash equilibrium; dominated strategies.

The author gratefully acknowledges financial support from the CNRS under grant no. REAL.NET–PEPS–JCJC–2016, and the French National Research Agency (ANR) under grants ANR–GAGA–13–JS01–0004–01 and ANR–NETLEARN–13–INFR–004.

gradient of $u_n$ at $x_n$), and updates $x_n$ with the goal of minimizing the instantaneous payoff gap $\epsilon_n = \max_{x\in\mathcal{X}} u_n(x) - u_n(x_n)$ or an averaged version thereof – such as the induced regret $R_n = \max_{x\in\mathcal{X}} \sum_{k=0}^{n} [u_k(x) - u_k(x_k)]$.

Game-theoretic learning is a multi-agent variant of the above framework in which a set of players interact at each stage $n = 0, 1, \dots$, and each player's payoff function is determined by the actions of all other players at stage $n$ – so the dependence of $u_n$ on $n$ is *implicit*, not explicit. Since the mechanism that determines the players' payoff functions is now fixed (though possibly unknown and/or opaque to the players), finer convergence criteria apply, chief among them being that of convergence to a Nash equilibrium. With this in mind, this paper focuses on the following question: if the players of a repeated game concurrently employ some learning rule to increase their individual payoffs (for instance, if they follow a no-regret algorithm), do their strategies converge to a Nash equilibrium of the underlying one-shot game?

This question is largely motivated by the extremely successful applications of game theory to data networks and distributed systems where fast convergence to a stable equilibrium state is essential. In this setting, players naturally have an imperfect, localized view of their environment, typically subject to random – and possibly unbounded – estimation errors and noise. Thus, to improve their individual payoffs as the game is repeated over time, we posit that players try to learn from their past experiences by employing adaptive algorithms that induce relatively small, careful adjustments at each step.

In the case of finite games, Hart and Mas-Colell [16] showed that learning based on (external) regret minimization leads to the so-called *Hannan set*, a set of correlated strategies which includes the game's set of Nash equilibria. However, as was recently shown by Viossat and Zapechelnyuk [50], the Hannan set also contains thoroughly non-rationalizable strategies that assign positive weight *only* on strictly dominated strategies. As a result, blanket no-regret statements in finite games do not indicate convergence to Nash equilibrium; worse still, they do not even imply the long-run elimination of dominated strategies.

In games with continuous, convex action sets (such as the ones we consider here), learning typically takes place at the level of pure strategies – as opposed to mixed or correlated strategies that are more common in finite games. In this general context, starting with the seminal work of Zinkevich [53] on online convex programming, the most widely used class of algorithms for no-regret learning is the family of *online mirror descent* (OMD) schemes pioneered by Shalev-Shwartz and Singer [45] and the closely related "Follow the Regularized Leader" (FoReL) method of Kalai and Vempala [22].[1] If the players' action spaces are convex and their payoff functions are individually concave, employing an OMD-based scheme guarantees that players have no regret in the long run. However, except for certain special cases, this does not imply that the induced sequence of play converges to Nash equilibrium – and indeed, in many cases, it doesn't.

In view of the above, our aim in this paper is to (*i*) analyze the equilibrium convergence properties of no-regret, mirror-based learning in concave games; and (*ii*) assess the speed and robustness of this convergence in the presence of noise,

[1]The terminology "mirror descent" dates back to Nemirovski and Yudin [33] who introduced these methods in ordinary (static) convex programming. This class includes the standard online gradient descent (OGD) method of Zinkevich [53] and the widely used exponential weights (EW) scheme of Vovk [52] and Littlestone and Warmuth [27] for online learning.

uncertainty, and other feedback impediments. To this end, instead of restricting our attention to a specific class of games (such as zero-sum or potential games), we introduce a notion of equilibrium stability which we call *variational stability* (VS), and which is formally similar to the influential notion of evolutionary stability that was introduced in population games by Maynard Smith and Price [28]. By means of this stability notion, we are able to establish a series of convergence results that hold irrespective of the magnitude of the errors affecting the players' observations.

**Paper organization and outline of results.** In Section 2, we describe in detail the class of $N$-person concave games under study and we introduce the notion of variational stability. The family of mirror-based learning (ML) methods that we consider is then presented in Section 3. In a nutshell, the main idea of ML is as follows: at each stage $n = 0, 1, \dots$, every player takes a step along (an estimate of) the individual gradient of their payoff function and the output is "mirrored" onto each player's action space by means of a "choice map" that is analogous to ordinary Euclidean projection – in fact, it is a natural generalization thereof. Regarding the players' gradient information, we only assume that players have access to unbiased, bounded-in-mean-square estimates of their true payoff gradients; apart from these bare-bones hypotheses, we make no other tameness or independence assumptions on the errors affecting the players' feedback process.

Our main results can be summarized as follows: First, in Section 4, we show that when it exists, the limit of the process is a Nash equilibrium of the underlying game (a.s.). Subsequently, we show that stable states are locally attracting with high probability while globally stable states are globally attracting with probability 1. As a corollary, if the game admits a concave potential or if it is diagonally concave in the sense of Rosen [41], ML converges to Nash equilibrium almost surely, no matter the level of uncertainty.

In Section 5, we briefly outline some applications to learning in finite games. Specifically, we show that: (*i*) dominated strategies become extinct; (*ii*) strict Nash equilibria are locally attracting with high probability; and (*iii*) in zero-sum games, the long-term average of the players' mixed strategies converges to Nash equilibrium.

Finally, in Section 6, we examine the convergence speed of ML schemes. To do so, we focus again on variationally stable states, and we show that the long-term average gap from such states decays as $\mathcal{O}(n^{-1/2})$ if the scheme's step-size is chosen appropriately. In a similar fashion, we also show that the algorithm's expected running length until players reach an $\varepsilon$-neighborhood of a stable state is $\mathcal{O}(1/\varepsilon^2)$. Up to factors that do not depend on $n$ or $\varepsilon$, these rates hold for all ML schemes. However, if such an algorithm is run with a surjective choice map and the underlying game admits a strict equilibrium (a direct extension of the notion of strict equilibrium in finite games), then, with high probability, players converge to equilibrium in a *finite* number of steps.

Our analysis relies heavily on tools and techniques from the theory of stochastic approximation, martingale limit theory and convex/variational analysis. In particular, with regard to the latter, we make heavy use of a "primal-dual divergence" measure between action and gradient variables, which we call the *Fenchel coupling*. This coupling is a primal-dual analogue of the well-known Bregman divergence, and thanks to its Lyapunov-like properties, it provides a potent tool for proving convergence – both pointwise and setwise.

**Related work.** Mirror descent methods were pioneered by Nemirovski and Yudin [33] and have since given rise to an extensive literature in mathematical optimization. In this context, the works that are closest to our paper are those of Nemirovski et al. [32], Nesterov [35] and Juditsky et al. [21], where sharp convergence rate estimates are derived for (stochastic) convex programs, variational inequalities (VIs) and saddle-point problems. Motivation and setting aside, a fundamental difference between these works and the current one is that the former focus almost exclusively on the averaged sequence $\bar{x}_n = \sum_{k=0}^{n} \gamma_k x_k / \sum_{k=0}^{n} \gamma_k$, where $\gamma_n$ is the step-size of the method. By contrast, in game theory and sequential optimization, the figure of merit is the *actual* sequence of play $x_n$ that determines the players' payoffs at each stage, and whose behavior may diverge considerably from that of $\bar{x}_n$. Specifically, because there is no inherent averaging in $x_n$, almost sure (or high probability) convergence requires a completely different analysis so, beyond our averaging results (Theorems 4.13 and 6.2), there is very little overlap with these works.

In finite games, mirror-based techniques are closely related to the family of smooth (or perturbed) best response maps which have been studied extensively in models of stochastic fictitious play by Fudenberg and Levine [14], Hofbauer and Sandholm [17], and many others. In particular, in a discrete-time setting, Leslie and Collins [26] and Coucheney et al. [10] showed that a discounted, mirror-like scheme based only on observations of the players' realized, in-game payoffs converges to $\varepsilon$-equilibrium in potential games.[2] More recently, Mertikopoulos and Sandholm [30] showed that a broad class of continuous-time, mirror-based learning dynamics also eliminates dominated strategies and converges to strict equilibria from all nearby initial conditions; our analysis in Section 5 extends these results to a bona fide discrete-time, stochastic setting.

In games with continuous action sets, Perkins and Leslie [37] and Perkins et al. [38] recently examined the convergence properties of a related class of logit-based learning algorithms. The key difference between their approach and ours is that they focus on mixed-strategy learning and obtain convergence to $\varepsilon$-equilibria that assign positive weight on all (pure) strategies. Otherwise, with regard to pure-action learning in concave games, several authors have considered VI-based approaches, Gauss–Seidel best-response schemes, and Nikaido–Isoda relaxation methods for solving generalized Nash equilibrium problems; for a survey, see Facchinei and Kanzow [11] and Scutari et al. [43]. The intersection of these works with the current paper is when the game at hand satisfies a global monotonicity condition similar to the diagonal strict concavity condition of Rosen [41]; in this case VI methods converge to Nash equilibrium globally. However, these works do not consider the implications for the players' regret, the impact of imperfect information and/or local convergence/stability issues, so there is minimal overlap with our analysis.

Finally, during the final preparation stages of this manuscript (a few days before the actual submission date), we were made aware of a preprint by Bervoets et al. [4] where the authors consider pure-strategy learning in concave games with one-dimensional action sets, and they establish convergence to Nash equilibrium in ordinal potential games and games with strategic complements. A key feature of Bervoets et al. [4] is that players are assumed to observe only their realized, in-game payoffs, and they choose actions based on how their payoff has varied from the previous period. The resulting mean dynamics boil down to an interior-point,

[2] For a related treatment, see also Cominetti et al. [9] and Bravo [6].

primal variant of mirror-based learning induced by the entropic kernel $\theta(x) = x \log x$ (cf. Section 3), suggesting several interesting links with our paper.

## 2. Preliminaries

In this section, we present some basic elements from game theory. First, in Section 2.1, we define the class of concave games under consideration and we present a few examples thereof. Subsequently, in Section 2.2, we discuss some variational/geometric properties of Nash equilibria in concave games, and we introduce the notion of *variational stability* (VS), an analogue of evolutionary stability which plays a central role throughout our paper.

**Notation.** If $\mathcal{V}$ is a finite-dimensional real space with norm $\|\cdot\|$, the *conjugate* (or *dual*) *norm* on the dual space $\mathcal{V}^*$ of linear functionals on $\mathcal{V}$ is defined as $\|y\|_* = \sup\{\langle y|x\rangle : \|x\| \leq 1\}$, where $\langle y|x\rangle$ denotes the canonical pairing between $y \in \mathcal{V}^*$ and $x \in \mathcal{V}$. Also, if $\mathcal{C}$ is a closed convex subset of $\mathcal{V}$, the *tangent cone* $\mathrm{TC}_{\mathcal{C}}(x)$ to $\mathcal{C}$ at $x \in \mathcal{C}$ is defined as the closure of the set of all rays emanating from $x$ and intersecting $\mathcal{C}$ in at least one other point. Building on this, the *polar cone* $\mathrm{PC}_{\mathcal{C}}(x)$ to $\mathcal{C}$ at $x$ is defined as the polar cone of $\mathrm{TC}_{\mathcal{C}}(x)$, i.e. $\mathrm{PC}_{\mathcal{C}}(x) = \{y \in \mathcal{V}^* : \langle y|z\rangle \leq 0 \text{ for all } z \in \mathrm{TC}_{\mathcal{C}}(x)\}$. For concision, when $\mathcal{C}$ is understood from the context, we drop it altogether and write $\mathrm{TC}(x)$ and $\mathrm{PC}(x)$ instead. Finally, we write $\mathcal{C}^\circ \equiv \mathrm{ri}(\mathcal{C})$ for the relative interior of $\mathcal{C}$, $\|\mathcal{C}\| = \sup\{\|x' - x\| : x, x' \in \mathcal{C}\}$ for the diameter of $\mathcal{C}$, and $\mathrm{dist}(\mathcal{C}, x) = \inf_{x' \in \mathcal{C}} \|x' - x\|$ for the distance between $x \in \mathcal{V}$ and $\mathcal{C}$.

### 2.1. Concave games and examples.

Throughout this paper, we focus on games played by a finite set of *players* $i \in \mathcal{N} = \{1, \dots, N\}$, each of whom selects an *action* $x_i$ from a compact convex subset $\mathcal{X}_i$ of a finite-dimensional space $\mathcal{V}_i$. The players' rewards are then determined by their *action profile* $x = (x_1, \dots, x_N)$ which we often denote as $x \equiv (x_i; x_{-i})$ to highlight the action $x_i$ of player $i$ against the ensemble of actions $x_{-i} = (x_j)_{j \neq i}$ of all other players. Specifically, writing $\mathcal{X} \equiv \prod_i \mathcal{X}_i$ for the game's *action space*, each player's *payoff* is determined by an associated *payoff function* $u_i \colon \mathcal{X} \to \mathbb{R}$ which is assumed to be *individually concave*, i.e.

$$u_i(x_i; x_{-i}) \text{ is concave in } x_i \text{ for all } x_{-i} \in \textstyle\prod_{j \neq i} \mathcal{X}_j,\ i \in \mathcal{N}. \tag{2.1}$$

In terms of regularity, we also assume that $u_i$ is continuously differentiable in $x_i$ and we write

$$v_i(x) \equiv \nabla_{x_i} u_i(x_i; x_{-i}) \tag{2.2}$$

for the *individual gradient* of $u_i$ at $x$.[3]

Putting all this together, a *concave game* is a tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ with players, actions and payoffs defined as above. Below, we briefly discuss some widely studied examples of such games:

*Example* 2.1 (Mixed extensions of finite games). In a *finite game* $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$, each player $i \in \mathcal{N}$ chooses an action $s_i$ from a finite set $\mathcal{S}_i$ of "pure strategies" and no assumptions are made on the players' payoff functions $u_i \colon \mathcal{S} \equiv \prod_j \mathcal{S}_j \to \mathbb{R}$. Players can further "mix" these choices by playing *mixed strategies*, i.e. probability

[3] In the above, it is tacitly assumed that $u_i$ is defined on an open neighborhood of $\mathcal{X}_i$; doing so allows us to use ordinary differential calculus (instead of more advanced subgradient notions) but none of our results depend on this device. We also note that $v_i(x)$ acts naturally on vectors $z_i \in \mathcal{V}_i$ via the directional derivative mapping $z_i \mapsto \langle v_i(x)|z_i\rangle \equiv u_i'(x; z_i) = d/d\tau|_{\tau=0}\, u_i(x_i + \tau z_i; x_{-i})$. In view of this, $v_i(x)$ is treated throughout as an element of the dual space $\mathcal{V}_i^*$ of $\mathcal{V}_i$.

distributions $x_i \in \Delta(\mathcal{S}_i)$ over their pure strategies $s_i \in \mathcal{S}_i$. In this case (and in a slight abuse of notation), the expected payoff of player $i$ in the mixed profile $x = (x_1, \dots, x_N)$ is

$$u_i(x) = \sum_{s_1 \in \mathcal{S}_1} \cdots \sum_{s_N \in \mathcal{S}_N} u_i(s_1, \dots, s_N)\, x_{1,s_1} \cdots x_{N,s_N}, \tag{2.3}$$

so the players' individual gradients are simply their payoff vectors:

$$v_i(x) = \nabla_{x_i} u_i(x) = (u_i(s_i; x_{-i}))_{s_i \in \mathcal{S}_i}. \tag{2.4}$$

Writing $\mathcal{X}_i = \Delta(\mathcal{S}_i)$ for the players' mixed strategy spaces, we will refer to the game $\mathcal{G} = \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ as the *mixed extension* of the finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$. Since $\mathcal{X}_i$ is convex and $u_i$ is linear in $x_i$, $\mathcal{G}$ is clearly concave in the sense of (2.1).

*Example* 2.2 (Cournot competition). Consider the following asymmetric Cournot oligopoly: There is a finite set $\mathcal{N} = \{1, \dots, N\}$ of *firms*, each supplying the market with a quantity $x_i \in [0, C_i]$ of the same good (or service) up to the firm's production capacity $C_i$. The good is then priced as a decreasing function $P(x)$ of each firm's production; for concreteness, we focus on the linear model $P(x) = a - \sum_i b_i x_i$ where $a$ is a positive constant and the coefficients $b_i > 0$ reflect the price-setting power of each firm. In this oligopoly model, the utility of firm $i$ is given by

$$u_i(x) = x_i P(x) - c_i x_i, \tag{2.5}$$

where $c_i$ represents the marginal production cost of firm $i$. Letting $\mathcal{X}_i = [0, C_i]$, the resulting game $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ is easily seen to be concave in the sense of (2.1).

*Example* 2.3 (Atomic splittable congestion games). Congestion games are game-theoretic models that arise in the study of traffic networks (such as the Internet). To define them, fix a set of players $\mathcal{N}$ that share a set of *resources* $r \in \mathcal{R}$, each associated with a nondecreasing convex *cost function* $c_r \colon \mathbb{R}_+ \to \mathbb{R}$ (for instance, links in a data network and their corresponding delay functions). Each player $i \in \mathcal{N}$ has a certain *resource load* $\rho_i > 0$ which is split over a collection $\mathcal{S}_i \subseteq 2^{\mathcal{R}}$ of resource subsets $s_i$ of $\mathcal{R}$ – e.g. sets of links that form paths in the network. Accordingly, the action space of player $i \in \mathcal{N}$ is the scaled simplex $\mathcal{X}_i = \rho_i \Delta(\mathcal{S}_i) = \{x_i \in \mathbb{R}_+^{|\mathcal{S}_i|} : \sum_{s_i \in \mathcal{S}_i} x_{is_i} = \rho_i\}$ of *load distributions* over $\mathcal{S}_i$.

Given a distribution profile $x = (x_1, \dots, x_N)$, costs are determined based on the utilization of each resource as follows: First, the *demand* $w_r$ of the $r$-th resource is defined as the total load $w_r = \sum_{i \in \mathcal{N}} \sum_{s_i \ni r} x_{is_i}$ induced on said resource from all players. This demand is then assumed to incur a cost $c_r(w_r)$ per unit of load to each player utilizing resource $r$, where $c_r \colon \mathbb{R}_+ \to \mathbb{R}$ is a convex, nondecreasing function. Therefore, the aggregate cost to player $i \in \mathcal{N}$ is given by

$$c_i(x) = \sum_{s_i \in \mathcal{S}_i} x_{is_i} c_{is_i}(x), \tag{2.6}$$

where $c_{is_i}(x) = \sum_{r \in s_i} c_r(w_r)$ is the total cost incurred to player $i$ by the utilization of $s_i \subseteq \mathcal{R}$. The resulting *atomic splittable congestion game* $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, -c)$ is then easily seen to be concave in the sense of (2.1).

2.2. **Nash equilibria: characterization and stability.** Our analysis focuses primarily on *Nash equilibria* (NE), i.e. strategy profiles $x^* \in \mathcal{X}$ that discourage unilateral deviations in the sense that

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

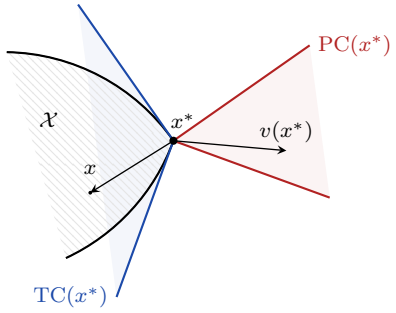


**Figure 1.** Geometric characterization of Nash equilibria in concave games.

With $u_i$ assumed concave and smooth in $x_i$, (NE) can be rewritten equivalently as

$$u_i'(x^*; z_i) = \langle v_i(x^*) \,|\, z_i \rangle \leq 0 \quad \text{for all } z_i \in \mathrm{TC}_i(x_i^*),\ i \in \mathcal{N}, \tag{2.7}$$

where $\mathrm{TC}_i(x_i^*)$ denotes the *tangent cone* to $\mathcal{X}_i$ at $x_i^*$. Thus, in geometric terms, $x^*$ is a Nash equilibrium if and only if each player's individual gradient $v_i(x^*)$ belongs to the *polar cone* $\mathrm{PC}_i(x_i^*)$ to $\mathcal{X}_i$ at $x_i^*$ (see also Fig. 1). Following Facchinei and Pang [12], this can be encoded even more concisely as follows:

**Proposition 2.1.** *$x^* \in \mathcal{X}$ is a Nash equilibrium if and only if $v(x^*) \in \mathrm{PC}(x^*)$, i.e.*

$$\langle v(x^*) \,|\, x - x^* \rangle \leq 0 \quad \textit{for all } x \in \mathcal{X}. \tag{2.8}$$

*Remark* 2.1. In the above (and what follows), $v = (v_i)_{i\in\mathcal{N}}$ denotes the collective profile of the players' individual payoff gradients and $\langle v \,|\, z \rangle \equiv \sum_{i\in\mathcal{N}} \langle v_i \,|\, z_i \rangle$ stands for the pairing between $v$ and the vector $z = (z_i)_{i\in\mathcal{N}} \in \prod_{i\in\mathcal{N}} \mathcal{V}_i$. For concision, we also write $\mathcal{V} \equiv \prod_i \mathcal{V}_i$ for the ambient space of $\mathcal{X} \equiv \prod_i \mathcal{X}_i$ and $\mathcal{V}^*$ for its dual.

Proposition 2.1 shows that Nash equilibria can be characterized as solutions to the variational inequality (2.8), so their existence follows from standard results. Using a similar variational characterization, Rosen [41] further provided the following sufficient condition for equilibrium uniqueness:

**Theorem 2.2** (Rosen, 1965)**.** *Assume that $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ satisfies the payoff monotonicity condition*

$$\langle v(x') - v(x) \,|\, x' - x \rangle \leq 0 \quad \textit{for all } x, x' \in \mathcal{X}, \tag{MC}$$

*with equality if and only if $x = x'$. Then, $\mathcal{G}$ admits a unique Nash equilibrium.*

*Remark* 2.2. In his original paper, Rosen [41] referred to (MC) as *diagonal strict concavity* (DSC). Later, in an evolutionary context, Hofbauer and Sandholm [18] used the term "stable" to describe games that satisfy a condition that is formally similar to (MC). More recently, Sandholm [42] has been advocating the use of the term "contractive" while Sorin and Wan [47] employ the term "dissipative". Our use of the term "monotonicity" is simply intended to highlight the fact that, under (MC), $v(x)$ is a *monotone operator* in the sense of convex/variational analysis – for a detailed discussion, see Rockafellar and Wets [40], Facchinei and Pang [12] and Facchinei and Kanzow [11].

Now, if (MC) holds and $x^*$ is a Nash equilibrium of $\mathcal{G}$, we readily obtain

$$\langle v(x) \,|\, x - x^* \rangle \leq \langle v(x^*) \,|\, x - x^* \rangle \leq 0, \tag{2.9}$$

where the second inequality follows from Proposition 2.1. In turn, (2.9) implies that the aggregate quantity $\sum_i u_i(x_i + t(x_i^* - x_i); x_{-i})$ increases with $t$ for small $t$, so, on average, players would tend to move unilaterally towards $x^*$. This variational property plays a key part in our analysis, so we formalize it as follows:

**Definition 2.3.** Let $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ be a concave game. A closed set $\mathcal{X}^* \subseteq \mathcal{X}$ is called *variationally stable* (or simply *stable*) if

$$\langle v(x) \,|\, x - x^* \rangle \leq 0 \quad \text{for all } x^* \in \mathcal{X}^* \text{ and all } x \text{ close to } \mathcal{X}^*, \tag{VS}$$

with equality if and only if $x \in \mathcal{X}^*$. In particular, if (VS) holds for all $x \in \mathcal{X}$, we say that $\mathcal{X}^*$ is *globally stable*.

As an immediate consequence of (2.9), we then get:

**Corollary 2.4.** *If a game satisfies* (MC)*, its* (*necessarily unique*) *Nash equilibrium is globally stable.*

The variational stability condition (VS) is strongly reminiscent of the seminal notion of *evolutionary stability* due to Maynard Smith and Price [28]. Specifically, if $v(x)$ denotes the payoff field of a population game and $\mathcal{X}^*$ is a singleton, (VS) is formally equivalent to the definition of an evolutionarily stable state (ESS).[4] As it turns out, just as evolutionary stability plays a crucial role in the convergence analysis of evolutionary dynamics, variational stability plays a similar part in ensuring the convergence of the class of learning schemes presented in the next section.

In addition, (VS) also has important consequences for the structure of the set of Nash equilibria of a concave game:

**Proposition 2.5.** *If $\mathcal{X}^* \subseteq \mathcal{X}$ is stable, it is an isolated convex set of Nash equilibria; in particular, if $\mathcal{X}^*$ is globally stable, the game admits no other equilibria.*

*Proof.* Assume $\mathcal{X}^*$ is stable, pick some $x^* \in \mathcal{X}^*$, and let $z_i = x_i' - x_i^*$ for some $x_i' \in \mathcal{X}_i$, $i \in \mathcal{N}$. Then, for all sufficiently small $\tau \geq 0$, (VS) gives

$$\frac{d}{d\tau} u_i(x_i^* + \tau z_i; x_{-i}^*) = \langle v_i(x_i^* + \tau z_i; x_{-i}^*) \,|\, z_i \rangle \leq 0. \tag{2.10}$$

Thus, letting $\tau \to 0$, (2.7) shows that $x^*$ is a Nash equilibrium (recall that $z_i = x_i' - x_i^*$ has been chosen arbitrarily).

Assume now that $x' \notin \mathcal{X}^*$ is a Nash equilibrium lying in a neighborhood $U$ of $\mathcal{X}^*$ where (VS) holds. By Proposition 2.1, we have $\langle v(x') \,|\, x - x' \rangle \leq 0$ for all $x \in \mathcal{X}$. However, since $x' \notin \mathcal{X}^*$, (VS) implies that $\langle v(x') \,|\, x^* - x' \rangle > 0$ for all $x^* \in \mathcal{X}^*$, a contradiction. We conclude that there are no other equilibria in $U$, i.e. $\mathcal{X}^*$ is an isolated set of Nash equilibria; the global version of our claim then follows by taking $U = \mathcal{X}$.

Finally, to prove the convexity of $\mathcal{X}^*$, take $x_0^*, x_1^* \in \mathcal{X}^*$ and set $x_\lambda^* = (1-\lambda)x_0^* + \lambda x_1^*$ for $\lambda \in [0,1]$. Substituting in (VS), we get $\langle v(x_\lambda^*) \,|\, x_\lambda^* - x_0^* \rangle = \lambda \langle v(x_\lambda^*) \,|\, x_1^* - x_0^* \rangle \leq 0$ and $\langle v(x_\lambda^*) \,|\, x_\lambda^* - x_1^* \rangle = -(1-\lambda)\langle v(x_\lambda^*) \,|\, x_1^* - x_0^* \rangle \leq 0$, showing that $\langle v(x_\lambda^*) \,|\, x_1^* - x_0^* \rangle = 0$. By (VS), this means that $x_\lambda^* \in \mathcal{X}^*$ for all $\lambda \in [0,1]$, i.e. $\mathcal{X}^*$ is convex. ∎

[4] This concise characterization of ESSs is due to Hofbauer et al. [19] and Taylor [49].

We close this section with a second derivative test that can be used to verify whether (VS) holds. To state it, define the *Hessian* of a game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ as the block matrix $H^{\mathcal{G}}(x) = (H^{\mathcal{G}}_{ij}(x))_{i,j\in\mathcal{N}}$ with

$$H^{\mathcal{G}}_{ij}(x) = \tfrac{1}{2}\nabla_{x_j}\nabla_{x_i}u_i(x) + \tfrac{1}{2}(\nabla_{x_i}\nabla_{x_j}u_j(x))^{\top}. \tag{2.11}$$

We then have:

**Proposition 2.6.** *If $H^{\mathcal{G}}(x)$ is negative-definite on $\mathrm{TC}(x)$ for all $x \in \mathcal{X}$, the game admits a unique, globally stable Nash equilibrium. More generally, if $x^*$ is a Nash equilibrium of $\mathcal{G}$ and $H^{\mathcal{G}}(x^*) \prec 0$ on $\mathrm{TC}(x^*)$, $x^*$ is stable and isolated.*

*Proof.* Assume first that $H^{\mathcal{G}}(x) \prec 0$ on $\mathrm{TC}(x)$ for all $x \in \mathcal{X}$. By Theorem 6 in Rosen [41], $\mathcal{G}$ satisfies (MC) so our claim follows from Corollary 2.4. For our second claim, if $H^{\mathcal{G}}(x^*) \prec 0$ on $\mathrm{TC}(x^*)$ for some Nash equilibrium $x^*$ of $\mathcal{G}$, we also have $H^{\mathcal{G}}(x) \prec 0$ for all $x$ in a product neighborhood $U = \prod_{i\in\mathcal{N}} U_i$ of $x^*$ in $\mathcal{X}$. The above reasoning shows that $x^*$ is the unique equilibrium of the restricted game $\mathcal{G}|_U(\mathcal{N}, U, u|_U)$, so $x^*$ is locally stable and isolated in $\mathcal{G}$. ∎

We provide two straightforward applications of Proposition 2.6 below:

*Example* 2.4 (Potential games). Following Monderer and Shapley [31], a game $\mathcal{G}$ is called a *potential game* if there exists a *potential function* $V\colon \mathcal{X} \to \mathbb{R}$ such that

$$u_i(x_i; x_{-i}) - u_i(x'_i; x_{-i}) = V(x_i; x_{-i}) - V(x'_i; x_{-i}) \quad \text{for all } x, x' \in \mathcal{X},\ i \in \mathcal{N}. \tag{2.12}$$

In potential games, local maximizers of $V$ are Nash equilibria and the converse also holds if $V$ is concave – cf. Neyman [36]. Moreover, by differentiating (2.12), it is easy to see that the Hessian of a potential game $\mathcal{G}$ is equal to the Hessian of its potential. Hence, if a game admits a concave potential $V$, the game's Nash set $\mathcal{X}^* = \arg\max_{x\in\mathcal{X}} V(x)$ is globally stable.

*Example* 2.5 (Cournot revisited). Consider again the Cournot oligopoly model of Example 2.2. A simple differentiation yields

$$H^{\mathcal{G}}_{ij}(x) = \frac{1}{2}\frac{\partial^2 u_i}{\partial x_i \partial x_j} + \frac{1}{2}\frac{\partial^2 u_j}{\partial x_j \partial x_i} = -b_i\delta_{ij} - \tfrac{1}{2}(b_i + b_j), \tag{2.13}$$

where $\delta_{ij} = \mathbb{1}\{i = j\}$ is the Kronecker delta. This shows that a Cournot oligopoly admits a unique, globally stable equilibrium whenever the RHS of (2.13) is negative-definite. This is always the case if the model is symmetric ($b_i = b$ for all $i \in \mathcal{N}$), but not necessarily otherwise: quantitatively, if the coefficients $b_i$ are independent and identically distributed (i.i.d.) on $[0, 1]$, a Monte Carlo simulation shows that (2.13) is negative-definite with probability between 65% and 75% for $N \in \{2, \dots, 100\}$.

## 3. A class of mirror-based learning schemes

In this section, we present a distributed learning scheme based on the method of mirror descent, a widely used optimization procedure pioneered by Nemirovski and Yudin [33] and studied further by (among others) Beck and Teboulle [2], Nesterov [35], Nemirovski et al. [32], Juditsky et al. [21] and Shalev-Shwartz [44]. Intuitively, the main idea of the method is as follows: At each stage $n = 0, 1, \dots$, each player $i \in \mathcal{N}$ estimates the individual gradient $v_i(x_n)$ of their payoff function at the current action profile $x_n \in \mathcal{X}$, possibly subject to noise and uncertainty. Subsequently, every player takes a step along this estimated gradient in the dual space $\mathcal{V}_i^*$ (where

gradients live), and they "mirror" the output back to the primal space $\mathcal{X}_i$ in order to choose an action $x_{i,n+1}$ for the next stage and continue playing.

Formally, this multi-agent *mirror-based learning* (ML) scheme can be written as

$$\begin{aligned} y_{i,n+1} &= y_{i,n} + \gamma_n \hat{v}_{i,n}, \\ x_{i,n+1} &= Q_i(y_{i,n+1}), \end{aligned} \tag{ML}$$

where:

1) $n = 0, 1, \dots$ denotes the stage of the process.
2) $\hat{v}_{i,n} \in \mathcal{V}_i^*$ is a stochastic estimate of the individual payoff gradient $v_i(x_n)$ of player $i$ at stage $n$ (more on this below).
3) $y_{i,n} \in \mathcal{V}_i^*$ is an auxiliary "score" variable that aggregates the player's individual gradient steps up to stage $n$.
4) $\gamma_n > 0$ is a nonincreasing step-size sequence, typically of the form $1/(n+1)^\beta$ for some $\beta \in [0, 1]$.
5) $Q_i \colon \mathcal{V}_i^* \to \mathcal{X}_i$ is the *mirror* (or *choice*) *map* that outputs the $i$-th player's action as a function of their score vector $y_i$ (see below for a rigorous definition).

In view of the above, the core components of (ML) are *a*) the players' gradient estimates $\hat{v}_i$; and *b*) the choice maps $Q_i$ that determine the players' actions. In the rest of this section, we discuss both in detail.

3.1. **Feedback and uncertainty.** Regarding the players' gradient observations, we will be assuming that each player $i \in \mathcal{N}$ has access to a "black box" feedback mechanism – an *oracle* – which returns an estimate $\hat{v}_i$ of the player's individual gradient $v_i(x)$ at a given action profile $x$. Of course, this information may be imperfect for a multitude of reasons: for instance *i*) gradient estimates may be susceptible to random measurement errors; *ii*) the transmission of this information could be subject to noise; and/or *iii*) the game's payoff functions may be stochastic expectations of the form

$$u_i(x) = \mathbb{E}[\hat{u}_i(x; \omega)] \quad \text{for some random variable } \omega, \tag{3.1}$$

and players may only observe the realized gradients $\nabla_{x_i} \hat{u}_i(x; \omega)$ of $\hat{u}_i(x; \omega)$.

With all this in mind, we will focus on the general model

$$\hat{v}_{i,n} = v_i(x_n) + \xi_{i,n}, \tag{3.2}$$

where the noise process $\xi_n = (\xi_{i,n})_{i \in \mathcal{N}}$ satisfies the following statistical hypotheses:

1. *Zero-mean:*

$$\mathbb{E}[\xi_n \,|\, \mathcal{F}_n] = 0 \quad \text{for all } n = 0, 1, \dots \text{ (a.s.).} \tag{H1}$$

2. *Finite mean squared error:* there exists some $\sigma_* \geq 0$ such that

$$\mathbb{E}[\|\xi_n\|_*^2 \,|\, \mathcal{F}_n] \leq \sigma_*^2 \quad \text{for all } n = 0, 1, \dots \text{ (a.s.).} \tag{H2}$$

In the above, $\mathcal{F}_n$ denotes the natural filtration induced by $x_n$, i.e. the history of $x_n$ up to stage $n$. Thus, (H1) and (H2) simply posit that the players' individual gradient estimates are *conditionally unbiased and bounded in mean square*, viz.

$$\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = v(x_n), \tag{3.3a}$$

$$\mathbb{E}[\|\hat{v}_n\|_*^2 \,|\, \mathcal{F}_n] \leq V_*^2 \quad \text{for some finite } V_* > 0 \text{ (a.s.).} \tag{3.3b}$$

The above allows for a broad range of error distributions, including all compactly supported, (sub-)Gaussian, (sub-)exponential and log-normal distributions.[5] In fact, both hypotheses can be relaxed (for instance, by assuming a vanishing bias or asking for finite moments up to order $q < 2$), but we do not do so for simplicity.

3.2. **Choosing actions.** Since the players' score variables $y_i$ aggregate gradient steps, a reasonable choice for $Q_i$ would be the arg max correspondence $y_i \mapsto \arg\max_{x_i \in \mathcal{X}_i} \langle y_i \,|\, x_i \rangle$ that outputs those actions which are most closely aligned with $y_i$. Notwithstanding, there are two problems with this approach: *a*) this assignment is too aggressive in the presence of uncertainty; and *b*) generically, the output action would be an extreme point of $\mathcal{X}$ so (ML) could never converge to an interior point. Thus, instead of taking a "hard" arg max approach, we will focus on regularized choice maps of the form

$$y_i \mapsto \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i \,|\, x_i \rangle - h(x_i)\}, \tag{3.4}$$

where the "regularization penalty" $h_i \colon \mathcal{X}_i \to \mathbb{R}$ satisfies the following requirements:

**Definition 3.1.** Let $\mathcal{C}$ be a compact convex subset of a finite-dimensional normed space $\mathcal{V}$. We say that $h \colon \mathcal{C} \to \mathbb{R}$ is a *penalty function* (or *regularizer*) on $\mathcal{C}$ if:

(1) $h$ is continuous.
(2) $h$ is *strongly convex*, i.e. there exists some $K > 0$ such that

$$h(tx + (1-t)x') \leq th(x) + (1-t)h(x') - \tfrac{1}{2}Kt(1-t)\|x' - x\|^2 \tag{3.5}$$

for all $x, x' \in \mathcal{C}$ and all $t \in [0,1]$.

The *choice* (or *mirror*) *map* $Q \colon \mathcal{V}^* \to \mathcal{C}$ *induced by* $h$ is then defined as

$$Q(y) = \arg\max\{\langle y \,|\, x \rangle - h(x) : x \in \mathcal{C}\}. \tag{3.6}$$

In what follows, we will be assuming that each player $i \in \mathcal{N}$ is endowed with an individual penalty function $h_i \colon \mathcal{X}_i \to \mathbb{R}$ that is $K_i$-strongly convex. Furthermore, to emphasize the interplay between primal and dual variables (the players' actions $x_i$ and their score vectors $y_i$ respectively), we will write $\mathcal{Y}_i \equiv \mathcal{V}_i^*$ for the dual space of $\mathcal{V}_i$ and $Q_i \colon \mathcal{Y}_i \to \mathcal{X}_i$ for the choice map induced by $h_i$. More concisely, this information can be encoded in the aggregate penalty function $h(x) = \sum_i h_i(x_i)$ with associated strong convexity constant $K \equiv \min_i K_i$.[6] The induced choice map is simply $Q \equiv (Q_1, \dots, Q_N)$ so we will write $x = Q(y)$ for the action profile induced by the score vector $y = (y_1, \dots, y_N) \in \mathcal{Y} \equiv \prod_i \mathcal{Y}_i$.

*Remark* 3.1. When $\mathcal{G}$ is the mixed extension of a finite game, McKelvey and Palfrey [29] originally referred to $Q_i$ as a "quantal response function" (the notation $Q$ alludes precisely to this terminology). In the same game-theoretic context, the composite map $Q_i \circ v_i$ is often called a smooth, perturbed, or regularized best response; for a detailed discussion, see Fudenberg and Levine [14], Hofbauer and Sandholm [17], and Mertikopoulos and Sandholm [30]. Finally, in online learning and optimization, $h$ is usually referred to as a "Bregman" or "prox" function, and the induced regularization process is variously known as a "link", "softmax", or "Bregman/prox projection"; for a comprehensive account, see Nemirovski and Yudin [33], Shalev-Shwartz [44] and references therein.

[5] In particular, we will not be assuming i.i.d. errors; this point is crucial for applications to distributed control where measurements are typically correlated with the state of the system.

[6] We assume here that $\mathcal{V} \equiv \prod_i \mathcal{V}_i$ is endowed with the product norm $\|x\|_{\mathcal{V}}^2 = \sum_i \|x_i\|_{\mathcal{V}_i}^2$.

**Algorithm 1.** Learning with lazy Euclidean projections (Example 3.1).

**Parameter:** step-size $\gamma_n \propto 1/n^\beta$, $0 < \beta \leq 1$.
**Initialization:** $n \leftarrow 0$; $y \leftarrow$ arbitrary.
**Repeat**
  $n \leftarrow n + 1$;
  **foreach** player $i \in \mathcal{N}$ **do**
    project $x_i \leftarrow \Pi_{\mathcal{X}_i}(y_i)$; # choose actions
    observe $\hat{v}_i$; # estimate gradients
    update $y_i \leftarrow y_i + \gamma_n \hat{v}_i$; # gradient step
  **until** convergence

We discuss below a few examples of this regularization process; for a more general treatment, see Kiwiel [23], Bolte and Teboulle [5] and Alvarez et al. [1].

*Example* 3.1 (Euclidean projections). Let $h(x) = \frac{1}{2}\|x\|_2^2$. Then, $h$ is 1-strongly convex with respect to $\|\cdot\|_2$ and the corresponding choice map is the closest point projection

$$\Pi_{\mathcal{X}}(y) \equiv \arg\max_{x\in\mathcal{X}} \{\langle y \,|\, x\rangle - \tfrac{1}{2}\|x\|_2^2\} = \arg\min_{x\in\mathcal{X}} \|y - x\|_2^2. \tag{3.7}$$

The induced learning scheme (cf. Algorithm 1) may thus be treated as a multi-agent variant of gradient ascent with lazy projections – cf. Zinkevich [53], Beck and Teboulle [2], and Shalev-Shwartz [44]. For future reference, note that $h$ is differentiable on $\mathcal{X}$ and $\Pi_{\mathcal{X}}$ is *surjective* (i.e. $\operatorname{im}\Pi_{\mathcal{X}} = \mathcal{X}$).

*Example* 3.2 (Entropic regularization). Motivated by mixed strategy learning in finite games (Example 2.1), let $\Delta = \{x \in \mathbb{R}^d_+ : \sum_{j=1}^d x_j = 1\}$ denote the unit simplex of $\mathbb{R}^d$. Then, a standard regularizer on $\Delta$ is provided by the (negative) Gibbs entropy

$$h(x) = \sum_{j=1}^d x_j \log x_j. \tag{3.8}$$

This penalty function is 1-strongly convex with respect to the $\ell_1$ norm on $\mathbb{R}^d$ and a straightforward calculation shows that the induced choice map is

$$\Lambda(y) = \frac{1}{\sum_{j=1}^d \exp(y_j)} (\exp(y_1), \dots, \exp(y_d)). \tag{3.9}$$

This model is known as *logit choice* and the associated learning scheme has been studied extensively in evolutionary game theory and online learning (where it is sometimes referred to as "hedging"); for a detailed account, see Vovk [52], Littlestone and Warmuth [27], Fudenberg and Levine [14], Laraki and Mertikopoulos [25] and references therein. In contrast to the previous example, $h$ is differentiable only on the relative interior $\Delta^\circ$ of $\Delta$ and $\operatorname{im}\Lambda = \Delta^\circ$ (i.e. $\Lambda$ is "essentially" surjective).

3.3. **Choice map surjectivity and steepness.** We close this section with an important dichotomy concerning the boundary behavior of penalty functions and the induced choice maps. To describe it, it will be convenient to treat $h$ as an

extended-real-valued function $h\colon \mathcal{V} \to \mathbb{R} \cup \{+\infty\}$ by setting $h = +\infty$ outside $\mathcal{X}$. The *subdifferential* of $h$ at $x \in \mathcal{V}$ is then defined as

$$\partial h(x) = \{y \in \mathcal{V}^* : h(x') \geq h(x) + \langle y | x' - x \rangle \text{ for all } x' \in \mathcal{V}\}, \tag{3.10}$$

and we say that $h$ is *subdifferentiable* at $x \in \mathcal{X}$ whenever $\partial h(x)$ is nonempty. This is always the case if $x \in \mathcal{X}^\circ$, so we have $\mathcal{X}^\circ \subseteq \operatorname{dom} \partial h \equiv \{x \in \mathcal{X} : \partial h(x) \neq \varnothing\} \subseteq \mathcal{X}$ (Rockafellar [39, Chap. 26]).

Intuitively, $h$ fails to be subdifferentiable at a boundary point $x \in \operatorname{bd}(\mathcal{X})$ if it becomes "infinitely steep" near $x$. We thus say that $h$ is *steep* at $x$ whenever $x \notin \operatorname{dom} h$; otherwise, $h$ is said to be *nonsteep* at $x$. The following result shows that penalty functions that are everywhere nonsteep (as in Example 3.1) induce choice maps that are surjective; on the other hand, penalty functions that are everywhere steep (cf. Example 3.2) induce choice maps that are interior-valued:

**Proposition 3.2.** *Let $h$ be a $K$-strongly convex penalty function with induced choice map $Q\colon \mathcal{Y} \to \mathcal{X}$, and let $h^*\colon \mathcal{Y} \to \mathbb{R}$ be the convex conjugate of $h$, i.e.*

$$h^*(y) = \max\{\langle y | x \rangle - h(x) : x \in \mathcal{X}\}, \quad y \in \mathcal{Y}. \tag{3.11}$$

*Then:*

1) *$x = Q(y)$ if and only if $y \in \partial h(x)$; in particular, $\operatorname{im} Q = \operatorname{dom} \partial h$.*
2) *$h^*$ is differentiable on $\mathcal{Y}$ and $\nabla h^*(y) = Q(y)$ for all $y \in \mathcal{Y}$.*
3) *$Q$ is $(1/K)$-Lipschitz continuous.*

Proposition 3.2 is essentially part of the oral tradition in optimization and convex analysis; for a rigorous proof of the various statements, see Rockafellar [39, Theorem 23.5] and Rockafellar and Wets [40, Theorem 12.60(b)].

## 4. Convergence analysis

A first important property of the mirror-based learning scheme (ML) is that it leads to *no regret*, viz.

$$\max_{x_i \in \mathcal{X}_i} \sum_{k=0}^{n} [u_i(x_i; x_{-i,n}) - u_i(x_n)] = o(n) \quad \text{for all } i \in \mathcal{N}, \tag{4.1}$$

provided that (H1)–(H2) hold and $\gamma_n$ is chosen appropriately – for a precise statement, see Shalev-Shwartz [44] and Kwon and Mertikopoulos [24]. In other words, (ML) has the desirable attribute that, in the long run, every player's average payoff matches – or exceeds – that of the best fixed action in hindsight.

In this section, we expand on this worst-case performance guarantee and we derive some general convergence results for the sequence of play induced by (ML). Specifically, in Section 4.1 we show that, if it exists, the limit of (ML) is a Nash equilibrium (a.s.). Subsequently, to obtain stronger convergence results, we introduce in Section 4.2 the so-called *Fenchel coupling*, a "divergence" measure between primal and dual variables – that is, between the players' actions $x_i \in \mathcal{X}_i$ and their score vectors $y_i \in \mathcal{Y}_i$. Using this coupling as a primal-dual Lyapunov function, we show in Sections 4.3 and 4.4 that globally (resp. locally) stable states are globally (resp. locally) attracting under (ML). Finally, in Section 4.5, we examine the convergence properties of (ML) in zero-sum games.

4.1. **Termination states.** Our first result is that if the sequence of play induced by (ML) converges, its limit is a Nash equilibrium:

**Theorem 4.1.** *Suppose that* (ML) *is run with imperfect gradient information satisfying* (H1)–(H2) *and a step-size sequence* $\gamma_n$ *such that*

$$\sum_{n=0}^{\infty} \left(\frac{\gamma_n}{\tau_n}\right)^2 < \sum_{n=0}^{\infty} \gamma_n = \infty, \tag{4.2}$$

*where* $\tau_n = \sum_{k=0}^{n} \gamma_k$. *If* $x_n \to x^*$ *as* $n \to \infty$, *then* $x^*$ *is a Nash equilibrium* (*a.s.*).

*Remark* 4.1. The summability requirement (4.2) is fairly mild and holds for every step-size policy of the form $\gamma_n \propto (n+1)^{-\beta}$, $\beta \leq 1$ (i.e. even for *increasing* $\gamma_n$).

*Proof of Theorem 4.1.* Let $v^* = v(x^*)$ and assume ad absurdum that $x^*$ is not a Nash equilibrium. Then, by the characterization (2.7) of Nash equilibria, there exists a player $i \in \mathcal{N}$ and a unilateral deviation $q_i \in \mathcal{X}_i$ such that $\langle v_i^* \,|\, q_i - x_i^* \rangle > 0$. Thus, by continuity, there exists some $a > 0$ and neighborhoods $U$, $V$ of $x^*$ and $v^*$ respectively, such that

$$\langle v_i' \,|\, q_i - x_i' \rangle \geq a \tag{4.3}$$

whenever $x' \in U$ and $v' \in V$.

Since $x_n \to x^*$, we may assume without loss of generality that $x_n \in U$ and $v(x_n) \in V$ for all $n$. Then, (ML) yields

$$y_{n+1} = y_0 + \sum_{k=0}^{n} \gamma_k \hat{v}_k = y_0 + \tau_n \bar{v}_n, \tag{4.4}$$

where we have set $\bar{v}_n = \tau_n^{-1} \sum_{k=0}^{n} \gamma_k \hat{v}_k = \tau_n^{-1} \sum_{k=0}^{n} \gamma_k \left[v(x_k) + \xi_k\right]$. We now claim that $\bar{v}_n \to v^*$ (a.s.). Indeed, combining (4.2) with (H2), we get

$$\sum_{n=0}^{\infty} \frac{1}{\tau_n^2} \mathbb{E}[\|\gamma_n \xi_n\|_*^2 \,|\, \mathcal{F}_n] \leq \sum_{n=0}^{\infty} \frac{\gamma_n^2}{\tau_n^2} \sigma_*^2 < \infty. \tag{4.5}$$

Therefore, by the law of large numbers for martingale differences (Hall and Heyde [15, Theorem 2.18]), we obtain $\tau_n^{-1} \sum_{k=0}^{n} \gamma_k \xi_k \to 0$ (a.s.). Moreover, since $v(x_n) \to v^*$ by assumption, we infer that $\bar{v}_n \to v^*$ as well.

Thus, given that $y_{i,n} \in \partial h_i(x_{i,n})$ by Proposition 3.2, we get

$$h_i(q_i) - h_i(x_{i,n}) \geq \langle y_{i,n} \,|\, q_i - x_{i,n} \rangle = \langle y_{i,0} \,|\, q_i - x_{i,n} \rangle + \tau_n \langle \bar{v}_{i,n} \,|\, q_i - x_{i,n} \rangle. \tag{4.6}$$

Since $\bar{v}_n \to v^*$, (4.3) readily yields $\langle \bar{v}_{i,n} \,|\, q_i - x_{i,n} \rangle \geq a > 0$ for all sufficiently large $n$. By substituting in (4.6), we then obtain $h_i(q_i) - h_i(x_{i,n}) \gtrsim a\tau_n \nearrow \infty$ (a.s.), a contradiction which implies that $x^*$ is a Nash equilibrium. ∎

4.2. **The Fenchel coupling.** A key tool in establishing the convergence properties of mirror descent schemes is the so-called *Bregman divergence* $D(p, x)$ between a given state $x \in \mathcal{X}$ and a base point $p \in \mathcal{X}$. Following Kiwiel [23], this is defined as the difference between $h(p)$ and the best linear approximation of $h(p)$ from $x$, i.e.

$$D(p, x) = h(p) - h(x) - h'(x; p - x), \tag{4.7}$$

where $h'(x; z) = \lim_{t \to 0^+} t^{-1}[h(x + tz) - h(x)]$ denotes the one-sided derivative of $h$ at $x$ along $z \in \mathrm{TC}(x)$. Owing to the (strict) convexity of $h$, we have $D(p, x) \geq 0$ and $x_n \to p$ whenever $D(p, x_n) \to 0$ (Kiwiel [23]); as a result, the convergence of a

sequence $x_n$ to a target point $p$ can be checked directly by means of the associated divergence $D(p, x_n)$.

Nevertheless, it is often impossible to glean any useful information on $D(p, x_n)$ from (ML) when $x_n = Q(y_n)$ is not interior. By this token, given that (ML) alternates between primal and dual variables (actions and scores respectively), it will be more convenient to use the following "primal-dual divergence" between dual vectors $y \in \mathcal{Y}$ and base points $p \in \mathcal{X}$:

**Definition 4.2.** Let $h\colon \mathcal{X} \to \mathbb{R}$ be a penalty function on $\mathcal{X}$. Then, the *Fenchel coupling* induced by $h$ is defined as

$$F(p, y) = h(p) + h^*(y) - \langle y \,|\, p \rangle \quad \text{for all } p \in \mathcal{X},\ y \in \mathcal{Y}. \tag{4.8}$$

The terminology "Fenchel coupling" is due to Mertikopoulos and Sandholm [30] and refers to the fact that (4.8) collects all terms of Fenchel's inequality. As a result, $F(p, y)$ is nonnegative and strictly convex in both arguments (though not jointly so). Moreover, it enjoys the following core properties:

**Proposition 4.3.** *Let $h$ be a $K$-strongly convex penalty function on $\mathcal{X}$. Then, for all $p \in \mathcal{X}$ and all $y, y' \in \mathcal{Y}$, we have:*

$$a)\ \ F(p, y) = D(p, Q(y)) \ \text{ if } Q(y) \in \mathcal{X}^\circ \ (\text{but not necessarily otherwise}). \tag{4.9a}$$

$$b)\ \ F(p, y) \geq \tfrac{1}{2} K \left\| Q(y) - p \right\|^2. \tag{4.9b}$$

$$c)\ \ F(p, y') \leq F(p, y) + \langle y' - y \,|\, Q(y) - p \rangle + \tfrac{1}{2K} \|y' - y\|_*^2. \tag{4.9c}$$

Proposition 4.3 (proven in Appendix A) justifies the description "primal-dual divergence" for $F(x^*, y)$ and plays a key role in our analysis. Specifically, the lower bound (4.9b) yields $Q(y_n) \to p$ for every sequence $y_n$ such that $F(p, y_n) \to 0$, so $F(p, y_n)$ can be used to check convergence of $Q(y_n)$ to $p$. For technical reasons, it is convenient to assume that the converse also holds, namely

$$F(p, y_n) \to 0 \quad \text{whenever} \quad Q(y_n) \to p. \tag{H3}$$

When (H3) holds (Examples 3.1 and 3.2 both satisfy it), Proposition 4.3 gives:

**Corollary 4.4.** *Under (H3), $F(p, y_n) \to 0$ if and only if $Q(y_n) \to p$.*

Finally, to extend the above to subsets of $\mathcal{X}$, we define the setwise coupling

$$F(\mathcal{C}, y) = \inf\{F(p, y) : p \in \mathcal{C}\}, \quad \mathcal{C} \subseteq \mathcal{X},\ y \in \mathcal{Y}. \tag{4.10}$$

In analogy to the pointwise case, we then have:

**Proposition 4.5.** *Let $\mathcal{C}$ be a closed subset of $\mathcal{X}$. Then, $Q(y_n) \to \mathcal{C}$ whenever $F(\mathcal{C}, y_n) \to 0$; in addition, the converse is also true under (H3).*

The proof of Proposition 4.5 is a straightforward exercise in point-set topology so we omit it. What's more important is that, thanks to Proposition 4.5, the Fenchel coupling can also be used to test for convergence to a set; in what follows, we employ this property freely.

4.3. **Global convergence.** In this section, we focus on globally stable Nash equilibria (and sets thereof). We begin with the perfect feedback case:

**Theorem 4.6.** *Suppose that (ML) is run with perfect gradient information ($\sigma_* = 0$), choice maps satisfying (H3), and a vanishing step-size sequence $\gamma_n$ such that $\sum_{k=0}^n \gamma_k^2 / \sum_{k=0}^n \gamma_k \to 0$. If $\mathcal{X}^*$ is globally stable, $x_n$ converges to $\mathcal{X}^*$.*

| | Hypothesis | Precise statement |
|---|---|---|
| (H1) | Zero-mean errors | $\mathbb{E}[\xi_n \,\vert\, \mathcal{F}_n] = 0$ |
| (H2) | Finite error variance | $\mathbb{E}[\|\xi_n\|_*^2 \,\vert\, \mathcal{F}_n] \leq \sigma_*^2$ |
| (H3) | Regular choice maps | $F(p, y_n) \to 0$ whenever $Q(y_n) \to p$ |
| (H4) | Lipschitz gradients | $v(x)$ is Lipschitz continuous |

**Table 1.** Overview of the various regularity hypotheses used in the paper.

*Proof.* Fix some $\varepsilon > 0$ and let $U_\varepsilon = \{x = Q(y) : F(\mathcal{X}^*, y) < \varepsilon\}$. Then, by Proposition 4.5, it suffices to show that $x_n \in U_\varepsilon$ for all sufficiently large $n$.

To that end, for all $x^* \in \mathcal{X}^*$, Proposition 4.3 yields

$$F(x^*, y_{n+1}) \leq F(x^*, y_n) + \gamma_n \langle v(x_n) \,\vert\, x_n - x^* \rangle + \frac{\gamma_n^2}{2K} \|v(x_n)\|_*^2. \tag{4.11}$$

Assume now by induction that $x_n \in U_\varepsilon$. By (H3), there exists some $\delta > 0$ such that $\mathrm{cl}(U_{\varepsilon/2})$ contains a $\delta$-neighborhood of $\mathcal{X}^*$.[7] Consequently, with $\mathcal{X}^*$ assumed globally stable, there exists some $a \equiv a(\varepsilon) > 0$ such that

$$\langle v(x) \,\vert\, x - x^* \rangle \leq -a \quad \text{for all } x \in U_\varepsilon - U_{\varepsilon/2},\ x^* \in \mathcal{X}^*. \tag{4.12}$$

Therefore, if $x_n \in U_\varepsilon - U_{\varepsilon/2}$ and $\gamma_n \leq 2Ka/V_*^2$, (4.11) yields $F(x^*, y_{n+1}) \leq F(x^*, y_n)$. Hence, minimizing over $x^* \in \mathcal{X}^*$, we get $F(\mathcal{X}^*, y_{n+1}) \leq F(\mathcal{X}^*, y_n) < \varepsilon$, so $x_{n+1} \in U_\varepsilon$. Otherwise, if $x_n \in U_{\varepsilon/2}$ and $\gamma_n^2 < K\varepsilon/V_*^2$, combining (VS) with (4.11) yields $F(x^*, y_{n+1}) \leq F(x^*, y_n) + \varepsilon/2$ so, again, $F(\mathcal{X}^*, y_{n+1}) \leq F(\mathcal{X}^*, y_n) + \varepsilon/2 \leq \varepsilon$, i.e. $x_{n+1} \in U_\varepsilon$. We thus conclude that $x_{n+1} \in U_\varepsilon$ whenever $x_n \in U_\varepsilon$ and $\gamma_n < \min\{2Ka/V_*^2, \sqrt{K\varepsilon}/V_*\}$.

To complete the proof, Lemma A.3 shows that $x_n$ visits $U_\varepsilon$ infinitely often under the stated assumptions. Since $\gamma_n \searrow 0$, our assertion follows. ∎

We next show that Theorem 4.6 extends to the case of imperfect feedback under the additional regularity requirement:

The gradient field $v(x)$ is Lipschitz continuous. (H4)

With this extra assumption, we have:

**Theorem 4.7.** *Suppose that (ML) is run with a step-size sequence $\gamma_n$ such that $\sum_{n=0}^\infty \gamma_n^2 < \sum_{n=0}^\infty \gamma_n = \infty$. If (H1)–(H4) hold and $\mathcal{X}^*$ is a globally stable set of Nash equilibria, we have $x_n \to \mathcal{X}^*$ (a.s.).*

**Corollary 4.8.** *If $\mathcal{G}$ satisfies the payoff monotonicity condition (MC), $x_n$ converges to the (necessarily unique) Nash equilibrium of $\mathcal{G}$ (a.s.).*

**Corollary 4.9.** *If $\mathcal{G}$ admits a concave potential, $x_n$ converges to the set of Nash equilibria of $\mathcal{G}$ (a.s.).*

Because of the noise affecting the players' gradient estimates, our proof strategy for Theorem 4.7 is quite different from that of Theorem 4.6. In particular, instead of working directly in discrete time, we first consider the continuous-time system

$$\begin{aligned} \dot{y} &= v(x), \\ x &= Q(y), \end{aligned} \tag{ML-C}$$

[7] Indeed, if this were not the case, there would exist a sequence $y_n'$ in $\mathcal{Y}$ such that $Q(y_n') \to \mathcal{X}^*$ but $F(\mathcal{X}^*, y_n') \geq \varepsilon/2$, in contradiction to (H3).

which can be seen as a "mean-field" approximation of the recursive scheme (ML). As it turns out, the orbits $x(t) = Q(y(t))$ of (ML-C) converge to $\mathcal{X}^*$ in a certain, "uniform" way. Moreover, under the assumptions of Theorem 4.7, the sequence $y_n$ generated by the discrete-time, stochastic process (ML) comprises an *asymptotic pseudotrajectory* (APT) of the dynamics (ML-C) in the sense of Benaïm [3]. APTs have the key property that, in the presence of a globally attracting set, they cannot stray too far from the flow of (ML-C); however, given that $Q$ often fails to be invertible, the trajectories $x(t) = Q(y(t))$ do *not* consitute a semiflow, so the theory of Benaïm [3] does not apply. Instead, to overcome this difficulty, we exploit the derived convergence bound for $x(t) = Q(y(t))$, and we then use an inductive shadowing argument to show that (ML) converges itself to $\mathcal{X}^*$.

*Proof of Theorem 4.7.* Fix some $\varepsilon > 0$, let $U_\varepsilon = \{x = Q(y) : F(\mathcal{X}^*, y) < \varepsilon\}$, and write $\Phi_t \colon \mathcal{Y} \to \mathcal{Y}$ for the semiflow induced by (ML-C) on $\mathcal{Y}$ – i.e. $(\Phi_t(y))_{t\geq 0}$ is the solution orbit of (ML-C) that starts at $y \in \mathcal{Y}$.[8]

We first claim there exists some finite $\tau \equiv \tau(\varepsilon)$ such that $F(\mathcal{X}^*, \Phi_\tau(y)) \leq \max\{\varepsilon, F(\mathcal{X}^*, y) - \varepsilon\}$ for all $y \in \mathcal{Y}$. Indeed, since $\mathrm{cl}(U_\varepsilon)$ is a closed neighborhood of $\mathcal{X}^*$ by (H3), (VS) implies that there exists some $a \equiv a(\varepsilon) > 0$ such that

$$\langle v(x) \,|\, x - x^* \rangle \leq -a \quad \text{for all } x^* \in \mathcal{X}^*,\ x \notin U_\varepsilon. \tag{4.13}$$

Consequently, if $\tau_y = \inf\{t > 0 : Q(\Phi_t(y)) \in U_\varepsilon\}$ denotes the first time at which an orbit of (ML-C) reaches $U_\varepsilon$, Lemma A.2 in Appendix A gives:

$$F(x^*, \Phi_t(y)) \leq F(x^*, y) - at \quad \text{for all } x^* \in \mathcal{X}^*,\ t \leq \tau_y. \tag{4.14}$$

In view of this, set $\tau = \varepsilon/a$ and consider the following two cases:

(1) $\tau_y \geq \tau$: then, (4.14) gives $F(x^*, \Phi_\tau(y)) \leq F(x^*, y) - \varepsilon$ for all $x^* \in \mathcal{X}^*$, so $F(\mathcal{X}^*, \Phi_\tau(y)) \leq F(\mathcal{X}^*, y) - \varepsilon$.
(2) $\tau_y < \tau$: then, $Q(\Phi_\tau(y)) \in U_\varepsilon$, so $F(\mathcal{X}^*, \Phi_\tau(y)) \leq \varepsilon$.

In both cases we have $F(\mathcal{X}^*, \Phi_\tau(y)) \leq \max\{\varepsilon, F(\mathcal{X}^*, y) - \varepsilon\}$, as claimed.

Now, let $(Y(t))_{t\geq 0}$ denote the affine interpolation of the sequence $y_n$ generated by (ML), i.e. $Y$ is the continuous curve which connects linearly the values $y_{n+1}$ at all times $\tau_n = \sum_{k=0}^{n} \gamma_k$. Under the stated assumptions, a standard result of Benaïm [3, Propositions 4.1 and 4.2] shows that $Y(t)$ is an asymptotic pseudotrajectory of $\Phi$, i.e.

$$\lim_{t\to\infty} \sup_{0\leq h\leq T} \|Y(t+h) - \Phi_h(Y(t))\|_* = 0 \quad \text{for all } T > 0 \text{ (a.s.)}. \tag{4.15}$$

Thus, with some hindsight, let $\delta \equiv \delta(\varepsilon)$ be such that $\delta\|\mathcal{X}\| + \delta^2/(2K) \leq \varepsilon$ and choose $t_0 \equiv t_0(\varepsilon)$ so that $\sup_{0\leq h\leq \tau} \|Y(t+h) - \Phi_h(Y(t))\|_* \leq \delta$ for all $t \geq t_0$. Then, for all $t \geq t_0$ and all $x^* \in \mathcal{X}^*$, Proposition 4.3 gives

$$\begin{aligned} F(x^*, Y(t+h)) &\leq F(x^*, \Phi_h(Y(t))) \\ &\quad + \langle Y(t+h) - \Phi_h(Y(t)) \,|\, Q(\Phi_h(Y(t))) - x^* \rangle \\ &\quad + \frac{1}{2K} \|Y(t+h) - \Phi_h(Y(t))\|_*^2 \\ &\leq F(x^*, \Phi_h(Y(t))) + \delta\|\mathcal{X}\| + \frac{\delta^2}{2K} \\ &\leq F(x^*, \Phi_h(Y(t))) + \varepsilon. \end{aligned} \tag{4.16}$$

[8] That such a trajectory exists and is unique is a consequence of (H4).

Hence, minimizing over $x^* \in \mathcal{X}^*$, we get

$$F(\mathcal{X}^*, Y(t+h)) \leq F(\mathcal{X}^*, \Phi_h(Y(t))) + \varepsilon \quad \text{for all } t \geq t_0. \tag{4.17}$$

By Lemma A.3 in Appendix A, there exists some $t \geq t_0$ such that $F(\mathcal{X}^*, Y(t)) \leq 2\varepsilon$ (a.s.). Thus, given that $F(\mathcal{X}^*, \Phi_h(Y(t)))$ is nonincreasing in $h$ by Lemma A.2, Eq. (4.17) yields $F(\mathcal{X}^*, Y(t+h)) \leq 2\varepsilon+\varepsilon = 3\varepsilon$ for all $h \in [0,\tau]$. However, by the definition of $\tau$, we also have $F(\mathcal{X}^*, \Phi_\tau(Y(t))) \leq \max\{\varepsilon, F(\mathcal{X}^*, Y(t))-\varepsilon\} \leq \varepsilon$, implying in turn that $F(\mathcal{X}^*, Y(t+\tau)) \leq F(\mathcal{X}^*, \Phi_\tau(Y(t))) + \varepsilon \leq 2\varepsilon$. Therefore, by repeating the above argument at $t+\tau$ and proceeding inductively, we get $F(\mathcal{X}^*, Y(t+h)) \leq 3\varepsilon$ for all $h \in [k\tau, (k+1)\tau]$, $k = 0, 1, \dots$ (a.s.). Since $\varepsilon$ has been chosen arbitrarily, we conclude that $F(\mathcal{X}^*, y_n) \to 0$, so $x_n \to \mathcal{X}^*$ by Proposition 4.5. ∎

Theorem 4.7 shows that (almost) all realizations of (ML) converge to equilibrium, but the summability requirement $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$ suggests that players must be more conservative in their gradient steps under uncertainty. To make this more precise, note that the step-size assumptions of Theorem 4.6 are satisfied for all step-size policies of the form $\gamma_n \propto (n+1)^{-\beta}$, $\beta \in (0,1]$; however, in the presence of errors and uncertainty, Theorem 4.7 guarantees convergence only when $\beta \in (1/2, 1]$.

The "critical" value $\beta = 1/2$ above is tied to the finite mean squared error hypothesis (H2). If the players' gradient observations have finite moments up to some order $q > 2$, a more refined stochastic approximation argument as in Benaïm [3, Proposition 4.2] can be used to show that Theorem 4.7 still holds under the lighter requirement $\sum_{n=0}^{\infty} \gamma_n^{1+q/2} < \infty$. Consequently, even in the presence of noise, (ML) can be used with any step-size sequence of the form $\gamma_n \propto (n+1)^{-\beta}$, $\beta \in (0,1]$, provided that the noise process $\xi_n$ has $\mathbb{E}[\|\xi_n\|_*^q \,|\, \mathcal{F}_n] < \infty$ for some $q > 2/\beta - 2$. In particular, if the noise affecting the players' observations has finite moments of all orders (for instance, if $\xi_n$ is sub-exponential or sub-Gaussian), it is possible to recover essentially all the step-size policies covered by Theorem 4.6.

4.4. **Local convergence.** The results of the previous section show that (ML) converges globally to states (or sets) that are variationally stable on $\mathcal{X}$, even under noise and uncertainty. In this section, we show that (ML) remains locally convergent to states that are only locally stable with probability arbitrarily close to 1.

For simplicity, we begin with the deterministic, perfect feedback case:

**Theorem 4.10.** *Suppose that (ML) is run with perfect gradient information ($\sigma_* = 0$), choice maps satisfying (H3), and a sufficiently small step-size sequence with $\sum_{k=0}^{n} \gamma_k^2 / \sum_{k=0}^{n} \gamma_k \to 0$. If $\mathcal{X}^*$ is stable, there exists a neighborhood $U$ of $\mathcal{X}^*$ such that $x_n$ converges to $\mathcal{X}^*$ whenever $x_0 \in U$.*

*Proof.* As in the proof of Theorem 4.6, let $U_\varepsilon = \{x = Q(y) : F(\mathcal{X}^*, y) < \varepsilon\}$. Since $\mathcal{X}^*$ is stable, there exists some $\varepsilon > 0$ and some $a > 0$ satisfying (4.12) and such that (VS) holds throughout $U_\varepsilon$. If $x_0 \in U_\varepsilon$ and $\gamma_0 \leq \min\{2Ka/V_*^2, \sqrt{K\varepsilon}/V_*\}$, the same induction argument as in the proof of Theorem 4.6 shows that $x_n \in U_\varepsilon$ for all $n$. Since (VS) holds throughout $U_\varepsilon$, Lemma A.3 shows that $x_n$ visits any neighborhood of $\mathcal{X}^*$ infinitely many times; thus, by repeating the same argument as in the proof of Theorem 4.6, we get $x_n \to \mathcal{X}^*$. ∎

The key idea in the proof of Theorem 4.10 is that if the step-size of (ML) is small enough, the process $x_n = Q(y_n)$ always remains within the "basin of attraction" of

$\mathcal{X}^*$. On the other hand, if the players' feedback is subject to estimation errors and uncertainty, a single unlucky instance could drive $x_n$ away from said basin, possibly never to return. As a result, any local convergence result in the presence of noise must be probabilistic in nature. This is seen clearly in our next result:

**Theorem 4.11.** *Fix a confidence level $\delta > 0$ and suppose that (ML) is run with a sufficiently small step-size $\gamma_n$ satisfying $\sum_{n=0}^{\infty} \gamma_n^2 < \sum_{n=0}^{\infty} \gamma_n = \infty$. If $\mathcal{X}^*$ is stable and (H1)–(H4) hold, then $\mathcal{X}^*$ is locally attracting with probability at least $1 - \delta$; more precisely, there exists a neighborhood $U$ of $\mathcal{X}^*$ such that*

$$\mathbb{P}(x_n \to \mathcal{X}^* \,|\, x_0 \in U) \geq 1 - \delta. \tag{4.18}$$

**Corollary 4.12.** *Let $x^*$ be a Nash equilibrium with negative-definite Hessian matrix $H^{\mathcal{G}}(x^*) \prec 0$. Then, with assumptions as above, $x^*$ is locally attracting with probability arbitrarily close to 1.*

*Proof of Theorem 4.11.* Let $U_\varepsilon = \{x = Q(y) : F(\mathcal{X}^*, y) < \varepsilon\}$ and pick $\varepsilon > 0$ small enough so that (VS) holds for all $x \in U_{3\varepsilon}$. Assume further that $x_0 \in U_\varepsilon$ so there exists some $x^* \in \mathcal{X}^*$ such that $F(x^*, y_0) < \varepsilon$. Then, for all $n$, Proposition 4.3 yields

$$F(x^*, y_{n+1}) \leq F(x^*, y_n) + \gamma_n \langle v(x_n) \,|\, x_n - x^* \rangle + \gamma_n \psi_n + \frac{1}{2K} \gamma_n^2 \|\hat{v}_n^2\|_*, \tag{4.19}$$

where we have set $\psi_n = \langle \xi_n \,|\, x_n - x^* \rangle$.

We first claim that $\sup_n \sum_{k=0}^{n} \gamma_k \psi_k \leq \varepsilon$ with probability at least $1 - \delta/2$ if $\gamma_n$ is chosen appropriately. Indeed, let $S_n = \sum_{k=0}^{n} \gamma_k \psi_k$ and let $E_{n,\varepsilon}$ denote the event $\{\sup_{0 \leq k \leq n} |S_k| \geq \varepsilon\}$. Since $S_n$ is a martingale, Doob's maximal inequality (Hall and Heyde [15, Theorem 2.1]) yields

$$\mathbb{P}(E_{n,\varepsilon}) \leq \frac{\mathbb{E}[|S_n|^2]}{\varepsilon^2} \leq \frac{\sigma_*^2 \|\mathcal{X}\|^2 \sum_{k=0}^{n} \gamma_k^2}{\varepsilon^2}, \tag{4.20}$$

where we used the variance estimate

$$\mathbb{E}[\psi_k^2] = \mathbb{E}[\mathbb{E}[|\langle \xi_k \,|\, x_k - x^* \rangle|^2 \,|\, \mathcal{F}_k]] \leq \mathbb{E}[\mathbb{E}[\|\xi_k\|_*^2 \|x_k - x^*\|^2 \,|\, \mathcal{F}_k]] \leq \sigma_*^2 \|\mathcal{X}\|^2, \tag{4.21}$$

and the fact that $\mathbb{E}[\psi_k \psi_\ell] = \mathbb{E}[\mathbb{E}[\psi_k \psi_\ell] \,|\, \mathcal{F}_{k \vee \ell}] = 0$ whenever $k \neq \ell$. Since $E_{n+1,\varepsilon} \supseteq E_{n,\varepsilon} \supseteq \dots$, it follows that the event $E_\varepsilon = \bigcup_{n=0}^{\infty} E_{n,\varepsilon}$ occurs with probability $\mathbb{P}(E_\varepsilon) \leq \Gamma_2 \sigma_*^2 \|\mathcal{X}\|^2 / \varepsilon^2$, where $\Gamma_2 \equiv \sum_{n=0}^{\infty} \gamma_n^2$. Thus, if $\gamma_n$ is chosen so that $\Gamma_2 \leq \delta \varepsilon^2 / (2\sigma_*^2 \|\mathcal{X}\|^2)$, we get $\mathbb{P}(E_\varepsilon) \leq \delta/2$.

We now claim that the process $R_n = \sum_{k=0}^{n} \gamma_k^2 \|\hat{v}_k\|_*^2$ is also bounded from above by $\varepsilon$ with probability at least $1 - \delta/2$ if $\gamma_n$ is chosen appropriately. Indeed, working as above, let $F_{n,\varepsilon}$ denote the event $\{\sup_{0 \leq k \leq n} R_k \geq \varepsilon\}$. Since $R_n$ is a nonnegative submartingale, Doob's maximal inequality again yields

$$\mathbb{P}(F_{n,\varepsilon}) \leq \frac{\mathbb{E}[R_n]}{\varepsilon} \leq \frac{V_*^2 \sum_{k=0}^{n} \gamma_k^2}{\varepsilon}. \tag{4.22}$$

Consequently, the event $F_\varepsilon = \bigcup_{n=0}^{\infty} F_{n,\varepsilon}$ occurs with probability $\mathbb{P}(F_\varepsilon) \leq \Gamma_2 V_*^2 / \varepsilon \leq \delta/2$ if $\gamma_n$ is chosen so that $\Gamma_2 \leq \delta\varepsilon / (2V_*^2)$.

Assume therefore that $\Gamma_2 \leq \min\{\delta\varepsilon^2 / (2\sigma_*^2 \|\mathcal{X}\|^2), \delta\varepsilon / (2V_*^2)\}$. The above shows that $\mathbb{P}(\bar{E}_\varepsilon \cap \bar{F}_\varepsilon) = 1 - \mathbb{P}(E_\varepsilon \cup F_\varepsilon) \geq 1 - \delta/2 - \delta/2 = 1 - \delta$, i.e. $S_n$ and $R_n$ are both bounded from above by $\varepsilon$ for all $n$ and all $x^*$ with probability at least $1 - \delta$. Since

$F(x^*, y_0) \leq \varepsilon$ by assumption, we readily get $F(x^*, y_1) \leq 3\varepsilon$ if $\bar{E}_\varepsilon$ and $\bar{F}_\varepsilon$ both hold. Furthermore, telescoping (4.19) yields

$$F(x^*, y_{n+1}) \leq F(x^*, y_0) + \sum_{k=0}^{n} \langle v(x_n) \,|\, x_n - x^* \rangle + S_n + R_n \quad \text{for all } n, \tag{4.23}$$

so if we assume inductively that $F(x^*, y_k) \leq 3\varepsilon$ for all $k \leq n$ (implying that $\langle v(x_k) \,|\, x_k - x^* \rangle \leq 0$ for all $k \leq n$), we also get $F(x^*, y_{n+1})$ if neither $E_\varepsilon$ nor $F_\varepsilon$ occur. Since $\mathbb{P}(E_\varepsilon \cup F_\varepsilon) \leq \delta$, we conclude that $x_n$ stays in $U_{3\varepsilon}$ for all $n$ with probability at least $1 - \delta$. In turn, when this is the case, Lemma A.3 shows that $\mathcal{X}^*$ is recurrent under $x_n$; hence, by repeating the same steps as in the proof of Theorem 4.7, we get $x_n \to \mathcal{X}^*$ with probability at least $1 - \delta$, as asserted. ∎

4.5. **Convergence in zero-sum games.** We close this section by examining the asymptotic behavior of (ML) in 2-player zero-sum games ($\mathcal{N} = \{1, 2\}$, $u_1 + u_2 = 0$). Letting $u \equiv u_1 = -u_2$, the *value* of such a game is defined as

$$u^* = \max_{x_1 \in \mathcal{X}_1} \min_{x_2 \in \mathcal{X}} u(x_1, x_2) = \min_{x_2 \in \mathcal{X}_2} \max_{x_1 \in \mathcal{X}_1} u(x_1, x_2), \tag{4.24}$$

and its existence is guaranteed by the original minmax theorem of von Neumann [51]. On that account, the solutions of the saddle-point problem (4.24) are the Nash equilibria of $\mathcal{G}$ and the players' equilibrium payoffs are $u^*$ and $-u^*$ respectively.

With this in mind, we show below that the long-term average of the sequence of play generated by (ML) converges to equilibrium with probability 1:

**Theorem 4.13.** *Let $\mathcal{G}$ be a 2-player zero-sum game. If (ML) is run with imperfect gradient information satisfying (H1)–(H2) and a step-size $\gamma_n$ such that $\sum_{n=0}^{\infty} \gamma_n^2 < \sum_{n=0}^{\infty} \gamma_n = \infty$, the long-term average $\bar{x}_n = \sum_{k=0}^{n} \gamma_k x_k / \sum_{k=0}^{n} \gamma_k$ of the induced sequence of play converges to the set of Nash equilibria of $\mathcal{G}$ (a.s.).*

Nemirovski et al. [32] and Juditsky et al. [21] showed that, under a "greedy" variant of (ML) run with similar noise and step-size assumptions, $\bar{x}_n$ converges to the Nash set of $\mathcal{G}$ in $L^1$ (see also Nesterov [35] for a deterministic, perfect feedback version of this result). Our proof is inspired by that of Nemirovski et al. [32], but we focus throughout on almost sure convergence instead of convergence in $L^1$.

*Proof of Theorem 4.13.* Consider the gap function

$$\epsilon(x) = u^* - \min_{p_2 \in \mathcal{X}_2} u(x_1, p_2) + \max_{p_1 \in \mathcal{X}_1} u(p_1, x_2) - u^* = \max_{p \in \mathcal{X}} \sum_{i \in \mathcal{N}} u_i(p_i; x_{-i}). \tag{4.25}$$

Obviously, $\epsilon(x) \geq 0$ with equality if and only if $x$ is a Nash equilibrium, so it suffices to show that $\epsilon(\bar{x}_n) \to 0$ (a.s.).

To that end, pick an arbitrary state $p \in \mathcal{X}$. Then, arguing as in the proof of Theorem 4.7, we get

$$F(p, y_{n+1}) \leq F(p, y_n) + \gamma_n \langle v(x_n) \,|\, x_n - p \rangle + \gamma_n \psi_n + \frac{1}{2K} \gamma_n^2 \|\hat{v}_n\|_*^2, \tag{4.26}$$

and hence, after rearranging and telescoping

$$\sum_{k=0}^{n} \gamma_k \langle v(x_k) \,|\, p - x_k \rangle \leq F(p, y_0) + \sum_{k=0}^{n} \gamma_k \psi_k + \frac{1}{2K} \sum_{k=0}^{n} \gamma_k^2 \|\hat{v}_k\|_*^2, \tag{4.27}$$

where $\psi_n = \langle \xi_n \,|\, x_n - p \rangle$ and we used the fact that $F(p, y_{n+1}) \geq 0$. Since $u_i$ is concave in $x_i$, we also have

$$\langle v(x) \,|\, p - x \rangle = \sum_{i \in \mathcal{N}} \langle v_i(x) \,|\, p_i - x_i \rangle \geq \sum_{i \in \mathcal{N}} [u_i(p_i; x_{-i}) - u_i(x)] = \sum_{i \in \mathcal{N}} u_i(p_i; x_{-i}), \tag{4.28}$$

for all $x \in \mathcal{X}$. Therefore, letting $\tau_n = \sum_{k=0}^{n} \gamma_k$, we get

$$\begin{aligned} \frac{1}{\tau_n} \sum_{k=0}^{n} \gamma_k \langle v(x_k) \,|\, p - x_k \rangle &\geq \frac{1}{\tau_n} \sum_{k=0}^{n} \gamma_k \sum_{i \in \mathcal{N}} u_i(p_i; x_{-i,k}) \\ &\geq u(p_1, \bar{x}_{2,n}) - u(\bar{x}_{1,n}, p_2) = \sum_{i \in \mathcal{N}} u_i(p_i; \bar{x}_{-i,n}), \end{aligned} \tag{4.29}$$

where we used the fact that $u$ is concave-convex in the second line. Thus, combining (4.27) and (4.29), we finally obtain

$$\sum_{i \in \mathcal{N}} u_i(p_i; \bar{x}_{-i,n}) \leq \frac{F(p, y_0) + \sum_{k=0}^{n} \gamma_k \psi_k + (2K)^{-1} \sum_{k=0}^{n} \gamma_k^2 \|\hat{v}_k\|_*^2}{\tau_n}. \tag{4.30}$$

As before, the law of large numbers (Hall and Heyde [15, Theorem 2.18]) yields $\tau_n^{-1} \sum_{k=0}^{n} \gamma_k \psi_k \to 0$ (a.s.). Furthermore, given that $\mathbb{E}[\|\hat{v}_n\|_*^2] \leq V_*^2$ and $\sum_{k=0}^{n} \gamma_k^2 < \infty$, we also get $\tau_n^{-1} \sum_{k=0}^{n} \gamma_k^2 \|\hat{v}_k\|_*^2 \to 0$ by Doob's martingale convergence theorem (Hall and Heyde [15, Theorem 2.5]), implying in turn that $\sum_{i \in \mathcal{N}} u_i(p_i; \bar{x}_{-i,n}) \to 0$ (a.s.). Since $p$ is arbitrary, we conclude that $\epsilon(\bar{x}_n) \to 0$ (a.s.), as claimed. ∎

## 5. Learning in finite games

As a concrete application of the analysis of the previous section, we turn to the asymptotic behavior of (ML) in *finite* games. Briefly recalling the setup of Example 2.1, each player in a finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$ chooses a pure strategy $s_i$ from a finite set $\mathcal{S}_i$ and receives a payoff of $u_i(s_1, \dots, s_N)$. Pure strategies are drawn based on the players' mixed strategies $x_i \in \mathcal{X}_i \equiv \Delta(\mathcal{S}_i)$, so each player's expected payoff is given by the multilinear expression (2.3). Accordingly, the individual payoff gradient of player $i \in \mathcal{N}$ in the mixed profile $x = (x_1, \dots, x_N)$ is the (mixed) payoff vector $v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}) = (u_i(s_i; x_{-i}))_{s_i \in \mathcal{S}_i}$ – cf. Eq. (2.4).

Consider now the following learning scheme: At each stage $n = 0, 1, \dots$, every player $i \in \mathcal{N}$ selects a pure strategy $s_{i,n} \in \mathcal{S}_i$ according to their individual mixed strategy $x_{i,n} \in \mathcal{X}_i$ at stage $n$. Subsequently, each player observes – or calculates in some other way – the payoffs of his pure strategies $s_i \in \mathcal{S}_i$ against the chosen actions $s_{-i,n}$ of all other players, possibly subject to some random estimation error. Specifically, we posit that each player observes the "noisy" payoff vector

$$\hat{v}_{i,n} = (u_i(s_i; s_{-i,n}))_{s_i \in \mathcal{S}_i} + \xi_{i,n}, \tag{5.1}$$

where the error process $\xi_n = (\xi_{i,n})_{i \in \mathcal{N}}$ is assumed to satisfy Hypotheses (H1) and (H2). Then, based on this feedback, players update their mixed strategies using (ML) and the process repeats (for a concrete example, see Algorithm 2).

In the rest of this section, we study the long-term behavior of this game-theoretic learning process. Specifically, we focus on: *a*) the elimination of dominated strategies; *b*) convergence to strict Nash equilibria; and *c*) convergence to equilibrium in 2-player, zero-sum games.

**Algorithm 2.** Logit-based learning (or "hedging") in finite games (Example 3.2).

**Parameter:** step-size $\gamma_n \propto 1/n^\beta$, $0 < \beta \leq 1$.
**Initialization:** $n \leftarrow 0$; $y_i \leftarrow$ chosen by player $i$.
**Repeat**
  $n \leftarrow n + 1$;
  **foreach** player $i \in \mathcal{N}$ **do**
    set $x_i \leftarrow \Lambda_i(y_i)$; # update mixed strategies
    draw $s_i \in \mathcal{S}_i$ based on $x_i$; # choose actions
    observe $\hat{v}_i$; # estimate payoffs
    update $y_i \leftarrow y_i + \gamma_n \hat{v}_i$; # update scores
  **until** convergence

5.1. **Dominated strategies.** A pure strategy $s_i \in \mathcal{S}_i$ of a finite game $\Gamma$ is said to be *dominated* by $s_i' \in \mathcal{S}_i$ (and written $s_i \prec s_i'$) if

$$u_i(s_i; x_{-i}) < u_i(s_i'; x_{-i}) \quad \text{for all } x_{-i} \in \mathcal{X}_{-i} \equiv \textstyle\prod_{j\neq i} \mathcal{X}_j. \tag{5.2}$$

Put differently, we have $s_i \prec s_i'$ if and only if $v_{is_i}(x) < v_{is_i'}(x)$ for all $x \in \mathcal{X}$. In turn, this implies that the payoff gradient of player $i$ points consistently towards the face $x_{is_i} = 0$ of $\mathcal{X}_i$, so it is natural to expect that $s_i$ is eliminated under (ML). Indeed, we have:

**Theorem 5.1.** *Suppose that* (ML) *is run with noisy payoff observations of the form* (5.1) *and a step-size sequence* $\gamma_n$ *satisfying the summability condition* (4.2). *If* $s_i \in \mathcal{S}_i$ *is dominated, then* $x_{is_i,n} \to 0$ *(a.s.).*

*Proof.* Suppose that $s_i \prec s_i'$ for some $s_i' \in \mathcal{S}_i$. Then, suppressing the player index $i$ for simplicity, (ML) gives

$$\begin{aligned} y_{s',n} - y_{s,n} &= c_{s's} + \sum_{k=0}^{n} \gamma_k \left[\hat{v}_{s',k} - \hat{v}_{s,k}\right] \\ &= c_{s's} + \sum_{k=0}^{n} \gamma_k \left[v_{s'}(x_k) - v_s(x_k)\right] + \sum_{k=0}^{n} \gamma_k \zeta_k, \end{aligned} \tag{5.3}$$

where we set $c_{s's} = y_{s',0} - y_{s,0}$ and $\zeta_k = \mathbb{E}[\hat{v}_{s',k} - \hat{v}_{s,k} \,|\, \mathcal{F}_k] - [v_{s'}(x_k) - v_s(x_k)]$. Since $s \prec s'$, there exists some $a > 0$ such that $v_{s'}(x) - v_s(x) \geq a$ for all $x \in \mathcal{X}$. Then, (5.3) yields

$$y_{s',n} - y_{s,n} \geq c_{s's} + \tau_n \left[a + \frac{\sum_{k=0}^{n} \gamma_k \zeta_k}{\tau_n}\right], \tag{5.4}$$

where $\tau_n = \sum_{k=0}^{n} \gamma_k$. As in the proof of Theorem 4.1, the law of large numbers for martingale differences (Hall and Heyde [15, Theorem 2.18]) implies that $\tau_n^{-1} \sum_{k=0}^{n} \gamma_k \zeta_k \to 0$ under the step-size assumption (4.2), so $y_{s',n} - y_{s,n} \to \infty$ (a.s.).

Suppose now that $\limsup_{n\to\infty} x_{s,n} = 2\varepsilon$ for some $\varepsilon > 0$. By descending to a subsequence if necessary, we may assume that $x_{s,n} \geq \varepsilon$ for all $n$, so if we let $x_n' = x_n + \varepsilon(e_{s'} - e_s)$, the definition of $Q$ gives

$$\langle y_n \,|\, x_n \rangle - h(x_n) \geq \langle y_n \,|\, x_n' \rangle - h(x_n') = \langle y_n \,|\, x_n \rangle + \varepsilon(y_{s,n} - y_{s',n}) - h(x_n'). \tag{5.5}$$

Therefore, after rearranging, we get $h(x'_n) - h(x_n) \leq \varepsilon(y_{s,n} - y_{s',n}) \to -\infty$, a contradiction. This implies that $x_{s,n} \to 0$ (a.s.), as asserted. ∎

5.2. **Strict equilibria.** A Nash equilibrium $x^*$ of a finite game is called *strict* when (NE) holds as a strict inequality for all $x_i \neq x_i^*$, i.e. when no player can deviate unilaterally from $x^*$ without *reducing* their payoff (or, equivalently, when every player has a unique best response to $x^*$). This implies that strict Nash equilibria are pure strategy profiles $x^* = (s_1^*, \dots, s_N^*)$ such that

$$u_i(s_i^*; s_{-i}^*) > u_i(s_i; s_{-i}^*) \quad \text{for all } s_i \in \mathcal{S}_i \setminus \{s_i^*\},\ i \in \mathcal{N}. \tag{5.6}$$

In particular, we have the following characterization of strict Nash equilibria:

**Proposition 5.2.** *Let $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ be the mixed extension of a finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$. Then, the following are equivalent:*

*a) $x^*$ is a strict Nash equilibrium.*
*b) $\langle v(x^*) \,|\, z\rangle \leq 0$ for all $z \in \mathrm{TC}(x^*)$ with equality if and only if $z = 0$.*
*c) $x^*$ is stable.*

Thanks to the above characterization of strict equilibria (proven in Appendix A), the convergence analysis of Section 4 yields:

**Proposition 5.3.** *Let $x^*$ be a strict equilibrium of a finite game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$. Suppose further that (ML) is run with noisy payoff observations of the form (5.1) and a sufficiently small step-size $\gamma_n$ satisfying $\sum_{n=0}^{\infty} \gamma_n^2 < \sum_{n=0}^{\infty} \gamma_n = \infty$. If (H1)–(H3) hold, $x^*$ is locally attracting with arbitrarily high probability; specifically, for all $\delta > 0$, there exists a neighborhood $U$ of $x^*$ such that*

$$\mathbb{P}(x_n \to x^* \,|\, x_0 \in U) \geq 1 - \delta. \tag{5.7}$$

*Proof.* We first show that the noisy payoff vector $\hat{v}_n$ of (5.1) satisfies $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = v(x_n)$. Indeed, for all $i \in \mathcal{N}$, $s_i \in \mathcal{S}_i$, we have

$$\mathbb{E}[\hat{v}_{is_i,n} \,|\, \mathcal{F}_n] = \sum_{s_{-i} \in \mathcal{S}_{-i}} u_i(s_i; s_{-i})\, x_{s_{-i},n} + \mathbb{E}[\xi_{is_i,n} \,|\, \mathcal{F}_n] = u_i(s_i; x_{-i,n}), \tag{5.8}$$

where, in a slight abuse of notation, we write $x_{s_{-i},n}$ for the joint probability assigned to the pure strategy profile $s_{-i}$ of all players other than $i$ at stage $n$. Thus, comparing with (2.4), we get $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = v(x_n)$.

The above shows that (5.1) is an unbiased estimator of $v(x_n)$ that satisfies (H1). Hypothesis (H2) can be verified similarly, so (5.1) satisfies (3.3). Since $x^*$ is stable by Proposition 5.2 and $v(x)$ is multilinear (meaning in particular that (H4) is satisfied automatically), our assertion follows from Theorem 4.11. ∎

*Remark* 5.1. In the special case of logit-based learning (cf. Example 3.2), Cohen et al. [8] recently showed that Algorithm 2 converges locally to strict Nash equilibria under similar information assumptions. Proposition 5.2 essentially extends this result to the entire class of regularized learning processes induced by (ML) in finite games, thus showing that the logit choice map (3.9) has no distinctive attributes in this respect. Cohen et al. [8] further showed that the convergence rate of logit-based learning is exponential in the algorithm's "running horizon" $\tau_n = \sum_{k=0}^{n} \gamma_k$. This rate is closely linked to the logit choice model, and different choice maps yield different convergence speeds; we discuss this issue in more detail in Section 6.

5.3. **Convergence in zero-sum games.** Proposition 5.2 shows that the learning scheme (ML) is locally convergent in generic finite games that admit a Nash equilibrium in pure strategies, even in the presence of observation noise and estimation errors. To complement this result, we now turn to zero-sum games where Nash equilibria are often interior – in a sense, the "opposite" of strict equilibria.

In this setting, the analysis of Section 4.5 readily yields:

**Corollary 5.4.** *Let $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$ be a finite 2-player zero-sum game. If (ML) is run with noisy payoff observations of the form (5.1) and a step-size $\gamma_n$ such that $\sum_{n=0}^{\infty} \gamma_n^2 < \sum_{n=0}^{\infty} \gamma_n = \infty$, the long-term average $\bar{x}_n = \sum_{k=0}^{n} \gamma_k x_k \big/ \sum_{k=0}^{n} \gamma_k$ of the players' mixed strategies converges to the set of Nash equilibria of $\Gamma$ (a.s.).*

*Proof.* As in the proof of Proposition 5.2, the estimator (5.1) satisfies $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = v(x_n)$, so (H1) and (H2) also hold in the sense of (3.3). Our claim then follows from Theorem 4.13. ∎

*Remark* 5.2. Extending a result of Hofbauer et al. [20] for the replicator dynamics, Mertikopoulos and Sandholm [30] recently showed that the long-term average $\bar{x}(t) = t^{-1} \int_0^t x(s)\, ds$ of the players' mixed strategies under (ML-C) converges to Nash equilibrium in 2-player, zero-sum games.[9] Under this light, Corollary 5.4 can be seen as an extension of the results of Hofbauer et al. [20] and Mertikopoulos and Sandholm [30] to a stochastic, discrete-time setting.

## 6. Speed of convergence

6.1. **Mean convergence rate.** In this section, we focus on the quantitative aspects of the long-run behavior of (ML) and, in particular, its rate of convergence to stable equilibrium states (and/or sets thereof). To that end, we will quantify the speed of convergence to a globally stable set $\mathcal{X}^* \subseteq \mathcal{X}$ via the *equilibrium gap function*

$$\epsilon(x) = \inf_{x^* \in \mathcal{X}^*} \langle v(x) \,|\, x^* - x \rangle. \tag{6.1}$$

By Definition 2.3, $\epsilon(x) \geq 0$ with equality if and only if $x \in \mathcal{X}^*$, so $\epsilon(x)$ can be seen as a (game-specific) measure of the distance between $x$ and the target set $\mathcal{X}^*$. This can be seen more clearly in the case of *strongly* stable equilibria, defined here as follows:

**Definition 6.1.** We say that $x^* \in \mathcal{X}$ is *strongly stable* if there exists some $L > 0$ such that

$$\langle v(x) \,|\, x - x^* \rangle \leq -L \|x - x^*\|^2 \qquad \text{for all } x \in \mathcal{X}. \tag{6.2}$$

More generally, a closed subset $\mathcal{X}^*$ of $\mathcal{X}$ is called strongly stable if

$$\langle v(x) \,|\, x - x^* \rangle \leq -L \operatorname{dist}(\mathcal{X}^*, x)^2 \quad \text{for all } x \in \mathcal{X},\ x^* \in \mathcal{X}^*. \tag{6.3}$$

Obviously, $\epsilon(x) \geq L \operatorname{dist}(\mathcal{X}^*, x)^2$ if $\mathcal{X}^*$ is $L$-strongly stable, i.e. $\epsilon(x)$ grows at least quadratically near strongly stable sets – just like strongly convex functions near their minimum points. With this in mind, we derive below an estimate of the mean decay rate of the average equilibrium gap $\bar{\epsilon}_n = \sum_{k=0}^{n} \gamma_k \epsilon(x_k) \big/ \sum_{k=0}^{n} \gamma_k$ in the spirit of Nemirovski et al. [32] and Juditsky et al. [21]:

[9] See also Bravo and Mertikopoulos [7] for a further extension of this result to the case where the continuous-time dynamics (ML-C) are subject to Brownian payoff disturbances.

**Theorem 6.2.** *Suppose that* (ML) *is run with imperfect gradient information satisfying* (H1)–(H2). *Then*

$$\mathbb{E}[\bar{\epsilon}_n] \leq \frac{F_0 + V_*^2/(2K)\sum_{k=0}^n \gamma_k^2}{\sum_{k=0}^n \gamma_k}, \tag{6.4}$$

*where* $F_0 = F(\mathcal{X}^*, y_0)$. *If, in addition,* $\sum_{n=0}^\infty \gamma_n^2 < \infty$, *we have*

$$\bar{\epsilon}_n \leq \frac{A}{\sum_{k=0}^n \gamma_k} \quad \text{for all } n \text{ (a.s.)}, \tag{6.5}$$

*where* $A > 0$ *is a finite random variable such that, with probability at least* $1 - \delta$,

$$A \leq F_0 + \sigma_* \|\mathcal{X}\| \alpha + V_*^2 \alpha^2, \tag{6.6}$$

*where* $\alpha^2 = 2\delta^{-1} \sum_{n=0}^\infty \gamma_n^2$.

**Corollary 6.3.** *Suppose that* (ML) *is initialized at* $y_0 = 0$ *and run for* $n$ *iterations with constant step-size* $\gamma = V_*^{-1}\sqrt{2K\Omega/n}$ *where* $\Omega = \max h - \min h$. *Then,*

$$\mathbb{E}[\bar{\epsilon}_n] \leq 2V_* \sqrt{\Omega/(Kn)}. \tag{6.7}$$

*In addition, if* $\mathcal{X}^*$ *is* $L$*-strongly stable, we also have* $\mathbb{E}[\bar{r}_n] \leq \sqrt[4]{4L^{-2}V_*^2\Omega/(Kn)}$.

*Proof of Theorem 6.2.* Let $x^* \in \mathcal{X}^*$. Rearranging (4.19) and telescoping yields

$$\sum_{k=0}^n \gamma_k \langle v(x_k) \,|\, x^* - x_k\rangle \leq F(x^*, y_0) + \sum_{k=0}^n \gamma_k \psi_k + \frac{1}{2K}\sum_{k=0}^n \gamma_k^2 \|\hat{v}_k\|_*^2, \tag{6.8}$$

where $\psi_k = \langle \xi_k \,|\, x_k - x^*\rangle$. Thus, taking expectations, we obtain

$$\sum_{k=0}^n \gamma_k \,\mathbb{E}[\langle v(x_k) \,|\, x^* - x_k\rangle] \leq F(x^*, y_0) + \frac{V_*^2}{2K}\sum_{k=0}^n \gamma_k^2, \tag{6.9}$$

and hence, minimizing both sides of (6.9) over $x^* \in \mathcal{X}^*$, we have

$$\sum_{k=0}^n \gamma_k \,\mathbb{E}[\epsilon(x_k)] \leq F_0 + \frac{V_*^2}{2K}\sum_{k=0}^n \gamma_k^2, \tag{6.10}$$

where we used Jensen's inequality to interchange the inf and $\mathbb{E}$ operations. The estimate (6.4) then follows immediately.

For the almost sure bound (6.5), set $S_n = \sum_{k=0}^n \gamma_k \psi_k$ and $R_n = \sum_{k=0}^n \gamma_k^2 \|v_k\|_*^2$. Then, (6.8) becomes

$$\sum_{k=0}^n \gamma_k \langle v(x_k) \,|\, x^* - x_k\rangle \leq F(x^*, y_0) + S_n + R_n, \tag{6.11}$$

Arguing as in the proof of Theorem 4.11, it follows that $\sup_n \mathbb{E}[|S_n|]$ and $\sup_n \mathbb{E}[R_n]$ are both finite, i.e. $S_n$ and $R_n$ are both bounded in $L^1$. Thus, Doob's (sub)martingale convergence theorem (Hall and Heyde [15, Theorem 2.5]) shows that $S_n$ and $R_n$ both converge (a.s.) to a random, finite limit $S_\infty$ and $R_\infty$ respectively. Consequently, by (6.11), there exists an a.s. finite random variable $A > 0$ such that

$$\sum_{k=0}^n \gamma_k \langle v(x_k) \,|\, x^* - x_k\rangle \leq A \quad \text{for all } n \text{ (a.s.)}. \tag{6.12}$$

The bound (6.5) then follows by taking the minimum of (6.12) over $x^* \in \mathcal{X}^*$ and dividing both sides by $\sum_{k=0}^n \gamma_k$.

Finally, applying Doob's maximal inequality to (4.20) and (4.22), we obtain $\mathbb{P}\big(\sup_n S_n \geq \sigma_* \|\mathcal{X}\| \, \alpha\big) \leq \delta/2$ and $\mathbb{P}\big(\sup_n R_n \geq V_*^2 \alpha^2\big) \leq \delta/2$. Combining these bounds with (6.11) then shows that $A$ can be taken to satisfy (6.6) with probability at least $1 - \delta$, as claimed. ∎

*Proof of Corollary 6.3.* By the definition (4.10) of the setwise Fenchel coupling, we have $F_0 \leq h(x^*) + h^*(0) \leq \max h - \min h = \Omega$. Our claim then follows by noting that $\mathbb{E}[\operatorname{dist}(\mathcal{X}^*, x_n)]^2 \leq \mathbb{E}[\operatorname{dist}(\mathcal{X}^*, x_n)^2] \leq L^{-1} \, \mathbb{E}[\epsilon(x_n)]$ and applying (6.4). ∎

Although the mean bound (6.4) is valid for any step-size sequence, the summability condition $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$ for the almost sure bound (6.5) rules out more aggressive step-size policies of the form $\gamma_n \propto (n+1)^{-\beta}$ for $\beta \leq 1/2$. Specifically, the "critical" value $\beta = 1/2$ is again tied to the finite mean squared error hypothesis (H2): if the players' gradient measurements have finite moments up to some order $q > 2$, a more refined application of Doob's inequality reveals that (6.5) still holds under the lighter summability requirement $\sum_{n=0}^{\infty} \gamma_n^{1+q/2} < \infty$. In this case, the exponent $\beta = 1/2$ is optimal with respect to the guarantee (6.4) and leads to an almost sure convergence rate of the order of $\mathcal{O}(n^{-1/2} \log n)$.

Except for this $\log n$ factor, the $\mathcal{O}(n^{-1/2})$ convergence rate of (ML) is the exact lower complexity bound for black-box subgradient schemes for convex problems (Nemirovski and Yudin [33]; Nesterov [34]). Thus, running (ML) with a step-size policy of the form $\gamma_n \propto n^{-1/2}$ leads to a convergence speed that is optimal in the mean, and near-optimal with high probability.

6.2. **Running length.** Intuitively, the main obstacle to achieving rapid convergence is that, even with an optimized step-size policy, players may end up oscillating around an equilibrium state because of the noise in their observations. To quantify this behavior, we focus below on the *running length* of (ML), defined as

$$\ell_n = \sum_{k=0}^{n-1} \|x_{k+1} - x_k\|. \tag{6.13}$$

Obviously, if $x_n$ converges to some limit point $x^*$, a shorter length signifies less oscillations of $x_n$ around $x^*$; thus, in a certain way, $\ell_n$ is a more refined convergence criterion than the induced equilibrium gap $\epsilon(x_n)$.

Our next result shows that the expected running length of (ML) until players reach an $\varepsilon$-neighborhood of a (strongly) stable set is at most $\mathcal{O}(1/\varepsilon^2)$:

**Theorem 6.4.** *Suppose that (ML) is run with imperfect gradient information satisfying (H1)–(H2) and a step-size sequence $\gamma_n$ such that $\sum_{n=0}^{\infty} \gamma_n^2 < \sum_{n=0}^{\infty} \gamma_n = \infty$. If $\mathcal{X}^*$ is $L$-strongly stable and $\ell_\varepsilon$ is the length of $x_n$ until $x_n$ gets within $\varepsilon$ of $\mathcal{X}^*$, we have*

$$\mathbb{E}[\ell_\varepsilon] \leq \frac{V_*}{KL} \frac{F_0 + (2K)^{-1} V_*^2 \sum_{k=0}^{\infty} \gamma_k^2}{\varepsilon^2}. \tag{6.14}$$

*Proof.* Define the stopping time $n_\varepsilon = \inf\{n \geq 0 : \operatorname{dist}(\mathcal{X}^*, x_n) \leq \varepsilon\}$ so $\ell_\varepsilon = \ell_{n_\varepsilon}$. Then, for all $x^* \in \mathcal{X}^*$ and all $n \in \mathbb{N}$, (4.19) yields

$$F(x^*, y_{n_\varepsilon \wedge n+1}) \leq F(x^*, y_0) - \sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k \langle v(x_k) \,|\, x_k - x^* \rangle + \sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k \psi_k + \frac{1}{2K} \sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k \|\hat{v}_k\|_*^2. \tag{6.15}$$

Hence, after taking expectations and minimizing over $x^* \in \mathcal{X}^*$, we get

$$0 \leq F_0 - L\varepsilon^2\,\mathbb{E}\left[\sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k\right] + \mathbb{E}\left[\sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k \psi_k\right] + \frac{V_*^2}{2K}\sum_{k=0}^{\infty}\gamma_k^2, \tag{6.16}$$

where we we used the fact that $\|x_k - x^*\| \geq \varepsilon$ for all $k \leq n_\varepsilon$.

Consider now the stopped process $S_{n_\varepsilon \wedge n} = \sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k \psi_k$. Since $n_\varepsilon \wedge n \leq n < \infty$, $S_{n_\varepsilon \wedge n}$ is a martingale and $\mathbb{E}[S_{n_\varepsilon \wedge n}] = 0$. Thus, by rearranging (6.16), we obtain

$$\mathbb{E}\left[\sum_{k=0}^{n_\varepsilon \wedge n} \gamma_k\right] \leq \frac{F_0 + (2K)^{-1}V_*^2 \sum_{k=0}^{\infty}\gamma_k^2}{L\varepsilon^2}. \tag{6.17}$$

Hence, with $n_\varepsilon \wedge n \to n_\varepsilon$ as $n \to \infty$, Lebesgue's monotone convergence theorem shows that the process $\tau_\varepsilon = \sum_{k=0}^{n_\varepsilon} \gamma_k$ is finite in expectation and

$$\mathbb{E}[\tau_\varepsilon] \leq \frac{F_0 + (2K)^{-1}V_*^2 \sum_{k=0}^{\infty}\gamma_k^2}{L\varepsilon^2}. \tag{6.18}$$

Furthermore, by Proposition 3.2 and the definition of $\ell_n$, we also have

$$\ell_n = \sum_{k=0}^{n-1}\|x_{k+1} - x_k\| \leq \frac{1}{K}\sum_{k=0}^{n-1}\|y_{k+1} - y_k\|_* = \frac{1}{K}\sum_{k=0}^{n-1}\gamma_k\|\hat{v}_k\|_*. \tag{6.19}$$

Now, let $\zeta_k = \|\hat{v}_k\|_*$ and $\Psi_n = \sum_{k=0}^{n}\gamma_k\,[\zeta_k - \mathbb{E}[\zeta_k \,|\, \mathcal{F}_k]]$. By construction, $\Psi_n$ is a martingale and $\mathbb{E}[\Psi_n^2] = \mathbb{E}[\sum_{k=0}^{n}\gamma_k^2\,[\zeta_k - \mathbb{E}[\zeta_k \,|\, \mathcal{F}_k]]^2] \leq 2V_*^2\sum_{k=0}^{\infty}\gamma_k^2 < \infty$ for all $n$. Thus, by the optional stopping theorem (Shiryaev [46, p. 485]), we get $\mathbb{E}[\Psi_{n_\varepsilon}] = \mathbb{E}[\Psi_0] = 0$, so

$$\mathbb{E}\left[\sum_{k=0}^{n_\varepsilon}\gamma_k\zeta_k\right] = \mathbb{E}\left[\sum_{k=0}^{n_\varepsilon}\gamma_k\,\mathbb{E}[\zeta_k \,|\, \mathcal{F}_k]\right] \leq V_*\,\mathbb{E}\left[\sum_{k=0}^{n_\varepsilon}\gamma_k\right] = V_*\,\mathbb{E}[\tau_\varepsilon]. \tag{6.20}$$

Our claim then follows by combining (6.19) and (6.20) with the bound (6.18). ∎

6.3. **Fast convergence to strict equilibria.** Because of the random shocks induced by the noise in the players' gradient observations, it is difficult to obtain an almost sure (or high probability) estimate for the convergence rate of the last iterate $x_n$ of (ML). Specifically, even with a rapidly decreasing step-size policy, a single realization of the error process $\xi_n$ may lead to an arbitrarily big jump of $x_n$ at any time, thus destroying any almost sure bound on the convergence rate of $x_n$.

On the other hand, in finite games, Cohen et al. [8] recently showed that logit-based learning (cf. Algorithm 2) achieves a quasi-linear convergence rate with high probability if the equilibrium in question is strict. Specifically, if $x^*$ is a strict Nash equilibrium and $x_0$ is not initialized too far from $x^*$, Cohen et al. [8] showed that, with high probability, $\|x_n - x^*\| = \mathcal{O}(-a\sum_{k=0}^{n}\gamma_k)$ for some positive constant $a > 0$ that depends only on the players' relative payoff differences.

To extend this result to our setting, we employ the characterization of Proposition 5.2 and define strict equilibria in concave games as follows:

**Definition 6.5.** We say that $x^* \in \mathcal{X}$ is a *strict Nash equilibrium* of $\mathcal{G}$ if

$$\langle v(x^*) \,|\, z\rangle \leq 0 \quad \text{for all } z \in \mathrm{TC}(x^*), \tag{6.21}$$

with equality if and only if $z = 0$.

In contrast to finite games, the adjective "strict" above does not directly apply to the definition (NE) of Nash equilibria, but to their variational characterization (2.8). The reason for this is that, if a player's payoff function is (individually) strictly concave, every Nash equilibrium would satisfy (NE) as a strict inequality, so the characterization "strict" would offer no new information if it were based on (NE). In finite games, payoff functions are multilinear, so these two characterizations of "strictness" coincide (cf. Proposition 5.2); however, in general concave games, it is the variational characterization (6.21) that is less fragile – and, hence, more suitable for characterizing equilibria with robust deviation disincentives.

Our next result shows that, with high probability, employing (ML) with surjective choice maps leads to strict Nash equilibrium in a *finite* number of steps:

**Theorem 6.6.** *Fix a tolerance level $\delta > 0$ and suppose that (ML) is run with surjective choice maps and a sufficiently small step-size $\gamma_n$ satisfying $\sum_{n=0}^{\infty} \gamma_n^2 < \sum_{n=0}^{\infty} \gamma_n = \infty$. If $x^*$ is strict and (ML) is not initialized too far from $x^*$, we have*

$$\mathbb{P}(x_n \text{ converges to } x^* \text{ in a } \textit{finite} \text{ number of steps}) \geq 1 - \delta, \tag{6.22}$$

*provided that (H1)–(H4) hold. In addition, if $x^*$ is also globally stable, $x_n$ converges to $x^*$ in a finite number of steps from every initial condition (a.s.).*

*Proof.* To begin with, note that $v(x^*)$ lies in the interior of the polar cone $\mathrm{PC}(x^*)$ to $\mathcal{X}$ at $x^*$.[10] Hence, by continuity, there exists a neighborhood $U^*$ of $x^*$ such that $v(x) \in \mathrm{int}(\mathrm{PC}(x^*))$ for all $x \in U^*$. In turn, this implies that $\langle v(x) \,|\, x - x^* \rangle < 0$ for all $x \in U^* \setminus \{x^*\}$, i.e. $x^*$ is stable. Thus, by Theorem 4.11, there exists a neighborhood $U$ of $x^*$ such that, under the stated assumptions, $x_n$ converges to $x^*$ with probability at least $1 - \delta$.

Now, let $U' \subseteq U^*$ be a sufficiently small neighborhood of $x^*$ such that $\langle v(x) \,|\, z \rangle \leq -a\|z\|$ for some $a > 0$ and for all $z \in \mathrm{TC}(x^*)$.[11] Then, with probability at least $1-\delta$, there exists some (random) $n_0$ such that $x_n \in U'$ for all $n \geq n_0$, so $\langle v(x_n) \,|\, z \rangle \leq -a\|z\|$ for all $n \geq n_0$. Thus, for all $z \in \mathrm{TC}(x^*)$ with $\|z\| = 1$, we have

$$\begin{aligned} \langle y_n \,|\, z \rangle &= \langle y_{n_0} \,|\, z \rangle + \sum_{k=n_0}^{n-1} \gamma_k \langle v(x_k) \,|\, z \rangle + \sum_{k=n_0}^{n-1} \gamma_k \langle \xi_k \,|\, z \rangle \\ &\leq \|y_{n_0}\|_* - a \sum_{k=n_0}^{n-1} \gamma_k + \sum_{k=n_0}^{n-1} \gamma_k \langle \xi_k \,|\, z \rangle. \end{aligned} \tag{6.23}$$

By the law of large numbers for martingale differences (Hall and Heyde [15, Theorem 2.18]), we also have $\sum_{k=n_0}^{n-1} \gamma_k \xi_k / \sum_{k=n_0}^{n-1} \gamma_k \to 0$ (a.s.), so there exists some $n^*$ such that $\|\sum_{k=n_0}^{n-1} \gamma_k \xi_k\|_* \leq (a/2) \sum_{k=n_0}^{n-1} \gamma_k$ for all $n \geq n^*$ (a.s.). We thus obtain

$$\langle y_n \,|\, z \rangle \leq \|y_{n_0}\|_* - a \sum_{k=0}^{n-1} \gamma_k + \frac{a}{2} \|z\| \sum_{k=0}^{n-1} \gamma_k \leq \|y_{n_0}\|_* - \frac{a}{2} \sum_{k=0}^{n-1} \gamma_k, \tag{6.24}$$

showing that $\langle y_n \,|\, z \rangle \to -\infty$ uniformly in $z$ with probability at least $1 - \delta$.

To proceed, Proposition A.1 in Appendix A shows that $y^* + \mathrm{PC}(x^*) \subseteq Q^{-1}(x^*)$ whenever $Q(y^*) = x^*$. Since $Q$ is surjective, there exists some $y^* \in Q^{-1}(x^*)$, so it suffices to show that, with probability at least $1 - \delta$, $y_n$ lies in the pointed cone

[10] Indeed, if this were not the case, we would have $\langle v(x^*) \,|\, z \rangle = 0$ for some nonzero $z \in \mathrm{TC}(x^*)$.

[11] That such a neighborhood exists is a direct consequence of Definition 6.5.

$y^* + \mathrm{PC}(x^*)$ for all sufficiently large $n$. To do so, simply note that $y_n - y^* \in \mathrm{PC}(x^*)$ if and only if $\langle y_n - y^* \,|\, z\rangle \leq 0$ for all $z \in \mathrm{TC}(x^*)$ with $\|z\| = 1$. Since $\langle y_n \,|\, z\rangle$ converges uniformly to $-\infty$ with probability at least $1 - \delta$, our assertion is immediate.

Finally, for the global case, simply recall that $x_n$ converges to $x^*$ with probability 1 from any initial condition by Theorem 4.7. The argument above shows that $x_n = x^*$ for all large $n$, so we conclude that $x_n$ converges to $x^*$ in a finite number of steps (a.s.). ∎

As we discussed above, Theorem 6.6 extends the results of Cohen et al. [8] to general concave games that admit a strict equilibrium. In addition, it also drastically improves on the exponential $\mathcal{O}(e^{-a\sum_{k=0}^{n}\gamma_k})$ convergence rate of logit-based learning (Algorithm 2) by showing that convergence can be achieved in a *finite* number of steps if players employ surjective choice maps instead.

## 7. Discussion

A key question in the implementation of mirror descent methods is the optimum choice of penalty function, which determines the players' choice maps $Q_i \colon \mathcal{Y}_i \to \mathcal{X}_i$. From a qualitative point of view, our analysis shows that the specifics of this regularization process do not matter too much: the convergence results of Sections 4 and 5 hold for all choice maps of the form (3.6). Quantitatively however, the specific choice map employed by each player impacts the algorithm's convergence speed, and different choice maps could lead to vastly different rates of convergence.

As noted above, in the case of strict equilibria, this choice seems to favor nonsteep penalty functions (that is, surjective choice maps). Nonetheless, in the general case, the situation is less clear because of the dimensional dependence hidden in the $\Omega/K$ factor that appears e.g. in the mean rate guarantee (6.7). This factor depends crucially on the geometry of the players' action spaces and the underlying norm, and its optimum value may be attained by *steep* penalty functions – for instance, the entropic regularizer (3.8) is well known to be asymptotically optimal in the case of simplex-like feasible regions (Shalev-Shwartz [44, p. 140]).

Another challenge that arises in practice is that players may only be able to estimate their individual payoff gradients via (possibly imperfect) derivative-free observations of their realized, in-game payoffs. When this is the case, it is possible to employ a *simultaneous* stochastic approximation estimator like the one proposed by Spall [48] and Flaxman et al. [13]: roughly speaking, the idea there is to estimate the gradient of an objective function at a given point by sampling it at a nearby, randomly perturbed point, and then multiply the observed value by this random perturbation vector. The resulting gradient estimate has bounded variance but it also introduces a (controllably) small bias, so (H2) holds while (H1) fails – although only by a hair. We believe our convergence analysis can be extended to this case by properly controlling this "bias-variance" tradeoff and using more refined stochastic approximation arguments; we intend to explore this direction in future work.[12]

[12]The very recent preprint of Bervoets et al. [4] essentially solves this estimation problem in (strictly) concave games with one-dimensional action spaces. We believe that the work of Bervoets et al. [4] provides a very encouraging first step along the direction outlined above.

## Appendix A. Auxiliary results

In this appendix, we collect some auxiliary results that would have otherwise disrupted the flow of the main text. We begin with the basic properties of the Fenchel coupling:

*Proof of Proposition 4.3.* For our first claim, let $x = Q(y)$. Then, by definition

$$F(p,y) = h(p) + \langle y \,|\, Q(y)\rangle - h(Q(y)) - \langle y \,|\, p\rangle = h(p) - h(x) - \langle y \,|\, p - x\rangle. \tag{A.1}$$

Since $y \in \partial h(x)$ by Proposition 3.2, we have $\langle y \,|\, p - x\rangle = h'(x; p - x)$ whenever $x \in \mathcal{X}^\circ$, thus proving (4.9a). Furthermore, the strong convexity of $h$ also yields

$$\begin{aligned} h(x) + t\langle y \,|\, p - x\rangle &\leq h(x + t(p - x)) \\ &\leq t h(p) + (1-t)h(x) - \tfrac{1}{2}Kt(1-t)\|x - p\|^2, \end{aligned} \tag{A.2}$$

leading to the bound

$$\tfrac{1}{2}K(1-t)\|x - p\|^2 \leq h(p) - h(x) - \langle y \,|\, p - x\rangle = F(p,y) \tag{A.3}$$

for all $t \in (0,1]$. Eq. (4.9b) then follows by letting $t \to 0^+$ in (A.3).

Finally, for our third claim, we have

$$\begin{aligned} F(p,y') &= h(p) + h^*(y') - \langle y' \,|\, p\rangle \\ &\leq h(p) + h^*(y) + \langle y' - y \,|\, \nabla h^*(y)\rangle + \frac{1}{2K}\|y' - y\|_*^2 - \langle y' \,|\, p\rangle \\ &= F(p,y) + \langle y' - y \,|\, Q(y) - p\rangle + \frac{1}{2K}\|y' - y\|_*^2, \end{aligned} \tag{A.4}$$

where the inequality in the second line follows from the fact that $h^*$ is $(1/K)$-strongly smooth (Rockafellar and Wets [40, Theorem 12.60(e)]). ∎

Complementing Proposition 4.3, our next result concerns the inverse images of the choice map $Q$:

**Proposition A.1.** *Let $h$ be a penalty function on $\mathcal{X}$, and let $x^* \in \mathcal{X}$. If $x^* = Q(y^*)$ for some $y^* \in \mathcal{Y}$, then $y^* + \mathrm{PC}(x^*) \subseteq Q^{-1}(x^*)$.*

*Proof.* By Proposition 3.2, we have $x^* = Q(y)$ if and only if $y \in \partial h(x^*)$, so it suffices to show that $y^* + v \in \partial h(x^*)$ for all $v \in \mathrm{PC}(x^*)$. Indeed, we have $\langle v \,|\, x - x^*\rangle \leq 0$ for all $x \in \mathcal{X}$, so

$$h(x) \geq h(x^*) + \langle y^* \,|\, x - x^*\rangle \geq h(x^*) + \langle y^* + v \,|\, x - x^*\rangle. \tag{A.5}$$

The above shows that $y^* + v \in \partial h(x^*)$, as claimed. ∎

Our next result concerns the evolution of the Fenchel coupling under the dynamics (ML-C):

**Lemma A.2.** *Let $x(t) = Q(y(t))$ be a solution orbit of (ML-C). Then, for all $p \in \mathcal{X}$, we have*

$$\frac{d}{dt}F(p, y(t)) = \langle v(x(t)) \,|\, x(t) - p\rangle. \tag{A.6}$$

*Proof.* By definition, we have

$$\begin{aligned} \frac{d}{dt}F(p, y(t)) &= \frac{d}{dt}\left[h(p) + h^*(y(t)) - \langle y(t) \,|\, p\rangle\right] \\ &= \langle \dot{y}(t) \,|\, \nabla h^*(y(t))\rangle - \langle \dot{y}(t) \,|\, p\rangle = \langle v(x(t)) \,|\, x(t) - p\rangle, \end{aligned} \tag{A.7}$$

where, in the last line, we used Proposition 3.2. ∎

Our last auxiliary result shows that, if the sequence of play generated by (ML) is contained in the "basin of attraction" of a stable set $\mathcal{X}^*$, then it admits an accumulation point in $\mathcal{X}^*$:

**Lemma A.3.** *Suppose that $\mathcal{X}^* \subseteq \mathcal{X}$ is stable and (ML) is run with a step-size sequence such that $\sum_{k=0}^n \gamma_k^2 < \sum_{k=0}^n \gamma_k = \infty$. Assume further that $(x_n)_{n=0}^\infty$ is contained in a region $\mathcal{R}$ of $\mathcal{X}$ such that (VS) holds for all $x \in \mathcal{R}$. Then, under (H1) and (H2), every neighborhood $U$ of $\mathcal{X}^*$ is recurrent; specifically, there exists a subsequence $x_{n_k}$ of $x_n$ such that $x_{n_k} \to \mathcal{X}^*$ (a.s.). Finally, if (ML) is run with perfect gradient information ($\sigma_* = 0$), the above holds under the lighter assumption $\sum_{k=0}^n \gamma_k^2 / \sum_{k=0}^n \gamma_k \to 0$.*

*Proof of Lemma A.3.* Let $U$ be a neighborhood of $\mathcal{X}^*$ and assume to the contrary that $x_n \notin U$ for all sufficiently large $n$ with positive probability. By starting the sequence at a later index if necessary, we may further assume that $x_n \notin U$ for all $n$ without loss of generality. Thus, with $\mathcal{X}^*$ stable and $x_n \in \mathcal{R}$ for all $n$ by assumption, there exists some $a > 0$ such that

$$\langle v(x_n) \,|\, x_n - x^* \rangle \leq -a \quad \text{for all } x^* \in \mathcal{X}^* \text{ and for all } n. \tag{A.8}$$

As a result, for all $x^* \in \mathcal{X}^*$, we get

$$\begin{aligned} F(x^*, y_{n+1}) &= F(x^*, y_n + \gamma_n \hat{v}_n) \\ &\leq F(x^*, y_n) + \gamma_n \langle v(x_n) + \xi_n \,|\, x_n - x^* \rangle + \frac{1}{2K} \gamma_n^2 \|\hat{v}_n\|_*^2 \\ &\leq F(x^*, y_n) - a\gamma_n + \gamma_n \psi_n + \frac{1}{2K} \gamma_n^2 \|\hat{v}_n\|_*^2, \end{aligned} \tag{A.9}$$

where we used Proposition 4.3 in the second line and we set $\psi_n = \langle \xi_n \,|\, x_n - x^* \rangle$ in the third. Telescoping (A.9) then gives

$$F(x^*, y_{n+1}) \leq F(x^*, y_0) - \tau_n \left[ a - \frac{\sum_{k=0}^n \gamma_k \psi_k}{\tau_n} - \frac{1}{2K} \frac{\sum_{k=0}^n \gamma_k^2 \|\hat{v}_k\|_*^2}{\tau_n} \right], \tag{A.10}$$

where $\tau_n = \sum_{k=0}^n \gamma_k$.

Since $\mathbb{E}[\psi_n \,|\, \mathcal{F}_n] = \langle \mathbb{E}[\xi_n \,|\, \mathcal{F}_n] \,|\, x_n - x^* \rangle = 0$ and $\mathbb{E}[|\psi_n|^2 \,|\, \mathcal{F}_n] \leq \mathbb{E}[\|\xi_n\|_*^2 \|x_n - x^*\|^2 \,|\, \mathcal{F}_n] \leq \sigma_*^2 \|\mathcal{X}\|^2 < \infty$ by (H1) and (H2) respectively, the law of large numbers for martingale differences yields $\tau_n^{-1} \sum_{k=0}^n \gamma_k \psi_k \to 0$ (Hall and Heyde [15, Theorem 2.18]). Furthermore, letting $R_n = \sum_{k=0}^n \gamma_k^2 \|\hat{v}_k\|_*^2$, we also get

$$\mathbb{E}[R_n] \leq \sum_{k=0}^n \gamma_k^2 \, \mathbb{E}[\hat{v}_k]^2 \leq V_*^2 \sum_{k=0}^\infty \gamma_k^2 < \infty \quad \text{for all } n, \tag{A.11}$$

so Doob's martingale convergence theorem shows that $R_n$ converges (a.s.) to some random, finite value (Hall and Heyde [15, Theorem 2.5]). Combining the above, (A.10) gives $F(x^*, y_n) \sim -a\tau_n \to -\infty$ (a.s.), a contradiction.

Finally, if $\sigma_* = 0$, we also have $\psi_n = 0$ and $\|\hat{v}_n\|_*^2 = \|v(x_n)\|_*^2 \leq V_*^2$ for all $n$, so (A.10) yields $F(x^*, y_n) \to -\infty$ provided that $\tau_n^{-1} \sum_{k=0}^n \gamma_k^2 \to 0$, a contradiction. ∎

Finally, we turn to the characterization of strict equilibria in finite games:

*Proof of Proposition 5.2.* We will show that $(a) \Longrightarrow (b) \Longrightarrow (c) \Longrightarrow (a)$.

$(a) \implies (b)$. Suppose that $x^* = (s_1^*, \dots, s_N^*)$ is a strict equilibrium. Then, the weak inequality $\langle v(x^*) \,|\, z\rangle \leq 0$ follows from Proposition 2.1. For the strict part, if $z_i \in \mathrm{TC}_i(x_i^*)$ is nonzero for some $i \in \mathcal{N}$, we readily get

$$\langle v_i(x^*) \,|\, z_i\rangle = \sum_{s_i \neq s_i^*} z_{i,s_i} \left[u_i(s_i^*; s_{-i}^*) - u_i(s_i; s_{-i}^*)\right] < 0, \tag{A.12}$$

where we used the fact that $z_i$ is tangent to $\mathcal{X}$ at $x_i^*$, so $\sum_{s_i \in \mathcal{S}_i} z_{is_i} = 0$ and $z_{is_i} \geq 0$ for $s_i \neq s_i^*$, with at least one of these inequalities being strict when $z_i \neq 0$.

$(b) \implies (c)$. Property $(b)$ implies that $v(x^*)$ lies in the interior of the polar cone $\mathrm{PC}(x^*)$ to $\mathcal{X}$ at $x^*$. Since $\mathrm{PC}(x^*)$ has nonempty interior, continuity implies that $v(x)$ also lies in $\mathrm{PC}(x^*)$ for $x$ sufficiently close to $x^*$. We thus get $\langle v(x) \,|\, x - x^*\rangle \leq 0$ for all $x$ in a neighborhood of $x^*$, i.e. $x^*$ is stable.

$(c) \implies (a)$. Assume that $x^*$ is stable but not strict, so $u_{is_i}(x^*) = u_{is_i'}(x^*)$ for some $i \in \mathcal{N}$, and some $s_i \in \mathrm{supp}(x_i^*)$, $s_i' \in \mathcal{S}_i$. Then, if we take $x_i = x_i^* + \lambda(e_{is_i'} - e_{is_i})$ and $x_{-i} = x_{-i}^*$ with $\lambda > 0$ small enough, we get

$$\langle v(x) \,|\, x - x^*\rangle = \langle v_i(x) \,|\, x_i - x_i^*\rangle = \lambda u_{is_i'}(x^*) - \lambda u_{is_i}(x^*) = 0, \tag{A.13}$$

contradicting the assumption that $x^*$ is stable. This shows that $x^*$ is strict. ∎

## References


[1] Alvarez, Felipe, Jérôme Bolte, Olivier Brahic. 2004. Hessian Riemannian gradient flows in convex programming. *SIAM Journal on Control and Optimization* **43**(2) 477–501.

[2] Beck, Amir, Marc Teboulle. 2003. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters* **31**(3) 167–175.

[3] Benaïm, Michel. 1999. Dynamics of stochastic approximation algorithms. Jacques Azéma, Michel Émery, Michel Ledoux, Marc Yor, eds., *Séminaire de Probabilités XXXIII*, *Lecture Notes in Mathematics*, vol. 1709. Springer Berlin Heidelberg, 1–68.

[4] Bervoets, Sebastian, Mario Bravo, Mathieu Faure. 2016. Learning and convergence to Nash in network games with continuous action set. preprint.

[5] Bolte, Jérôme, Marc Teboulle. 2003. Barrier operators and associated gradient-like dynamical systems for constrained minimization problems. *SIAM Journal on Control and Optimization* **42**(4) 1266–1292.

[6] Bravo, Mario. 2016. An adjusted payoff-based procedure for normal form games. *Mathematics of Operations Research* forthcoming.

[7] Bravo, Mario, Panayotis Mertikopoulos. 2016. On the robustness of learning in games with stochastically perturbed payoff observations. *Games and Economic Behavior* doi:10.1016/j.geb.2016.06.004.

[8] Cohen, Johanne, Amélie Héliou, Panayotis Mertikopoulos. 2016. Exponentially fast convergence to (strict) equilibrium via hedging. http://arxiv.org/abs/1607.08863.

[9] Cominetti, Roberto, Emerson Melo, Sylvain Sorin. 2010. A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior* **70**(1) 71–83.

[10] Coucheney, Pierre, Bruno Gaujal, Panayotis Mertikopoulos. 2015. Penalty-regulated dynamics and robust learning procedures in games. *Mathematics of Operations Research* **40**(3) 611–633.

[11] Facchinei, Francisco, Christian Kanzow. 2007. Generalized Nash equilibrium problems. *4OR* **5**(3) 173–210.

[12] Facchinei, Francisco, Jong-Shi Pang. 2003. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research, Springer.

[13] Flaxman, Abraham D., Adam Tauman Kalai, H. Brendan McMahan. 2005. Online convex optimization in the bandit setting: gradient descent without a gradient. *SODA '05: Proceedings of the 16th annual ACM-SIAM symposium on discrete algorithms*. 385–394.

[14] Fudenberg, Drew, David K. Levine. 1998. *The Theory of Learning in Games*, *Economic learning and social evolution*, vol. 2. MIT Press, Cambridge, MA.

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

[16] Hart, Sergiu, Andreu Mas-Colell. 2000. A simple adaptive procedure leading to correlated equilibrium. *Econometrica* **68**(5) 1127–1150.

[17] Hofbauer, Josef, William H. Sandholm. 2002. On the global convergence of stochastic fictitious play. *Econometrica* **70**(6) 2265–2294.

[18] Hofbauer, Josef, William H. Sandholm. 2009. Stable games and their dynamics. *Journal of Economic Theory* **144**(4) 1665–1693.

[19] Hofbauer, Josef, Peter Schuster, Karl Sigmund. 1979. A note on evolutionarily stable strategies and game dynamics. *Journal of Theoretical Biology* **81**(3) 609–612.

[20] Hofbauer, Josef, Sylvain Sorin, Yannick Viossat. 2009. Time average replicator and best reply dynamics. *Mathematics of Operations Research* **34**(2) 263–269.

[21] Juditsky, Anatoli, Arkadi Semen Nemirovski, Claire Tauvel. 2011. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems* **1**(1) 17–58.

[22] Kalai, Adam Tauman, Santosh Vempala. 2005. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences* **71**(3) 291–307.

[23] Kiwiel, Krzysztof C. 1997. Free-steering relaxation methods for problems with strictly convex costs and linear constraints. *Mathematics of Operations Research* **22**(2) 326–349.

[24] Kwon, Joon, Panayotis Mertikopoulos. 2014. A continuous-time approach to online optimization. http://arxiv.org/abs/1401.6956.

[25] Laraki, Rida, Panayotis Mertikopoulos. 2013. Higher order game dynamics. *Journal of Economic Theory* **148**(6) 2666–2695.

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

[27] Littlestone, Nick, Manfred K. Warmuth. 1994. The weighted majority algorithm. *Information and Computation* **108**(2) 212–261.

[28] Maynard Smith, John, George R. Price. 1973. The logic of animal conflict. *Nature* **246** 15–18.

[29] McKelvey, Richard D., Thomas R. Palfrey. 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior* **10**(6) 6–38.

[30] Mertikopoulos, Panayotis, William H. Sandholm. 2016. Learning in games via reinforcement and regularization. *Mathematics of Operations Research* **41**(4) 1297–1324.

[31] Monderer, Dov, Lloyd S. Shapley. 1996. Potential games. *Games and Economic Behavior* **14**(1) 124 – 143.

[32] Nemirovski, Arkadi Semen, Anatoli Juditsky, Guangui (George) Lan, Alexander Shapiro. 2009. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization* **19**(4) 1574–1609.

[33] Nemirovski, Arkadi Semen, David Berkovich Yudin. 1983. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York, NY.

[34] Nesterov, Yurii. 2004. *Introductory Lectures on Convex Optimization: A Basic Course*. No. 87 in Applied Optimization, Kluwer Academic Publishers.

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

[36] Neyman, Abraham. 1997. Correlated equilibrium and potential games. *International Journal of Game Theory* **26**(2) 223–227.

[37] Perkins, Steven, David S. Leslie. 2012. Asynchronous stochastic approximation with differential inclusions. *Stochastic Systems* **2**(2) 409–446.

[38] Perkins, Steven, Panayotis Mertikopoulos, David S. Leslie. 2016. Mixed-strategy learning with continuous action sets. *IEEE Trans. Autom. Control* To appear.

[39] Rockafellar, Ralph Tyrrell. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

[40] Rockafellar, Ralph Tyrrell, Roger J. B. Wets. 1998. *Variational Analysis*, *A Series of Comprehensive Studies in Mathematics*, vol. 317. Springer-Verlag, Berlin.

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

[42] Sandholm, William H. 2015. Population games and deterministic evolutionary dynamics. H. Peyton Young, Shmuel Zamir, eds., *Handbook of Game Theory IV*. Elsevier, 703–778.

[43] Scutari, Gesualdo, Francisco Facchinei, Daniel Pérez Palomar, Jong-Shi Pang. 2010. Convex optimization, game theory, and variational inequality theory in multiuser communication systems. *IEEE Signal Process. Mag.* **27**(3) 35–49.

[44] Shalev-Shwartz, Shai. 2011. Online learning and online convex optimization. *Foundations and Trends in Machine Learning* **4**(2) 107–194.

[45] Shalev-Shwartz, Shai, Yoram Singer. 2007. Convex repeated games and Fenchel duality. *Advances in Neural Information Processing Systems 19*. MIT Press, 1265–1272.

[46] Shiryaev, Albert N. 1995. *Probability*. 2nd ed. Springer, Berlin.

[47] Sorin, Sylvain, Cheng Wan. 2016. Finite composite games: Equilibria and dynamics. *Journal of Dynamics and Games* **3**(1) 101–120.

[48] Spall, James C. 1997. A one-measurement form of simultaneous stochastic approximation. *Automatica* **33**(1) 109–112.

[49] Taylor, Peter D. 1979. Evolutionarily stable strategies with two types of player. *Journal of Applied Probability* **16**(1) 76–83.

[50] Viossat, Yannick, Andriy Zapechelnyuk. 2013. No-regret dynamics and fictitious play. *Journal of Economic Theory* **148**(2) 825–842.

[51] von Neumann, John. 1928. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* **100** 295–320. Translated by S. Bargmann as "On the Theory of Games of Strategy" in A. Tucker and R. D. Luce, editors, *Contributions to the Theory of Games IV*, volume 40 of *Annals of Mathematics Studies*, pages 13-42, 1957, Princeton University Press, Princeton.

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

[53] Zinkevich, Martin. 2003. Online convex programming and generalized infinitesimal gradient ascent. *ICML '03: Proceedings of the 20th International Conference on Machine Learning*. 928–936.

CNRS (French National Center for Scientific Research) and Univ. Grenoble Alpes, LIG, F–38000, Grenoble, France

*E-mail address*: panayotis.mertikopoulos@imag.fr