Math. Program., Ser. A (2019) 173:465–507
https://doi.org/10.1007/s10107-018-1254-8

FULL LENGTH PAPER

Learning in games with continuous action sets and unknown payoff functions

Panayotis Mertikopoulos[1] · **Zhengyuan Zhou**[2]

Received: 29 August 2016 / Accepted: 29 December 2017 / Published online: 12 March 2018

Abstract This paper examines the convergence of no-regret learning in games with continuous action sets. For concreteness, we focus on learning via "dual averaging", a widely used class of no-regret learning schemes where players take small steps along their individual payoff gradients and then "mirror" the output back to their action sets. In terms of feedback, we assume that players can only estimate their payoff gradients up to a zero-mean error with bounded variance. To study the convergence of the induced sequence of play, we introduce the notion of *variational stability*, and we show that stable equilibria are locally attracting with high probability whereas globally stable equilibria are globally attracting with probability 1. We also discuss some applications to mixed-strategy learning in finite games, and we provide explicit estimates of the method's convergence speed.

Keywords Continuous games · Dual averaging · Variational stability · Fenchel coupling · Nash equilibrium

The authors are indebted to the associate editor and two anonymous referees for their detailed suggestions and remarks. The paper has also benefited greatly from thoughtful comments by Jérôme Bolte, Nicolas Gast, Jérôme Malick, Mathias Staudigl, and the audience of the Paris Optimization Seminar.
P. Mertikopoulos was partially supported by the French National Research Agency (ANR) project ORACLESS (ANR–GAGA–13–JS01–0004–01) and the Huawei Innovation Research Program ULTRON.

✉ Panayotis Mertikopoulos
panayotis.mertikopoulos@imag.fr

Zhengyuan Zhou
zyzhou@stanford.edu

[1] CNRS, Inria, LIG, Univ. Grenoble Alpes, 38000 Grenoble, France

[2] Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

Mathematics Subject Classification Primary 91A26 · 90C15; Secondary 90C33 · 68Q32

1 Introduction

The prototypical setting of online optimization can be summarized as follows: at every stage $n = 1, 2, \dots$, of a repeated decision process, an agent selects an action X_n from some set $\mathcal{X}$ (assumed here to be convex and compact), and obtains a reward $u_n(X_n)$ determined by an *a priori* unknown payoff function $u_n \colon \mathcal{X} \to \mathbb{R}$. Subsequently, the agent receives some problem-specific feedback (for instance, an estimate of the gradient of u_n at X_n), and selects a new action with the goal of maximizing the obtained reward. Aggregating over the stages of the process, this is usually quantified by asking that the agent's *regret* $R_n \equiv \max_{x\in\mathcal{X}} \sum_{k=1}^{n} [u_k(x) - u_k(X_k)]$ grow sublinearly in n, a property known as "no regret".

In this general setting, the most widely used class of no-regret policies is the *online mirror descent* (OMD) method of Shalev-Shwartz [43] and its variants— such as "Following the Regularized Leader" [45], dual averaging [33,50], etc. Specifically, if the problem's payoff functions are concave, mirror descent guarantees an $\mathcal{O}(\sqrt{n})$ regret bound which is well-known to be tight in a "black-box" environment (i.e., without any further assumptions on u_n). Thus, owing to these guarantees, this class of first-order methods has given rise to an extensive literature in online learning and optimization; for a survey, see Shalev-Shwartz [44], Bubeck and Cesa-Bianchi [8], Hazan [15], and references therein.

In this paper, we consider a multi-agent extension of the above framework where the agents' rewards are determined by their individual actions and the actions of all other agents via a fixed mechanism: a *non-cooperative game*. Even though this mechanism may be unknown and/or opaque to the players, the additional structure it provides means that finer convergence criteria apply, chief among them being that of convergence to a *Nash equilibrium* (NE). We are thus led to the following fundamental question: *if all players of a repeated game employ a no-regret updating policy, do their actions converge to a Nash equilibrium of the underlying game?*

1.1 Summary of contributions

In general, the answer to this question is a resounding "no". Even in simple, finite games, no-regret learning may cycle [27] and its limit set may contain highly non-rationalizable strategies that assign positive weight *only* to strictly dominated strategies [48]. As such, our aim in this paper is twofold:

(i) to provide sufficient conditions under which no-regret learning converges to equilibrium; and
(ii) to assess the speed and robustness of this convergence in the presence of uncertainty, feedback noise, and other learning impediments.

Our contributions along these lines are as follows: First, in Sect. 2, we introduce an equilibrium stability notion which we call *variational stability* (VS), and which is

formally similar to (and inspired by) the seminal notion of *evolutionary stability* in population games [24]. [1] This stability notion extends the standard notion of operator monotonicity, so it applies in particular to all *monotone games* (that is, concave games that satisfy Rosen's [40] diagonal strict concavity condition). In fact, going beyond concave games, variational stability allows us to treat convergence questions in general games with continuous action spaces *without* having to restrict ourselves to a specific subclass (such as potential or common interest games).

Our second contribution is a detailed analysis of the long-run behavior of no-regret learning under variational stability. Regarding the information available to the players, our only assumption is that they have access to unbiased, bounded-variance estimates of their individual payoff gradients at each step; beyond this, we assume no prior knowledge of their payoff functions and/or the game. Despite this lack of information, variational stability guarantees that *(i)* the induced sequence of play converges globally to globally stable equilibria with probability 1 (Theorem 4.7); and *(ii)* it converges locally to locally stable equilibria with high probability (Theorem 4.11). As a corollary, if the game admits a (pseudo-)concave potential or if it is monotone, the players' actions converge to Nash equilibrium no matter the level of uncertainty affecting the players' feedback. In Sect. 5, we further extend these results to learning with imperfect feedback in finite games.

Our third contribution concerns the method's convergence speed. Mirroring a known result of Nesterov [33] for variational inequalities, we show that the gap from a stable state decays ergodically as $\mathcal{O}(1/\sqrt{n})$ if the method's step-size is chosen appropriately. Dually to this, we also show that the algorithm's expected running length until players reach an ε-neighborhood of a stable state is $\mathcal{O}(1/\varepsilon^2)$. Finally, if the stage game admits a *sharp* equilibrium (a straightforward extension of the notion of strict equilibrium in finite games), we show that, with probability 1, the process reaches an equilibrium in a *finite* number of steps.

Our analysis relies on tools and techniques from stochastic approximation, martingale limit theory and convex 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 hybridization of the Bregman divergence which provides a potent tool for proving convergence thanks to its Lyapunov properties.

1.2 Related work

Originally, mirror descent was introduced by Nemirovski and Yudin [29] for solving offline convex programs. The *dual averaging* (DA) variant that we consider here was pioneered by Nesterov [33] and proceeds as follows:[2] at each stage, the method takes

[1] Heuristically, variational stability is to games with a finite number of players and a continuum of actions what evolutionary stability is to games with a continuum of players and a finite action space. Our choice of terminology reflects precisely this analogy.

[2] In the online learning literature, dual averaging is sometimes called lazy mirror descent and can be seen as a linearized "Follow the Regularized Leader" (FTRL) scheme— for more details, we refer the reader to Beck and Teboulle [3], Xiao [50], and Shalev-Shwartz [44].

a gradient step in a dual space (where gradients live); the result is then mapped (or "mirrored") back to the problem's feasible region, a new gradient is generated, and the process repeats. The "mirroring" step above is itself determined by a strongly convex regularizer (or "distance generating") function: the squared Euclidean norm gives rise to Zinkevich's [51] online gradient descent algorithm, while the (negative) Gibbs entropy on the simplex induces the well-known exponential weights (EW) algorithm [2,49].

Nesterov [33] and Nemirovski et al. [30] provide several convergence results for dual averaging in (stochastic) convex programs and saddle-point problems, while Xiao [50] provides a thorough regret analysis for online optimization problems. In addition to treating the interactions of several competing agents at once, the fundamental difference of our paper with these works is that the convergence analysis in the latter is "ergodic", i.e., it concerns the time-averaged sequence $\bar{X}_n = \sum_{k=1}^{n} \gamma_k X_k / \sum_{k=1}^{n} \gamma_k$, and *not* the actual sequence of actions X_n employed by the players.

In online optimization, this averaging comes up naturally because the focus is on the players' regret. In the offline case, the points where an oracle is called during the execution of an algorithm do not carry any particular importance, so averaging provides a convenient way of obtaining convergence. However, in a game-theoretic setting, the figure of merit is the *actual sequence of play*, which determines the players' payoffs at each stage. The behavior of X_n may differ drastically from that of $\bar{X}_n$, so our treatment requires a completely different set of tools and techniques (especially in the stochastic regime).

Much of our analysis boils down to solving in an online way a (stochastic) variational inequality (VI) characterizing the game's Nash equilibria. Nesterov [32] and Juditsky et al. [19] provide efficient offline methods to do this, relying on an "extra-gradient" step to boost the convergence rate of the ergodic sequence $\bar{X}_n$. In our limited-feedback setting, we do not assume that players can make an extra oracle call to actions that were not actually employed, so the extrapolation results of Nesterov [32] and Juditsky et al. [19] do not apply. The single-call results of Nesterov [33] are closer in spirit to our paper but, again, they focus exclusively on monotone variational inequalities and the ergodic sequence $\bar{X}_n$— not the actual sequence of play X_n. All the same, for completeness, we make the link with ergodic convergence in Theorems 4.13 and 6.2.

When applied to mixed-strategy learning in finite games, the class of algorithms studied here has very close ties to the family of perturbed best response maps that arise in models of fictitious play and reinforcement learning [11,16,22]. Along these lines, Mertikopoulos and Sandholm [26] recently showed that a continuous-time version of the dynamics studied in this paper eliminates dominated strategies and converges to strict equilibria from all nearby initial conditions. Our analysis in Sect. 5 extends these results to a discrete-time, stochastic setting.

In games with continuous action sets, Perkins and Leslie [35] and Perkins et al. [36] examined a mixed-strategy actor-critic algorithm which converges to a probability distribution that assigns most weight to equilibrium states. At the pure strategy level, several authors have considered VI-based and Gauss–Seidel methods for solving generalized Nash equilibrium problems (GNEPs); for a survey, see Facchinei and Kanzow [12] and Scutari et al. [42]. The intersection of these works with the current

paper is when the game satisfies a global monotonicity condition similar to the diagonal strict concavity condition of Rosen [40]. However, the literature on GNEPs does not consider the implications for the players' regret, the impact of uncertainty and/or local convergence/stability issues, so there is no overlap with our results.

Finally, during the final preparation stages of this paper (a few days before the actual submission), we were made aware of a preprint by Bervoets et al. [5] examining the convergence of pure-strategy learning in strictly concave games with one-dimensional action sets. A key feature of the analysis of Bervoets et al. [5] is that players only observe their realized, in-game payoffs, and they choose actions based on their payoffs' variation from the previous period. The resulting mean dynamics boil down to an instantiation of dual averaging induced by the entropic regularization penalty $h(x) = x \log x$ (cf. Sect. 3), suggesting several interesting links with the current work.

1.3 Notation

Given a finite-dimensional vector space $\mathcal{V}$ with norm $\|\cdot\|$, we write $\mathcal{V}^*$ for its dual, $\langle y, x\rangle$ for the pairing between $y \in \mathcal{V}^*$ and $x \in \mathcal{V}$, and $\|y\|_* \equiv \sup\{\langle y, x\rangle : \|x\| \leq 1\}$ for the dual norm of y in $\mathcal{V}^*$. If $\mathcal{C} \subseteq \mathcal{V}$ is convex, we also write $\mathcal{C}^\circ \equiv \operatorname{ri}(\mathcal{C})$ for the relative interior of $\mathcal{C}$, $\|\mathcal{C}\| = \sup\{\|x' - x\| : x, x' \in \mathcal{C}\}$ for its diameter, and $\operatorname{dist}(\mathcal{C}, x) = \inf_{x' \in \mathcal{C}} \|x' - x\|$ for the distance between $x \in \mathcal{V}$ and $\mathcal{C}$.

For a given $x \in \mathcal{C}$, the *tangent cone* $\operatorname{TC}_{\mathcal{C}}(x)$ is defined as the closure of the set of all rays emanating from x and intersecting $\mathcal{C}$ in at least one other point; dually, the *polar cone* $\operatorname{PC}_{\mathcal{C}}(x)$ to $\mathcal{C}$ at x is defined as $\operatorname{PC}_{\mathcal{C}}(x) = \{y \in \mathcal{V}^* : \langle y, z\rangle \leq 0 \text{ for all } z \in \operatorname{TC}_{\mathcal{C}}(x)\}$. For concision, when $\mathcal{C}$ is clear from the context, we will drop it altogether and write $\operatorname{TC}(x)$ and $\operatorname{PC}(x)$ instead.

2 Preliminaries

2.1 Basic definitions and examples

Throughout this paper, we focus on games played by a finite set of *players* $i \in \mathcal{N} = \{1, \dots, N\}$. During play, each player selects an *action* x_i from a compact convex subset $\mathcal{X}_i$ of a finite-dimensional normed space $\mathcal{V}_i$, and their reward is determined by the profile $x = (x_1, \dots, x_N)$ of all players' actions— often denoted as $x \equiv (x_i; x_{-i})$ when we seek 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.

In more detail, 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* $u_i : \mathcal{X} \to \mathbb{R}$. In terms of regularity, we assume that u_i is continuously differentiable in x_i, and we write

$$v_i(x) \equiv \nabla_{x_i} u_i(x_i; x_{-i}) \tag{2.1}$$

for the *individual gradient* of u_i at x; we also assume that u_i and v_i are both continuous in x.[3] Putting all this together, a *continuous game* is a tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, (\mathcal{X}_i)_{i\in\mathcal{N}}, (u_i)_{i\in\mathcal{N}})$ with players, actions and payoffs defined as above.

As a special case, we will sometimes consider payoff functions that are *individually (pseudo-)concave* in the sense that

$$u_i(x_i; x_{-i}) \text{ is (pseudo-)concave in } x_i \text{ for all } x_{-i} \in \textstyle\prod_{j\neq i} \mathcal{X}_j,\ i \in \mathcal{N}. \tag{2.2}$$

When this is the case, we say that the game itself is (pseudo-)concave. Below, we briefly discuss some well-known examples of such games:

Example 2.1 (Mixed extensions of finite games) In a *finite game* $\Gamma \equiv \Gamma(\mathcal{N}, (\mathcal{A}_i)_{i\in\mathcal{N}}, (u_i)_{i\in\mathcal{N}})$, each player $i \in \mathcal{N}$ chooses an action α_i from a finite set $\mathcal{A}_i$ of "pure strategies" and no assumptions are made on the players' payoff functions $u_i : \mathcal{A} \equiv \prod_j \mathcal{A}_j \to \mathbb{R}$. Players can "mix" these choices by playing *mixed strategies*, i.e., probability distributions x_i drawn from the simplex $\mathcal{X}_i \equiv \Delta(\mathcal{A}_i)$. In this case (and in a slight abuse of notation), the expected payoff to player i in the mixed profile $x = (x_1, \dots, x_N)$ can be written as

$$u_i(x) = \sum_{\alpha_1\in\mathcal{A}_1} \cdots \sum_{\alpha_N\in\mathcal{A}_N} u_i(\alpha_1, \dots, \alpha_N)\, x_{1,\alpha_1} \cdots x_{N,\alpha_N}, \tag{2.3}$$

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

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

The resulting continuous game is called the *mixed extension* of Γ. Since $\mathcal{X}_i = \Delta(\mathcal{A}_i)$ is convex and u_i is linear in x_i, $\mathcal{G}$ is itself concave in the sense of (2.2).

Example 2.2 (Cournot competition) Consider the following Cournot oligopoly model: 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. This 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 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 is easily seen to be concave in the sense of (2.2).

[3] In the above, we tacitly assume that u_i is defined on an open neighborhood of $\mathcal{X}_i$. This allows us to use ordinary derivatives, 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 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 as an element of $\mathcal{V}_i^*$, the dual of $\mathcal{V}_i$.

Example 2.3 (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* $c_r : \mathbb{R}_+ \to \mathbb{R}$ (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 $\mathcal{A}_i \subseteq 2^{\mathcal{R}}$ of resource subsets α_i of $\mathcal{R}$— e.g., sets of links that form paths in the network. Then, the action space of player $i \in \mathcal{N}$ is the scaled simplex $\mathcal{X}_i = \rho_i \Delta(\mathcal{A}_i) = \{x_i \in \mathbb{R}_+^{\mathcal{A}_i} : \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} = \rho_i\}$ of *load distributions* over $\mathcal{A}_i$.

Given a load 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 $w_r = \sum_{i \in \mathcal{N}} \sum_{\alpha_i \ni r} x_{i\alpha_i}$ on said resource. This demand incurs a *cost* $c_r(w_r)$ per unit of load to each player utilizing resource r, where $c_r : \mathbb{R}_+ \to \mathbb{R}$ is a nondecreasing convex function. Accordingly, the total cost to player $i \in \mathcal{N}$ is

$$c_i(x) = \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} c_{i\alpha_i}(x), \tag{2.6}$$

where $c_{i\alpha_i}(x) = \sum_{r \in \alpha_i} c_r(w_r)$ denotes the cost incurred to player i by the utilization of $\alpha_i \subseteq \mathcal{R}$. The resulting *atomic splittable congestion game* $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, -c)$ is easily seen to be concave in the sense of (2.2).

2.2 Nash equilibrium

Our analysis focuses primarily on *Nash equilibria* (NE), i.e., strategy profiles that discourage unilateral deviations. Formally, $x^* \in \mathcal{X}$ is a *Nash equilibrium* if

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

Obviously, if x^* is a Nash equilibrium, we have the first-order condition

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

where $\mathrm{TC}_i(x_i^*)$ denotes the *tangent cone* to $\mathcal{X}_i$ at x_i^*. Therefore, if x^* is a Nash equilibrium, 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^* (cf. Fig. 1); moreover, the converse also holds if the game is pseudo-concave. We encode this more concisely as follows:

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

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

The converse also holds if the game is (pseudo-)concave in the sense of (2.2).

Remark 2.1 In the above (and in what follows), $v = (v_i)_{i \in \mathcal{N}}$ denotes the ensemble 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

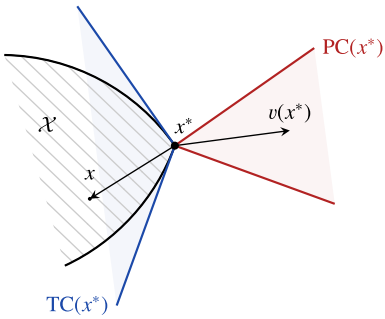

Fig. 1 Geometric characterization of Nash equilibria

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.

Proof of Proposition 2.1 If x^* is a Nash equilibrium, (2.8) is obtained by setting $z_i = x_i - x_i^*$ in (2.7) and summing over all $i \in \mathcal{N}$. Conversely, if (2.8) holds and the game is (pseudo-)concave, pick some $x_i \in \mathcal{X}_i$ and let $x = (x_i; x_{-i}^*)$ in (2.8). This gives $\langle v_i(x^*), x_i - x_i^*\rangle \leq 0$ for all $x_i \in \mathcal{X}_i$ so (NE) follows by the basic properties of (pseudo-)concave functions. □

Proposition 2.1 shows that Nash equilibria of concave games are precisely the solutions of the variational inequality (2.8), so existence follows from standard results. Using a similar variational characterization, Rosen [40] proved the following sufficient condition for equilibrium uniqueness:

Theorem 2.2 [40] *Assume that $\mathcal{G}$ 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.

Games satisfying (MC) are called *(strictly) monotone* and they enjoy properties similar to those of (strictly) convex functions.[4] In particular, letting $x'_{-i} = x_{-i}$, (MC) gives

$$\langle v_i(x_i'; x_{-i}) - v_i(x_i; x_{-i}), x_i' - x_i\rangle \leq 0 \quad \text{for all } x_i, x_i' \in \mathcal{X}_i, x_{-i} \in \mathcal{X}_{-i}, \tag{2.9}$$

implying in turn that $u_i(x)$ is (strictly) concave in x_i for all i. Therefore, *any game satisfying* (MC) *is also concave.*

[4] Rosen [40] originally referred to (MC) as diagonal strict concavity; Hofbauer and Sandholm [17] use the term "stable" for population games that satisfy a formal analogue of (MC), while Sandholm [41] and Sorin and Wan [47] call such games "contractive" and "dissipative" respectively. In all cases, the adverb "strictly" refers to the "only if" requirement in (MC).

Table 1 Monotonicity, stability, and Nash equilibrium: the existence of a concave potential implies monotonicity; monotonicity implies the existence of a globally stable point; and globally stable points are equilibria

		First-order requirement	Second-order test
Nash equilibrium	(NE)	$\langle v(x^*), x - x^* \rangle \leq 0$	N/A
Variational stability	(VS)	$\langle v(x), x - x^* \rangle \leq 0$	$H^{\mathcal{G}}(x^*) \prec 0$
Monotonicity	(MC)	$\langle v(x') - v(x), x' - x \rangle \leq 0$	$H^{\mathcal{G}}(x) \prec 0$
Concave potential	(PF)	$v(x) = \nabla f(x)$	$\nabla^2 f(x) \prec 0$

2.3 Variational stability

Combining Proposition 2.1 and (MC), it follows that the (necessarily unique) Nash equilibrium of a monotone game satisfies the inequality

$$\langle v(x), x - x^* \rangle \leq \langle v(x^*), x - x^* \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}. \tag{2.10}$$

In other words, if x^* is a Nash equilibrium of a monotone game, the players' individual payoff gradients "point towards" x^* in the sense that $v(x)$ forms an acute angle with $x^* - x$. Motivated by this, we introduce below the following relaxation of the monotonicity condition (MC) (Table 1):

Definition 2.3 We say that $x^* \in \mathcal{X}$ is *variationally stable* (or simply *stable*) if there exists a neighborhood U of x^* such that

$$\langle v(x), x - x^* \rangle \leq 0 \quad \text{for all } x \in U,$$

with equality if and only if $x = x^*$. In particular, if U can be taken to be all of $\mathcal{X}$, we say that x^* is *globally variationally stable* (or *globally stable) for short*.

Remark 2.2 The terminology "variational stability" alludes to the seminal notion of *evolutionary stability* introduced by Maynard Smith and Price [24] for population games (i.e., games with a continuum of players and a common, finite set of actions $\mathcal{A}$). Specifically, if $v(x) = (v_\alpha(x))_{\alpha \in \mathcal{A}}$ denotes the payoff field of such a game (with $x \in \Delta(\mathcal{A})$ denoting the state of the population), Definition 2.6 boils down to the variational characterization of evolutionarily stable state due to Hofbauer et al. [18]. As we show in the next sections, variational stability plays the same role for learning in games with continuous action spaces as evolutionary stability plays for evolution in games with a continuum of players.

By (2.10), a first example of variational stability is provided by the class of monotone games:

Corollary 2.4 *If $\mathcal{G}$ satisfies* (MC)*, its (unique) Nash equilibrium is globally stable.*

The converse to Corollary 2.4 does not hold, even partially. For instance, consider the single-player game with payoffs given by the function

$$u(x) = 1 - \sum_{\ell=1}^{d} \sqrt{1 + x_\ell}, \quad x \in [0, 1]^d. \tag{2.11}$$

In this simple example, the origin is the unique maximizer (and hence unique Nash equilibrium) of u. Moreover, we trivially have $\langle v(x), x\rangle = -2\sum_{\ell=1}^{d} x_\ell/\sqrt{1+x_\ell} \leq 0$ with equality if and only if $x = 0$, so the origin satisfies the global version of (VS); however, u is not even pseudo-concave if $d \geq 2$, so the game cannot be monotone. In words, (MC) is a sufficient condition for the existence of a (globally) stable state, but not a necessary one.

Nonetheless, even in this (non-monotone) example, variational stability characterizes the game's unique Nash equilibrium. We make this link precise below:

Proposition 2.5 *Suppose that $x^* \in \mathcal{X}$ is variationally stable. Then:*

a) If $\mathcal{G}$ is (pseudo-)concave, x^* is an isolated Nash equilibrium of $\mathcal{G}$.
b) If x^* is globally stable, it is the game's unique Nash equilibrium.

Proposition 2.5 indicates that variationally stable states are isolated (for the proof, see that of Proposition 2.7 below). However, this also means that Nash equilibria of games that admit a concave – but not *strictly* concave— potential may fail to be stable. To account for such cases, we will also consider the following setwise version of variational stability:

Definition 2.6 Let $\mathcal{X}^* \subseteq \mathcal{X}$ be closed and nonempty. We say that $\mathcal{X}^*$ is *variationally stable* (or simply *stable*) if there exists a neighborhood U of $\mathcal{X}^*$ such that

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

with equality for a given $x^* \in \mathcal{X}^*$ if and only if $x \in \mathcal{X}^*$. In particular, if U can be taken to be all of $\mathcal{X}$, we say that $\mathcal{X}^*$ is *globally variationally stable* (or *globally stable) for short*.

Obviously, Definition 2.6 subsumes Definition 2.3: if $x^* \in \mathcal{X}$ is stable in the pointwise sense of Definition 2.3, then it is also stable when viewed as a singleton set. In fact, when this is the case, it is also easy to see that x^* cannot belong to some larger variationally stable set,[5] so the notion of variational stability tacitly incorporates a certain degree of maximality. This is made clearer in the following:

Proposition 2.7 *Suppose that $\mathcal{X}^* \subseteq \mathcal{X}$ is variationally stable. Then:*

a) $\mathcal{X}^*$ is convex.
b) If $\mathcal{G}$ is concave, $\mathcal{X}^*$ is an isolated component of Nash equilibria.
c) If $\mathcal{X}^*$ is globally stable, it coincides with the game's set of Nash equilibria.

[5] In that case (VS) would give $\langle v(x'), x' - x^*\rangle = 0$ for some $x' \neq x^*$, a contradiction.

Proof of Proposition 2.7 To show that $\mathcal{X}^*$ is convex, 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 $\langle v(x_\lambda^*), x_\lambda^* - x_1^* \rangle = -(1-\lambda)\langle v(x_\lambda^*), x_1^* - x_0^* \rangle \leq 0$, implying that $\langle v(x_\lambda^*), x_1^* - x_0^* \rangle = 0$. Writing $x_1^* - x_0^* = \lambda^{-1}(x_\lambda^* - x_0^*)$, we then get $\langle v(x_\lambda^*), x_\lambda^* - x_0^* \rangle = 0$. By (VS), we must have $x_\lambda^* \in \mathcal{X}^*$ for all $\lambda \in [0,1]$, implying in turn that $\mathcal{X}^*$ is convex.

We now proceed to show that $\mathcal{X}^*$ only consists of Nash equilibria. To that end, asssume first that $\mathcal{X}^*$ is globally 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 $\tau \in [0,1]$, we have

$$\begin{aligned} \frac{d}{d\tau} u_i(x_i^* + \tau z_i; x_{-i}^*) &= \langle v_i(x_i^* + \tau z_i; x_{-i}^*), z_i \rangle \\ &= \frac{1}{\tau} \langle v_i(x_i^* + \tau z_i; x_{-i}^*), x_i^* + \tau z_i - x_i^* \rangle \leq 0, \end{aligned} \tag{2.12}$$

where the last inequality follows from (VS). In turn, this shows that $u_i(x_i^*; x_{-i}^*) \geq u_i(x_i^* + z_i; x_{-i}^*) = u_i(x_i; x_{-i}^*)$ for all $x_i \in \mathcal{X}_i$, $i \in \mathcal{N}$, i.e., x^* is a Nash equilibrium. Our claim for locally stable sets then follows by taking $\tau = 0$ above and applying Proposition 2.1.

We are left to show that there are no other Nash equilibria close to $\mathcal{X}^*$ (locally or globally). To do so, assume first that $\mathcal{X}^*$ is locally stable and let $x' \notin \mathcal{X}^*$ be 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 of $\mathcal{G}$ in U, i.e., $\mathcal{X}^*$ is an isolated set of Nash equilibria; the global version of our claim then follows by taking $U = \mathcal{X}$. □

2.4 Tests for variational stability

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

$$H_{ij}^{\mathcal{G}}(x) = \tfrac{1}{2} \nabla_{x_j} \nabla_{x_i} u_i(x) + \tfrac{1}{2} (\nabla_{x_i} \nabla_{x_j} u_j(x))^\top. \tag{2.13}$$

We then have:

Proposition 2.8 *If x^* is a Nash equilibrium of $\mathcal{G}$ and $H^{\mathcal{G}}(x^*) \prec 0$ on $\mathrm{TC}(x^*)$, then x^* is stable— and hence an isolated Nash equilibrium. In particular, if $H^{\mathcal{G}}(x) \prec 0$ on $\mathrm{TC}(x)$ for all $x \in \mathcal{X}$, x^* is globally stable— so it is the unique equilibrium of $\mathcal{G}$.*

Remark The requirement "$H^{\mathcal{G}}(x^*) \prec 0$ on $\mathrm{TC}(x^*)$" above means that $z^\top H^{\mathcal{G}}(x^*) z < 0$ for every nonzero tangent vector $z \in \mathrm{TC}(x^*)$.

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 [40], $\mathcal{G}$ 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 neighborhood $U = \prod_{i\in\mathcal{N}} U_i$ of x^* in $\mathcal{X}$. By the same theorem in Rosen [40], we get that (MC) holds locally in U, so the above reasoning shows that x^* is the unique equilibrium of the restricted game $\mathcal{G}|_U(\mathcal{N}, U, u|_U)$. Hence, x^* is locally stable and isolated in $\mathcal{G}$. □

We provide two straightforward applications of Proposition 2.8 below:

Example 2.4 (Potential games) Following Monderer and Shapley [28], a game $\mathcal{G}$ is called a *potential game* if it admits a *potential function* $f\colon \mathcal{X} \to \mathbb{R}$ such that

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

Local maximizers of f are Nash equilibria and the converse also holds if f is concave [34]. By differentiating (PF), it is easy to see that the Hessian of $\mathcal{G}$ is just the Hessian of its potential. Hence, if a game admits a concave potential f, the game's Nash set $\mathcal{X}^* = \arg\max_{x\in\mathcal{X}} f(x)$ is globally stable.

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

$$H_{ij}^{\mathcal{G}}(x) = \frac{1}{2}\frac{\partial^2 u_i}{\partial x_i \partial x_j} + \frac{1}{2}\frac{\partial^2 u_j}{\partial x_j \partial x_i} = -b_i\delta_{ij} - \tfrac{1}{2}(b_i + b_j), \tag{2.14}$$

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 (2.14) 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.[6] Quantitatively, if the coefficients b_i are independent and identically distributed (i.i.d.) on $[0, 1]$, a Monte Carlo simulation shows that (2.14) is negative-definite with probability between 65% and 75% for $N \in \{2, \dots, 100\}$.

3 Learning via dual averaging

In this section, we adapt the widely used *dual averaging* (DA) method of Nesterov [33] to our game-theoretic setting.[7] Intuitively, the main idea is as follows: At each stage of the process, every player $i \in \mathcal{N}$ gets an estimate $\hat{v}_i$ of the individual gradient of their payoff function at the current action profile, possibly subject to noise and uncertainty. Subsequently, they take a step along this estimate 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 for the next stage and continue playing (for a schematic illustration, see Fig. 2).

[6] This is so because, in the symmetric case, the RHS of (2.14) is a circulant matrix with eigenvalues $-b$ and $-(N+1)b$.

[7] In optimization, the roots of the method can be traced back to Nemirovski and Yudin [29]; see also [3], Nemirovski et al. [30] and Shalev-Shwartz [44].

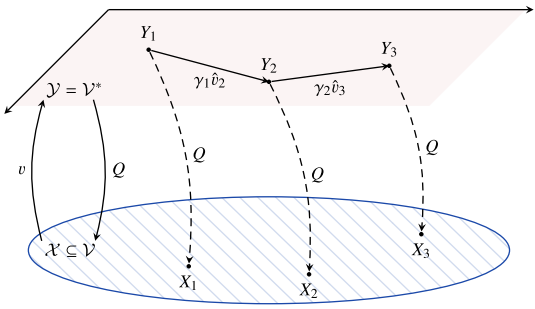

Fig. 2 Schematic representation of dual averaging

Formally, starting with some arbitrary (and possibly uninformed) gradient estimate $Y_1 = \hat{v}_1$ at $n = 1$, this scheme can be described via the recursion

$$\begin{aligned} X_{i,n} &= Q_i(Y_{i,n}), \\ Y_{i,n+1} &= Y_{i,n} + \gamma_n \hat{v}_{i,n+1}, \end{aligned} \tag{DA}$$

where:

(1) n denotes the stage of the process.
(2) $\hat{v}_{i,n+1} \in \mathcal{V}_i^*$ is an 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 i-th player's individual gradient steps.
(4) $\gamma_n > 0$ is a nonincreasing step-size sequence, typically of the form $1/n^\beta$ for some $\beta \in (0, 1]$.
(5) $Q_i : \mathcal{V}_i^* \to \mathcal{X}_i$ is the *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 (DA) are *a)* the players' gradient estimates; and *b)* the choice maps that determine the players' actions. In the rest of this section, we discuss both in detail.

3.1 Feedback and uncertainty

Regarding the players' individual gradient observations, we assume that each player $i \in \mathcal{N}$ has access to a "black box" feedback mechanism— an *oracle*— which returns an estimate of their payoff gradients at their current action profile. Of course, this information may be imperfect for a multitude of reasons: for instance *(i)* 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 be able to observe the realized gradients $\nabla_{x_i} \hat{u}_i(x; \omega)$.

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

$$\hat{v}_{i,n+1} = v_i(X_n) + \xi_{i,n+1}, \tag{3.2}$$

where the noise process $\xi_n = (\xi_{i,n})_{i\in\mathcal{N}}$ is an L^2-bounded martingale difference adapted to the history $(\mathcal{F}_n)_{n=1}^{\infty}$ of X_n (i.e., ξ_n is $\mathcal{F}_n$-measurable but ξ_{n+1} isn't).[8] More explicitly, this means that ξ_n satisfies the statistical hypotheses:

1. *Zero-mean:*

$$\mathbb{E}\big[\xi_{n+1} \,|\, \mathcal{F}_n\big] = 0 \qquad \text{for all } n = 1, 2, \dots \text{(a.s.)}. \tag{H1}$$

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

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

Alternatively, (H1) and (H2) simply posit that the players' individual gradient estimates are *conditionally unbiased and bounded in mean square*, viz.

$$\mathbb{E}[\hat{v}_{n+1} \,|\, \mathcal{F}_n] = v(X_n), \tag{3.3a}$$

$$\mathbb{E}[\|\hat{v}_{n+1}\|_*^2 \,|\, \mathcal{F}_n] \leq V_*^2 \quad \text{for some finite } V_* > 0. \tag{3.3b}$$

The above allows for a broad range of error processes, including all compactly supported, (sub-)Gaussian, (sub-)exponential and log-normal distributions.[9] In fact, both hypotheses can be relaxed (for instance, by assuming a small bias or asking for finite moments up to some order $q < 2$), but we do not do so to keep things simple.

3.2 Choosing actions

Given that the players' score variables 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 would be an extreme point of $\mathcal{X}$, so (DA) could never converge to an interior point. Thus, instead of taking a "hard" arg max approach, we will focus on *regularized* maps of the form

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

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

[8] Indices have been chosen so that all relevant processes are $\mathcal{F}_n$-measurable at stage n.

[9] 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.

Definition 3.1 Let $\mathcal{C}$ be a compact convex subset of a finite-dimensional normed space $\mathcal{V}$. We say that $h : \mathcal{C} \to \mathbb{R}$ is a *regularizer* (or *penalty function*) 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* $Q : \mathcal{V}^* \to \mathcal{C}$ *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 $h_i : \mathcal{X}_i \to \mathbb{R}$ 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 $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$ 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 $K \equiv \min_i K_i$.[10] The induced choice map is simply $Q \equiv (Q_1, \ldots, Q_N)$ so we will write $x = Q(y)$ for the action profile induced by the score vector $y = (y_1, \ldots, y_N) \in \mathcal{Y} \equiv \prod_i \mathcal{Y}_i$.

Remark 3.1 In finite games, McKelvey and Palfre [25] 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 Hofbauer and Sandholm [16] and Mertikopoulos and Sandholm [26].

We discuss below a few examples of this regularization process:

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

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

The induced learning scheme (cf. Algorithm 1) may thus be viewed as a multi-agent variant of gradient ascent with lazy projections [51]. For future reference, note that h is differentiable on $\mathcal{X}$ and $\Pi_{\mathcal{X}}$ is *surjective* (i.e., 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 Δ is provided by the (negative) Gibbs entropy

[10] 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 Dual averaging with Euclidean projections (Example 3.1).

Require: step-size sequence $\gamma_n \propto 1/n^\beta$, $\beta \in (0, 1]$; initial scores $Y_i \in \mathcal{Y}_i$

for $n = 1, 2, \dots$ **do**
  for every player $i \in \mathcal{N}$ **do**
    play $X_i \leftarrow \Pi_{\mathcal{X}_i}(Y_i)$; {choose an action}
    observe $\hat{v}_i$; {estimate gradient}
    update $Y_i \leftarrow Y_i + \gamma_n \hat{v}_i$; {take gradient step}
  end for
end for

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

The entropic regularizer (3.8) is 1-strongly convex with respect to the L^1-norm on $\mathbb{R}^d$. Moreover, a straightforward calculation shows that the induced choice map is

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

This model is known as *logit choice* and the associated learning scheme has been studied extensively in evolutionary game theory and online learning; for a detailed account, see Vovk [49], Littlestone and Warmuth [23], Laraki and Mertikopoulos [21], and references therein. In contrast to the previous example, h is differentiable *only* on the relative interior Δ° of Δ and im $\Lambda = \Delta^\circ$ (i.e., Λ is "essentially" surjective).

3.3 Surjectivity versus steepness

We close this section with an important link between the boundary behavior of penalty functions and the surjectivity of 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 h is called *subdifferentiable* at $x \in \mathcal{X}$ whenever $\partial h(x)$ is nonempty. This is always the case if $x \in \mathcal{X}^\circ$, so $\mathcal{X}^\circ \subseteq \operatorname{dom} \partial h \equiv \{x \in \mathcal{X} : \partial h(x) \neq \varnothing\} \subseteq \mathcal{X}$ [38, Chap. 26].

Intuitively, h fails to be subdifferentiable at a boundary point $x \in \operatorname{bd}(\mathcal{X})$ only 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 proposition shows that regularizers that are everywhere nonsteep (as in Example 3.1) induce choice maps that are surjective; on the other hand, regularizers 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 regularizer 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 folklore in optimization and convex analysis; for a proof, see Rockafellar [38, Theorem 23.5] and Rockafellar and Wets [39, Theorem 12.60(b)].

4 Convergence analysis

A key property of (DA) in concave games is that it leads to *no regret*, viz.

$$\max_{x_i \in \mathcal{X}_i} \sum_{k=1}^{n} \big[u_i(x_i; X_{-i,k}) - u_i(X_k)\big] = o(n) \quad \text{for all } i \in \mathcal{N}, \tag{4.1}$$

provided that the algorithm's step-size is chosen appropriately— for a precise statement, see Xiao [50] and Shalev-Shwartz [44]. As such, under (DA), every player's average payoff matches asymptotically that of the best fixed action in hindsight (though, of course, this does not take into account changes to other players' actions due to a change in a given player's chosen action).

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

4.1 Limit states

We first show that if the sequence of play induced by (DA) converges to some $x^* \in \mathcal{X}$ with positive probability, this limit is a Nash equilibrium:

Theorem 4.1 *Suppose that* (DA) *is run with imperfect gradient information satisfying* (H1)–(H2) *and a step-size sequence γ_n such that*

$$\sum_{n=1}^{\infty} \left(\frac{\gamma_n}{\tau_n}\right)^2 < \sum_{n=1}^{\infty} \gamma_n = \infty, \tag{4.2}$$

where $\tau_n = \sum_{k=1}^{n} \gamma_k$. *If the game is (pseudo-)concave and* X_n *converges to* $x^* \in \mathcal{X}$ *with positive probability,* x^* *is a Nash equilibrium.*

Remark 4.1 Note here that the requirement (4.2) holds for every step-size policy of the form $\gamma_n \propto 1/n^\beta$, $\beta \leq 1$ (i.e. even for *increasing* γ_n).

Proof of Theorem 4.1 Let $v^* = v(x^*)$ and assume ad absurdum that x^* is not a Nash equilibrium. By the characterization (2.7) of Nash equilibria, there exists a player $i \in \mathcal{N}$ and a 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

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

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

Now, let Ω_0 be the event that X_n converges to x^*, so $\mathbb{P}(\Omega_0) > 0$ by assumption. Within Ω_0, we may assume for simplicity that $X_n \in U$ and $v(X_n) \in V$ for all n, so (DA) yields

$$Y_{n+1} = Y_1 + \sum_{k=1}^{n} \gamma_k \hat{v}_{k+1} = Y_{n_0} + \sum_{k=1}^{n} \gamma_k \left[v(X_k) + \xi_{k+1}\right] = Y_1 + \tau_n \bar{v}_{n+1}, \tag{4.4}$$

where we set $\bar{v}_{n+1} = \tau_n^{-1} \sum_{k=1}^{n} \gamma_k \hat{v}_{k+1} = \tau_n^{-1} \sum_{k=1}^{n} \gamma_k \left[v(X_k) + \xi_{k+1}\right]$.

We now claim that $\mathbb{P}(\bar{v}_n \to v^* \mid \Omega_0) = 1$. Indeed, by (4.2) and (H2), we have

$$\sum_{n=1}^{\infty} \frac{1}{\tau_n^2} \mathbb{E}[\|\gamma_n \xi_{n+1}\|_*^2 \mid \mathcal{F}_n] \leq \sum_{n=1}^{\infty} \frac{\gamma_n^2}{\tau_n^2} \sigma^2 < \infty. \tag{4.5}$$

Therefore, by the law of large numbers for martingale difference [14, Theorem 2.18], we obtain $\tau_n^{-1} \sum_{k=1}^{n} \gamma_k \xi_{k+1} \to 0$ (a.s.). Given that $v(X_n) \to v^*$ in Ω_0 and $\mathbb{P}(\Omega_0) > 0$, we infer that $\mathbb{P}(\bar{v}_n \to v^* \mid \Omega_0) = 1$, as claimed.

Now, with $Y_{i,n} \in \partial h_i(X_{i,n})$ by Proposition 3.2, we also have

$$h_i(q_i) - h_i(X_{i,n}) \geq \langle Y_{i,n}, q_i - X_{i,n} \rangle = \langle Y_{i,1}, q_i - X_{i,n} \rangle + \tau_{n-1} \langle \bar{v}_{i,n}, q_i - X_{i,n} \rangle. \tag{4.6}$$

Since $\bar{v}_n \to v^*$ almost surely on Ω_0, (4.3) yields $\langle \bar{v}_{i,n}, q_i - X_{i,n} \rangle \geq c > 0$ for all sufficiently large n. However, given that $|\langle Y_{i,1}, q_i - X_{i,n} \rangle| \leq \|Y_{i,1}\|_* \|q_i - X_{i,n}\| \leq \|Y_{i,1}\|_* \|\mathcal{X}\| = \mathcal{O}(1)$, a simple substitution in (4.6) yields $h_i(q_i) - h_i(X_{i,n}) \gtrsim c\tau_n \to \infty$ with positive probability, a contradiction. We conclude that x^* is a Nash equilibrium of $\mathcal{G}$, as claimed. □

4.2 The Fenchel coupling

A key tool in establishing the convergence properties of (DA) is the so-called *Bregman divergence* $D(p, x)$ between a given base point $p \in \mathcal{X}$ and a test state $x \in \mathcal{X}$. Following Kiwiel [20], $D(p, x)$ is defined as the difference between $h(p)$ and the best linear approximation of $h(p)$ from x, viz.

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

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 $D(p, X_n) \to 0$ [20]. Accordingly, 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 (DA) when $X_n = Q(Y_n)$ is not interior. Instead, given that (DA) mixes 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 [26] 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 key 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:*

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

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

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

Proposition 4.3 (proven in Appendix 1) justifies the terminology "primal-dual divergence" and plays a key role in our analysis. Specifically, given a sequence Y_n in $\mathcal{Y}$, (4.9b) yields $Q(Y_n) \to p$ whenever $F(p, Y_n) \to 0$, meaning that $F(p, Y_n)$ can be used to test the convergence of $Q(Y_n)$ to p.

For technical reasons, it is convenient to also assume the converse, namely that

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

When h is steep, we have $F(p, y) = D(p, Q(y))$ for all $y \in \mathcal{Y}$, so (H3) boils down to the requirement

$$D(p, x_n) \to 0 \quad \text{whenever} \quad X_n \to p. \tag{4.10}$$

This so-called "reciprocity condition" is well known in the theory of Bregman functions [1,9,20]: essentially, it means that the sublevel sets of $D(p, \cdot)$ are neighborhoods of p in $\mathcal{X}$. Hypothesis (H3) posits that the *images* of the sublevel sets of $F(p, \cdot)$ under Q are neighborhoods of p in $\mathcal{X}$, so it may be seen as a "primal-dual" variant of Bregman reciprocity. Under this light, it is easy to check that Example 3.1 and 3.2 both satisfy (H3).

Obviously, when (H3) holds, Proposition 4.3 gives:

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

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

$$F(\mathcal{C}, y) = \inf\{F(p, y) : p \in \mathcal{C}\}, \quad \mathcal{C} \subseteq \mathcal{X}, y \in \mathcal{Y}. \tag{4.11}$$

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, if* (H3) *holds, the converse is also true.*

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* (DA) *is run with perfect feedback* ($\sigma = 0$)*, choice maps satisfying* (H3)*, and a step-size* γ_n *such that* $\sum_{k=1}^{n} \gamma_k^2 \big/ \sum_{k=1}^{n} \gamma_k \to 0$*. If the set* $\mathcal{X}^*$ *of the game's Nash equilibria is globally stable,* X_n *converges to* $\mathcal{X}^*$*.*

Proof Let $\mathcal{X}^*$ be the game's set of Nash equilibria, fix some arbitrary $\varepsilon > 0$, 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

$$F(x^*, Y_{n+1}) \leq F(x^*, Y_n) + \gamma_n \langle v(X_n), X_n - x^* \rangle + \frac{\gamma_n^2}{2K} \|v(X_n)\|_*^2. \tag{4.12}$$

Table 2 Overview of the various regularity hypotheses used in the paper

	Hypothesis	Precise statement
(H1)	Zero-mean errors	$\mathbb{E}[\xi_{n+1} \mid \mathcal{F}_n] = 0$
(H2)	Finite error variance	$\mathbb{E}[\|\xi_{n+1}\|_*^2 \mid \mathcal{F}_n] \leq \sigma^2$
(H3)	Bregman reciprocity	$F(p, y_n) \to 0$ whenever $Q(y_n) \to p$
(H4)	Lipschitz gradients	$v(x)$ is Lipschitz continuous

To proceed, assume inductively that $X_n \in U_\varepsilon$. By (H3), there exists some $\delta > 0$ such that $\mathrm{cl}(U_{\varepsilon/2})$ contains a δ-neighborhood of $\mathcal{X}^*$.[11] Consequently, with $\mathcal{X}^*$ globally stable, there exists some $c \equiv c(\varepsilon) > 0$ such that

$$\langle v(x), x - x^* \rangle \leq -c \quad \text{for all } x \in U_\varepsilon - U_{\varepsilon/2}, x^* \in \mathcal{X}^*. \tag{4.13}$$

If $X_n \in U_\varepsilon - U_{\varepsilon/2}$ and $\gamma_n \leq 2cK/V_*^2$, (4.12) yields $F(x^*, Y_{n+1}) \leq F(x^*, Y_n)$.[12] Hence, minimizing over $x^* \in \mathcal{X}^*$, we get $F(\mathcal{X}^*, Y_{n+1}) \leq F(\mathcal{X}^*, Y_n) < \varepsilon$, so $X_{n+1} = Q(Y_{n+1}) \in U_\varepsilon$. Otherwise, if $X_n \in U_{\varepsilon/2}$ and $\gamma_n^2 < \varepsilon K/V_*^2$, combining (VS) with (4.12) 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 $\gamma_n < \min\{2cK/V_*^2, \sqrt{\varepsilon K}/V_*\}$.

To complete the proof, Lemma A.3 shows that X_n visits U_ε infinitely often under the stated assumptions. Since $\gamma_n \to 0$, 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 (Table 2):

Theorem 4.7 *Suppose that* (DA) *is run with a step-size sequence* γ_n *such that* $\sum_{n=1}^\infty \gamma_n^2 < \infty$ *and* $\sum_{n=1}^\infty \gamma_n = \infty$. *If* (H1)–(H4) *hold and the set* $\mathcal{X}^*$ *of the game's Nash equilibria is globally stable,* X_n *converges to* $\mathcal{X}^*$ *(a.s.).*

Corollary 4.8 *If* $\mathcal{G}$ *satisfies* (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 start with the continuous-time system

[11] 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).

[12] Since $\sigma = 0$, we can take here $V_* = \max_{x \in \mathcal{X}} \|v(x)\|_*$.

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

which can be seen as a "mean-field" approximation of the recursive scheme (DA). As we show in Appendix 1, the orbits $x(t) = Q(y(t))$ of (DA-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 (DA) comprises an *asymptotic pseudotrajectory* (APT) of the dynamics (DA-c), i.e. Y_n asymptotically tracks the flow of (DA-c) with arbitrary accuracy over windows of arbitrary length Benaïm [4].[13] APTs have the key property that, in the presence of a global attractor, they cannot stray too far from the flow of (DA-c); however, given that Q may fail to be invertible, the trajectories $x(t) = Q(y(t))$ do *not* consitute a semiflow, so it is not possible to leverage the general stochastic approximation theory of Benaïm [4]. 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 (DA) 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 $\Phi_t : \mathcal{Y} \to \mathcal{Y}$ for the semiflow induced by (DA-c) on $\mathcal{Y}$ – i.e. $(\Phi_t(y))_{t\geq 0}$ is the solution orbit of (DA-c) that starts at $y \in \mathcal{Y}$.[14]

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 $c \equiv c(\varepsilon) > 0$ such that

$$\langle v(x), x - x^* \rangle \leq -c \quad \text{for all } x^* \in \mathcal{X}^*, x \notin U_\varepsilon. \tag{4.14}$$

Consequently, if $\tau_y = \inf\{t > 0 : Q(\Phi_t(y)) \in U_\varepsilon\}$ denotes the first time at which an orbit of (DA-c) reaches U_ε, Lemma A.2 in Appendix 1 gives:

$$F(x^*, \Phi_t(y)) \leq F(x^*, y) - ct \quad \text{for all } x^* \in \mathcal{X}^*, t \leq \tau_y. \tag{4.15}$$

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

1. $\tau_y \geq \tau$: then, (4.15) 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 (DA), i.e. Y is the continuous curve which joins the values Y_n at all times $\tau_n = \sum_{k=1}^{n} \gamma_k$. Under the stated assumptions, a standard result of Benaïm ([4], Propositions 4.1 and 4.2) shows that $Y(t)$ is an asymptotic pseudotrajectory of Φ, i.e.

$$\lim_{t\to\infty} \sup_{0\leq h\leq T} \|Y(t+h) - \Phi_h(Y(t))\|_* = 0 \quad \text{for all } T > 0 \text{ (a.s.)}. \tag{4.16}$$

[13] For a precise definition, see (4.16) below.

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

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

$$\begin{aligned} F(x^*, Y(t+h)) &\leq F(x^*, \Phi_h(Y(t))) \\ &\quad + \langle Y(t+h) - \Phi_h(Y(t)), Q(\Phi_h(Y(t))) - x^* \rangle \\ &\quad + \frac{1}{2K}\|Y(t+h) - \Phi_h(Y(t))\|_*^2 \\ &\leq F(x^*, \Phi_h(Y(t))) + \delta\|\mathcal{X}\| + \frac{\delta^2}{2K} \\ &\leq F(x^*, \Phi_h(Y(t))) + \varepsilon. \end{aligned} \tag{4.17}$$

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

$$F(\mathcal{X}^*, Y(t+h)) \leq F(\mathcal{X}^*, \Phi_h(Y(t))) + \varepsilon \quad \text{for all } t \geq t_0. \tag{4.18}$$

By Lemma A.3, 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. 4.18 yields $F(\mathcal{X}^*, Y(t+h)) \leq 2\varepsilon + \varepsilon = 3\varepsilon$ for all $h \in [0, \tau]$. However, by the definition of τ, 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 $h \in [k\tau, (k+1)\tau]$, $k = 1, 2, \dots$ (a.s.). Since ε has been chosen arbitrarily, we conclude that $F(\mathcal{X}^*, Y_n) \to 0$, so $X_n \to \mathcal{X}^*$ by Proposition 4.5. □

We close this section with a few remarks:

Remark 4.2 In the above, the Lipschitz continuity assumption (H4) is used to show that the sequence X_n comprises an APT of the continuous-time dynamics (DA-c). Since any continuous functions on a compact set is uniformly continuous, the proof of Proposition 4.1 in Benaïm [4, p. 14] shows that (H4) can be dropped altogether if (DA-c) is well-posed (which, in turn, holds if $v(x)$ is only *locally* Lipschitz). Albeit less general, Lipschitz continuity is more straightforward as an assumption, so we do not go into the details of this relaxation.

We should also note that several classic convergence results for dual averaging and mirror descent do not require Lipschitz continuity at all (see e.g. [30,33]). The reason for this is that these results focus on the convergence of the averaged sequence $\bar{X}_n = \sum_{k=1}^n \gamma_k X_k \big/ \sum_{k=1}^n \gamma_k$, whereas the figure of merit here is the *actual* sequence of play X_n. The latter sequence is more sensitive to noise, hence the need for additional regularity; in our ergodic analysis later in the paper, (H4) is not invoked.

Remark 4.3 Theorem 4.7 shows that (DA) converges to equilibrium, but the summability requirement $\sum_{n=1}^\infty \gamma_n^2 < \infty$ suggests that players must be more conservative 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 $\gamma_n \propto 1/n^\beta$, $\beta \in (0, 1]$;

however, in the presence of errors and uncertainty, Theorem 4.7 guarantees convergence only when $\beta \in (1/2, 1]$.

The "critical" value $\beta = 1/2$ 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 can be used to show that Theorem 4.7 still holds under the lighter requirement $\sum_{n=1}^{\infty} \gamma_n^{1+q/2} < \infty$. Thus, even in the presence of noise, it is possible to employ (DA) with any step-size sequence of the form $\gamma_n \propto 1/n^{\beta}$, $\beta \in (0, 1]$, provided that the noise process ξ_n has $\mathbb{E}[\|\xi_{n+1}\|_*^q \mid \mathcal{F}_n] < \infty$ for some $q > 2/\beta - 2$. In particular, if the noise affecting the players' observations has finite moments of all orders (for instance, if ξ_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 (DA) converges globally to states (or sets) that are globally stable, even under noise and uncertainty. In this section, we show that (DA) 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* (DA) *is run with perfect feedback* ($\sigma = 0$)*, choice maps satisfying* (H3)*, and a sufficiently small step-size with* $\sum_{k=1}^{n} \gamma_k^2 \big/ \sum_{k=1}^{n} \gamma_k \to 0$*. If* $\mathcal{X}^*$ *is a stable set of Nash equilibria, there exists a neighborhood* U *of* $\mathcal{X}^*$ *such that* X_n *converges to* $\mathcal{X}^*$ *whenever* $X_1 \in U$*.*

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 $c > 0$ satisfying (4.13) and such that (VS) holds throughout U_ε. If $X_1 \in U_\varepsilon$ and $\gamma_1 \leq \min\{2cK/V_*^2, \sqrt{\varepsilon K}/V_*\}$, 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_ε, Lemma A.3 shows that X_n visits any neighborhood of $\mathcal{X}^*$ infinitely many times. Thus, by 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 (DA) is small enough, $X_n = Q(Y_n)$ always remains within the "basin of attraction" of $\mathcal{X}^*$; hence, local convergence can be obtained in the same way as global convergence for a game with smaller action spaces. However, 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. Consequently, any local convergence result in the presence of noise is necessarily probabilistic in nature.

Conditioning on the event that X_n stays close to $\mathcal{X}^*$, local convergence can be obtained as in the proof of Theorem 4.7. Nevertheless, showing that this event occurs with controllably high probability requires a completely different analysis. This is the essence of our next result:

Theorem 4.11 *Fix a confidence level* $\delta > 0$ *and suppose that* (DA) *is run with a sufficiently small step-size* γ_n *satisfying* $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$ *and* $\sum_{n=1}^{\infty} \gamma_n = \infty$*. 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*

$$\mathbb{P}(X_n \to \mathcal{X}^* \,|\, X_1 \in U) \geq 1-\delta. \tag{4.19}$$

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 $X_1 \in U_\varepsilon$ so there exists some $x^* \in \mathcal{X}^*$ such that $F(x^*, Y_1) < \varepsilon$. Then, for all n, Proposition 4.3 yields

$$F(x^*, Y_{n+1}) \leq F(x^*, Y_n) + \gamma_n \langle v(X_n), X_n - x^* \rangle + \gamma_n \psi_{n+1} + \frac{\gamma_n^2}{2K} \|\hat{v}_{n+1}\|_*^2, \tag{4.20}$$

where we have set $\psi_{n+1} = \langle \xi_{n+1}, X_n - x^* \rangle$.

We first claim that $\sup_n \sum_{k=1}^n \gamma_k \psi_{k+1} \leq \varepsilon$ with probability at least $1-\delta/2$ if γ_n is chosen appropriately. Indeed, set $S_{n+1} = \sum_{k=1}^n \gamma_k \psi_{k+1}$ and let $E_{n,\varepsilon}$ denote the event $\{\sup_{1\leq k\leq n+1} |S_k| \geq \varepsilon\}$. Since S_n is a martingale, Doob's maximal inequality ([14], Theorem 2.1) yields

$$\mathbb{P}(E_{n+1,\varepsilon}) \leq \frac{\mathbb{E}[|S_{n+1}|^2]}{\varepsilon^2} \leq \frac{\sigma^2 \|\mathcal{X}\|^2 \sum_{k=1}^n \gamma_k^2}{\varepsilon^2}, \tag{4.21}$$

where we used the variance estimate

$$\begin{aligned} \mathbb{E}[\psi_{k+1}^2] &= \mathbb{E}[\mathbb{E}[|\langle \xi_{k+1}, X_k - x^* \rangle|^2 \,|\, \mathcal{F}_k]] \\ &\leq \mathbb{E}[\mathbb{E}[\|\xi_{k+1}\|_*^2 \|X_k - x^*\|^2 \,|\, \mathcal{F}_k]] \leq \sigma^2 \|\mathcal{X}\|^2, \end{aligned} \tag{4.22}$$

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

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

$$\mathbb{P}(F_{n+1,\varepsilon}) \leq \frac{\mathbb{E}[R_{n+1}]}{\varepsilon} \leq \frac{V_*^2 \sum_{k=1}^n \gamma_k^2}{2K\varepsilon}. \tag{4.23}$$

Consequently, the event $F_\varepsilon = \bigcup_{n=1}^\infty F_{n,\varepsilon}$ occurs with probability $\mathbb{P}(F_\varepsilon) \leq \Gamma_2 V_*^2/\varepsilon \leq \delta/2$ if γ_n is chosen so that $\Gamma_2 \leq K\delta\varepsilon/V_*^2$.

Assume therefore that $\Gamma_2 \leq \min\{\delta\varepsilon^2/(2\sigma^2\|\mathcal{X}\|^2),\ K\delta\varepsilon/V_*^2\}$. 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 ε for all n and all x^* with probability at least $1 - \delta$. Since $F(x^*, Y_1) \leq \varepsilon$ by assumption, we readily get $F(x^*, Y_1) \leq 3\varepsilon$ if $\bar{E}_\varepsilon$ and $\bar{F}_\varepsilon$ both hold. Furthermore, telescoping (4.20) yields

$$F(x^*, Y_{n+1}) \leq F(x^*, Y_1) + \sum_{k=1}^{n} \langle v(X_k), X_k - x^* \rangle + S_{n+1} + R_{n+1} \quad \text{for all } n, \tag{4.24}$$

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 $F(x^*, Y_{n+1}) \leq 3\varepsilon$ if neither E_ε nor F_ε 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 claimed. □

4.5 Convergence in zero-sum concave games

We close this section by examining the asymptotic behavior of (DA) in 2-player, concave-convex zero-sum games. To do so, let $\mathcal{N} = \{A, B\}$ denote the set of players with corresponding payoff functions $u_A = -u_B$ respectively concave in x_A and x_B. Letting $u \equiv u_A = -u_B$, the *value* of the game is defined as

$$u^* = \max_{x_A \in \mathcal{X}_A} \min_{x_B \in \mathcal{X}_B} u(x_A, x_B) = \min_{x_B \in \mathcal{X}_B} \max_{x_A \in \mathcal{X}_A} u(x_A, x_B). \tag{4.25}$$

The solutions of the concave-convex saddle-point problem (4.25) are the Nash equilibria of $\mathcal{G}$ and the players' equilibrium payoffs are u^* and $-u^*$ respectively.

In the "perfect feedback" case ($\sigma = 0$), Nesterov [33] showed that the ergodic average

$$\bar{X}_n = \frac{\sum_{k=1}^{n} \gamma_k X_k}{\sum_{k=1}^{n} \gamma_k} \tag{4.26}$$

of the sequence of play generated by (DA) converges to equilibrium. With imperfect feedback and steep h,[15] Nemirovski et al. [30] further showed that $\bar{X}_n$ converges in expectation to the game's set of Nash equilibria, provided that (H1) and (H2) hold. Our next result provides an almost sure version of this result which is also valid for nonsteep h:

Theorem 4.13 *Let $\mathcal{G}$ be a concave 2-player zero-sum game. If (DA) is run with imperfect feedback satisfying (H1)–(H2) and a step-size γ_n such that $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$*

[15] When h is steep, the mirror descent algorithm examined by Nemirovski et al. [30] is a special case of the dual averaging method of Nesterov [33]. This is no longer the case if h is not steep, so the analysis of Nemirovski et al. [30] does not apply to (DA). In the online learning literature, this difference is sometimes referred to as "greedy" vs. "lazy" mirror descent.

and $\sum_{n=1}^{\infty}\gamma_n = \infty$*, the ergodic average* $\bar{X}_n$ *of* X_n *converges to the set of Nash equilibria of* $\mathcal{G}$ *(a.s.).*

Proof of Theorem 4.13 Consider the gap function

$$\epsilon(x) = u^* - \min_{p_B\in\mathcal{X}_B} u(x_A, p_B) + \max_{p_A\in\mathcal{X}_A} u(p_A, x_B) - u^* = \max_{p\in\mathcal{X}} \sum_{i\in\mathcal{N}} u_i(p_i; x_{-i}). \tag{4.27}$$

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 do so, pick some $p \in \mathcal{X}$. Then, as in the proof of Theorem 4.7, we have

$$F(p, Y_{n+1}) \leq F(p, Y_n) + \gamma_n\langle v(X_n), X_n - p\rangle + \gamma_n\psi_{n+1} + \frac{1}{2K}\gamma_n^2\|\hat{v}_{n+1}\|_*^2. \tag{4.28}$$

Hence, after rearranging and telescoping, we get

$$\sum_{k=1}^{n}\gamma_k\langle v(X_k), p - X_k\rangle \leq F(p, Y_1) + \sum_{k=1}^{n}\gamma_k\psi_{k+1} + \frac{1}{2K}\sum_{k=1}^{n}\gamma_k^2\|\hat{v}_{k+1}\|_*^2, \tag{4.29}$$

where $\psi_{n+1} = \langle\xi_{n+1}, X_n - p\rangle$ and we used the fact that $F(p, Y_n) \geq 0$. By concavity, we also have

$$\langle v(x), p-x\rangle = \sum_{i\in\mathcal{N}}\langle v_i(x), p_i - x_i\rangle \geq \sum_{i\in\mathcal{N}}\big[u_i(p_i; x_{-i}) - u_i(x)\big] = \sum_{i\in\mathcal{N}} u_i(p_i; x_{-i}), \tag{4.30}$$

for all $x \in \mathcal{X}$. Therefore, letting $\tau_n = \sum_{k=1}^{n}\gamma_k$, we get

$$\begin{aligned}\frac{1}{\tau_n}\sum_{k=1}^{n}\gamma_k\langle v(X_k), p - X_k\rangle &\geq \frac{1}{\tau_n}\sum_{k=1}^{n}\gamma_k\sum_{i\in\mathcal{N}} u_i(p_i; X_{-i,k})\\ &\geq u(p_A, \bar{X}_{B,n}) - u(\bar{X}_{A,n}, p_B)\\ &= \sum_{i\in\mathcal{N}} u_i(p_i; \bar{X}_{-i,n}),\end{aligned} \tag{4.31}$$

where we used the fact that u is concave-convex in the second line. Thus, combining (4.29) and (4.31), we finally obtain

$$\sum_{i\in\mathcal{N}} u_i(p_i; \bar{X}_{-i,n}) \leq \frac{F(p, Y_1) + \sum_{k=1}^{n}\gamma_k\psi_{k+1} + (2K)^{-1}\sum_{k=1}^{n}\gamma_k^2\|\hat{v}_{k+1}\|_*^2}{\tau_n}. \tag{4.32}$$

As before, the law of large numbers [14, Theorem 2.18] yields $\tau_n^{-1}\sum_{k=1}^{n}\gamma_k\psi_{k+1} \to 0$ (a.s.). Furthermore, given that $\mathbb{E}[\|\hat{v}_{n+1}\|_*^2 \,|\, \mathcal{F}_n] \leq V_*^2$ and $\sum_{k=1}^{n}\gamma_k^2 < \infty$, we also get $\tau_n^{-1}\sum_{k=1}^{n}\gamma_k^2\|\hat{v}_{k+1}\|_*^2 \to 0$ by Doob's martingale convergence theorem [14, 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 (DA) in *finite* games. Briefly recalling the setup of Example 2.1, each player in a finite game $\Gamma \equiv \Gamma(\mathcal{N}, (\mathcal{A}_i)_{i\in\mathcal{N}}, (u_i)_{i\in\mathcal{N}})$ chooses a pure strategy α_i from a finite set $\mathcal{A}_i$ and receives a payoff of $u_i(\alpha_1, \dots, \alpha_N)$. Pure strategies are drawn based on the players' mixed strategies $x_i \in \mathcal{X}_i \equiv \Delta(\mathcal{A}_i)$, 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 $v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}) = (u_i(\alpha_i; x_{-i}))_{\alpha_i\in\mathcal{A}_i}$ of Eq. 2.4.

Consider now the following learning scheme: At stage n, every player $i \in \mathcal{N}$ selects a pure strategy $\alpha_{i,n} \in \mathcal{A}_i$ according to their individual mixed strategy $X_{i,n} \in \mathcal{X}_i$. Subsequently, each player observes— or otherwise calculates— the payoffs of their pure strategies $\alpha_i \in \mathcal{A}_i$ against the chosen actions $\alpha_{-i,n}$ of all other players (possibly subject to some random estimation error). Specifically, we posit that each player receives as feedback the "noisy" payoff vector

$$\hat{v}_{i,n+1} = (u_i(\alpha_i; \alpha_{-i,n}))_{\alpha_i\in\mathcal{A}_i} + \xi_{i,n+1}, \tag{5.1}$$

where the error process $\xi_n = (\xi_{i,n})_{i\in\mathcal{N}}$ is assumed to satisfy (H1) and (H2). Then, based on this feedback, players update their mixed strategies 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 adaptive 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.

5.1 Dominated strategies

We say that a pure strategy $\alpha_i \in \mathcal{A}_i$ of a finite game Γ is *dominated* by $\beta_i \in \mathcal{A}_i$ (and we write $\alpha_i \prec \beta_i$) if

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

Algorithm 2 Logit-based learning in finite games (Example 3.2).

Require: step-size sequence $\gamma_n \propto 1/n^\beta$, $\beta \in (0, 1]$; initial scores $Y_i \in \mathbb{R}^{\mathcal{A}_i}$

for $n = 1, 2, \dots$ **do**
  for every player $i \in \mathcal{N}$ **do**
    set $X_i \leftarrow \Lambda_i(Y_i)$; {mixed strategy}
    play $\alpha_i \sim X_i$; {choose action}
    observe $\hat{v}_i$; {estimate payoffs}
    update $Y_i \leftarrow Y_i + \gamma_n \hat{v}_i$; {update scores}
  end for
end for

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

Theorem 5.1 *Suppose that* (DA) *is run with noisy payoff observations of the form* (5.1) *and a step-size sequence* γ_n *satisfying* (4.2)*. If* $\alpha_i \in \mathcal{A}_i$ *is dominated, then* $X_{i\alpha_i,n} \to 0$ *(a.s.).*

Proof Suppose that $\alpha_i \prec \beta_i$ for some $\beta_i \in \mathcal{A}_i$. Then, suppressing the player index i for simplicity, (DA) gives

$$\begin{aligned} Y_{\beta,n+1} - Y_{\alpha,n+1} &= c_{\beta\alpha} + \sum_{k=1}^{n} \gamma_k \left[\hat{v}_{\beta,k+1} - \hat{v}_{\alpha,k+1}\right] \\ &= c_{\beta\alpha} + \sum_{k=1}^{n} \gamma_k \left[v_\beta(X_k) - v_\alpha(X_k)\right] + \sum_{k=1}^{n} \gamma_k \zeta_{k+1}, \end{aligned} \tag{5.3}$$

where we set $c_{\beta\alpha} = Y_{\beta,1} - Y_{\alpha,1}$ and

$$\zeta_{k+1} = \mathbb{E}[\hat{v}_{\beta,k+1} - \hat{v}_{\alpha,k+1} \,|\, \mathcal{F}_k] - [v_\beta(X_k) - v_\alpha(X_k)]. \tag{5.4}$$

Since $\alpha \prec \beta$, there exists some $c > 0$ such that $v_\beta(x) - v_\alpha(x) \geq c$ for all $x \in \mathcal{X}$. Then, (5.3) yields

$$Y_{\beta,n+1} - Y_{\alpha,n+1} \geq c_{\beta\alpha} + \tau_n \left[c + \frac{\sum_{k=1}^{n} \gamma_k \zeta_{k+1}}{\tau_n}\right], \tag{5.5}$$

where $\tau_n = \sum_{k=1}^{n} \gamma_k$. As in the proof of Theorem 4.1, the law of large numbers for martingale difference [14, Theorem 2.18] implies that $\tau_n^{-1} \sum_{k=1}^{n} \gamma_k \zeta_{k+1} \to 0$ under the step-size assumption (4.2), so $Y_{\beta,n} - Y_{\alpha,n} \to \infty$ (a.s.).

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

$$h(X'_n) \geq h(X_n) + \langle Y_n, X'_n - X_n \rangle = h(X_n) + \varepsilon(Y_{\beta,n} - Y_{\alpha,n}) \to \infty, \tag{5.6}$$

a contradiction. This implies that $X_{\alpha,n} \to 0$ (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 $x^* = (\alpha_1^*, \dots, \alpha_N^*)$ such that

$$u_i(\alpha_i^*; \alpha_{-i}^*) > u_i(\alpha_i; \alpha_{-i}^*) \quad \text{for all } \alpha_i \in \mathcal{A}_i \setminus \{\alpha_i^*\}, i \in \mathcal{N}. \tag{5.7}$$

Strict Nash equilibria can be characterized further as follows:

Proposition 5.2 *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 1), the convergence analysis of Sect. 4 yields:

Proposition 5.3 *Let* x^* *be a strict equilibrium of a finite game* Γ. *Suppose further that* (DA) *is run with noisy payoff observations of the form* (5.1) *and a sufficiently small step-size* γ_n *such that* $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$ *and* $\sum_{n=1}^{\infty} \gamma_n = \infty$. *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*

$$\mathbb{P}(X_n \to x^* \mid X_1 \in U) \geq 1 - \delta. \tag{5.8}$$

Proof We first show that $\mathbb{E}[\hat{v}_{n+1} \mid \mathcal{F}_n] = v(X_n)$. Indeed, for all $i \in \mathcal{N}$, $\alpha_i \in \mathcal{A}_i$, we have

$$\mathbb{E}[\hat{v}_{i\alpha_i,n+1} \mid \mathcal{F}_n] = \sum_{\alpha_{-i} \in \mathcal{A}_{-i}} u_i(\alpha_i; \alpha_{-i})\, X_{\alpha_{-i},n} + \mathbb{E}[\xi_{i\alpha_i,n+1} \mid \mathcal{F}_n] = u_i(\alpha_i; X_{-i,n}), \tag{5.9}$$

where, in a slight abuse of notation, we set $X_{\alpha_{-i},n}$ for the joint probability assigned to the pure strategy profile α_{-i} of all players other than i at stage n.

By (2.4), it follows that $\mathbb{E}[\hat{v}_{n+1} \mid \mathcal{F}_n] = v(X_n)$ so the estimator (5.1) is unbiased in the sense of (H1). Hypothesis (H2) can be verified similarly, so the estimator (5.1) satisfies (3.3). Since x^* is stable by Proposition 5.2 and $v(x)$ is multilinear (so (H4) is satisfied automatically), our assertion follows from Theorem 4.11. □

In the special case of logit-based learning (Example 3.2), Cohen et al. [10] 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 (DA) in finite games, showing that the logit choice map (3.9) has no special properties in this regard. Cohen et al. [10] further showed that the convergence rate of logit-based learning is exponential in the algorithm's "running horizon" $\tau_n = \sum_{k=1}^{n} \gamma_k$. 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 Sect. 6.

5.3 Convergence in zero-sum games

We close this section with a brief discussion of the ergodic convergence properties of (DA) in finite two-player zero-sum games. In this case, the analysis of Sect. 4.5 readily yields:

Corollary 5.4 *Let Γ be a finite 2-player zero-sum game. If (DA) is run with noisy payoff observations of the form (5.1) and a step-size γ_n such that $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$ and $\sum_{n=1}^{\infty} \gamma_n = \infty$, the ergodic average $\bar{X}_n = \sum_{k=1}^{n} \gamma_k X_k \big/ \sum_{k=1}^{n} \gamma_k$ of the players' mixed strategies converges to the set of Nash equilibria of Γ (a.s.).*

Proof As in the proof of Proposition 5.3, the estimator (5.1) satisfies $\mathbb{E}[\hat{v}_{n+1} \mid \mathcal{F}_n] = v(X_n)$, so (H1) and (H2) also hold in the sense of (3.3). Our claim then follows from Theorem 4.13. □

Remark 5.1 In a very recent paper, Bravo and Mertikopoulos [7] showed that the time average $\bar{X}(t) = t^{-1} \int_0^t X(s)\, ds$ of the players' mixed strategies under (DA-c) with Brownian payoff shocks converges to Nash equilibrium in 2-player, zero-sum games. Corollary 5.4 may be seen as a discrete-time version of this result.

6 Speed of convergence

6.1 Ergodic convergence rate

In this section, we focus on the rate of convergence of (DA) to stable equilibrium states (and/or sets thereof). To that end, we will measure 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 2.6, $\epsilon(x) \geq 0$ with equality if and only if $x \in \mathcal{X}^*$, so $\epsilon(x)$ can be seen as a (game-dependent) 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

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

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 grow quadratically around their minimum points. With this in mind, we provide

below an explicit estimate for the decay rate of the average equilibrium gap $\bar{\epsilon}_n = \sum_{k=1}^n \gamma_k \epsilon(X_k) \big/ \sum_{k=1}^n \gamma_k$ in the spirit of Nemirovski et al. [30]:

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

$$\mathbb{E}[\bar{\epsilon}_n] \leq \frac{F_1 + V_*^2/(2K) \sum_{k=1}^n \gamma_k^2}{\sum_{k=1}^n \gamma_k}, \tag{6.4}$$

where $F_1 = F(\mathcal{X}^*, Y_1)$. *If, in addition,* $\sum_{n=1}^\infty \gamma_n^2 < \infty$, *we have*

$$\bar{\epsilon}_n \leq \frac{A}{\sum_{k=1}^n \gamma_k} \quad \text{for all } n \text{(a.s.)}, \tag{6.5}$$

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

$$A \leq F_1 + \sigma \|\mathcal{X}\| \kappa + \kappa^2 V_*^2, \tag{6.6}$$

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

Corollary 6.3 *Suppose that* (DA) *is initialized at* $Y_1 = 0$ *and is 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, the long-run average distance to equilibrium* $\bar{r}_n = \sum_{k=1}^n \operatorname{dist}(\mathcal{X}^*, X_n) \big/ \sum_{k=1}^n \gamma_k$ *satisfies*

$$\mathbb{E}[\bar{r}_n] \leq \sqrt[4]{4L^{-2} V_*^2 \Omega/(Kn)}. \tag{6.8}$$

Proof of Theorem 6.2 Let $x^* \in \mathcal{X}^*$. Rearranging (4.20) and telescoping yields

$$\sum_{k=1}^n \gamma_k \langle v(X_k), x^* - X_k \rangle \leq F(x^*, Y_1) + \sum_{k=1}^n \gamma_k \psi_{k+1} + \frac{1}{2K} \sum_{k=1}^n \gamma_k^2 \|\hat{v}_{k+1}\|_*^2, \tag{6.9}$$

where $\psi_{k+1} = \langle \xi_{k+1}, X_k - x^* \rangle$. Thus, taking expectations on both sides, we obtain

$$\sum_{k=1}^n \gamma_k \, \mathbb{E}[\langle v(X_k), x^* - X_k \rangle] \leq F(x^*, Y_1) + \frac{V_*^2}{2K} \sum_{k=1}^n \gamma_k^2. \tag{6.10}$$

Subsequently, minimizing both sides of (6.10) over $x^* \in \mathcal{X}^*$ yields

$$\sum_{k=1}^n \gamma_k \, \mathbb{E}[\epsilon(X_k)] \leq F_1 + \frac{V_*^2}{2K} \sum_{k=1}^n \gamma_k^2, \tag{6.11}$$

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

To establish the almost sure bound (6.5), set $S_{n+1} = \sum_{k=1}^{n} \gamma_k \psi_{k+1}$ and $R_{n+1} = (2K)^{-1} \sum_{k=1}^{n} \gamma_k^2 \|\hat{v}_{k+1}\|_*^2$. Then, (6.9) becomes

$$\sum_{k=1}^{n} \gamma_k \langle v(X_k), x^* - X_k \rangle \leq F(x^*, Y_1) + S_n + R_n, \tag{6.12}$$

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. By Doob's (sub)martingale convergence theorem [14, Theorem 2.5], it also follows that S_n and R_n both converge to an (a.s.) finite limit S_∞ and R_∞ respectively. Consequently, by (6.12), there exists a finite (a.s.) random variable $A > 0$ such that

$$\sum_{k=1}^{n} \gamma_k \langle v(X_k), x^* - X_k \rangle \leq A \quad \text{for all } n \text{ (a.s.)}. \tag{6.13}$$

The bound (6.5) follows by taking the minimum of (6.13) over $x^* \in \mathcal{X}^*$ and dividing both sides by $\sum_{k=1}^{n} \gamma_k$. Finally, applying Doob's maximal inequality to (4.21) and (4.23), we obtain $\mathbb{P}\big(\sup_n S_n \geq \sigma \|\mathcal{X}\| \kappa\big) \leq \delta/2$ and $\mathbb{P}\big(\sup_n R_n \geq V_*^2 \kappa^2\big) \leq \delta/2$. Combining these bounds with (6.12) 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 (4.11) of the setwise Fenchel coupling, we have $F_1 \leq h(x^*) + h^*(0) \leq \max h - \min h = \Omega$. Our claim then follows by invoking Jensen's inequality, 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 $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$ for the almost sure bound (6.5) rules out more aggressive step-size policies of the form $\gamma_n \propto 1/n^\beta$ 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 $\sum_{n=1}^{\infty} \gamma_n^{1+q/2} < \infty$. 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 (DA) is the exact lower complexity bound for black-box subgradient schemes for convex problems ([29, 31]). Thus, running (DA) 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. It is also worth noting that, when the horizon of play is known in advance (as in Corollary 6.3), the constant $\Omega = \max h - \min h$ that results from the initialization $Y_1 = 0$ is essentially the same as the constant that appears in the stochastic mirror descent analysis of Nemirovski et al. [30] and Nesterov [33].

6.2 Running length

Intuitively, the main obstacle to achieving rapid convergence is that, even with an optimized step-size policy, the sequence of play may end up oscillating around an equilibrium state because of the noise in the players' observations. To study such phenomena, we focus below on the *running length* of (DA), defined as

$$\ell_n = \sum_{k=1}^{n-1} \|X_{k+1} - X_k\|. \tag{6.14}$$

Obviously, if X_n converges to some $x^* \in \mathcal{X}$, a shorter length signifies less oscillations of X_n around x^*. Thus, in a certain way, ℓ_n is a more refined convergence criterion than the induced equilibrium gap $\epsilon(X_n)$.

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

Theorem 6.4 *Suppose that* (DA) *is run with imperfect feedback satisfying* (H1)–(H2) *and a step-size* γ_n *such that* $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$ *and* $\sum_{n=1}^{\infty} \gamma_n = \infty$. *Also, given a closed subset* $\mathcal{X}^*$ *of* $\mathcal{X}$, *consider the stopping time* $n_\varepsilon = \inf\{n \geq 0 : \operatorname{dist}(\mathcal{X}^*, X_n) \leq \varepsilon\}$ *and let* $\ell_\varepsilon \equiv \ell_{n_\varepsilon}$ *denote the running length of* (DA) *until* X_n *reaches an* ε*-neighborhood of* $\mathcal{X}^*$. *If* $\mathcal{X}^*$ *is* L*-strongly stable, we have*

$$\mathbb{E}[\ell_\varepsilon] \leq \frac{V_*}{KL} \frac{F_1 + (2K)^{-1} V_*^2 \sum_{k=1}^{\infty} \gamma_k^2}{\varepsilon^2}. \tag{6.15}$$

Proof For all $x^* \in \mathcal{X}^*$ and all $n \in \mathbb{N}$, (4.20) yields

$$\begin{aligned} F(x^*, Y_{n_\varepsilon \wedge n+1}) &\leq F(x^*, Y_1) - \sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k \langle v(X_k), X_k - x^* \rangle \\ &\quad + \sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k \psi_{k+1} + \frac{1}{2K} \sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k^2 \|\hat{v}_{k+1}\|_*^2. \end{aligned} \tag{6.16}$$

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

$$0 \leq F_1 - L\varepsilon^2 \, \mathbb{E}\left[\sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k\right] + \mathbb{E}\left[\sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k \psi_{k+1}\right] + \frac{V_*^2}{2K} \sum_{k=1}^{\infty} \gamma_k^2, \tag{6.17}$$

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

Consider now the stopped process $S_{n_\varepsilon \wedge n} = \sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k \psi_{k+1}$. 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 (6.17), we obtain

$$\mathbb{E}\left[\sum_{k=1}^{n_\varepsilon \wedge n} \gamma_k\right] \leq \frac{F_1 + (2K)^{-1} V_*^2 \sum_{k=1}^{\infty} \gamma_k^2}{L\varepsilon^2}. \tag{6.18}$$

Hence, with $n_\varepsilon \wedge n \to n_\varepsilon$ as $n \to \infty$, Lebesgue's monotone convergence theorem shows that the process $\tau_\varepsilon = \sum_{k=1}^{n_\varepsilon} \gamma_k$ is finite in expectation and

$$\mathbb{E}[\tau_\varepsilon] \leq \frac{F_1 + (2K)^{-1} V_*^2 \sum_{k=1}^{\infty} \gamma_k^2}{L\varepsilon^2}. \tag{6.19}$$

Furthermore, by Proposition 3.2 and the definition of ℓ_n, we also have

$$\ell_n = \sum_{k=1}^{n-1} \|X_{k+1} - X_k\| \leq \frac{1}{K} \sum_{k=1}^{n-1} \|Y_k - Y_{k-1}\|_* = \frac{1}{K} \sum_{k=1}^{n-1} \gamma_k \|\hat{v}_{k+1}\|_*. \tag{6.20}$$

Now, let $\zeta_{k+1} = \|\hat{v}_{k+1}\|_*$ and $\Psi_{n+1} = \sum_{k=1}^{n} \gamma_k \left[\zeta_{k+1} - \mathbb{E}[\zeta_{k+1} \mid \mathcal{F}_k]\right]$. By construction, Ψ_n is a martingale and

$$\mathbb{E}[\Psi_{n+1}^2] = \mathbb{E}\left[\sum_{k=1}^{n} \gamma_k^2 \left[\zeta_{k+1} - \mathbb{E}[\zeta_{k+1} \mid \mathcal{F}_k]\right]^2\right] \leq 2V_*^2 \sum_{k=1}^{\infty} \gamma_k^2 < \infty \quad \text{for all } n. \tag{6.21}$$

Thus, by the optional stopping theorem [46, p. 485], we get $\mathbb{E}[\Psi_{n_\varepsilon}] = \mathbb{E}[\Psi_1] = 0$, so

$$\mathbb{E}\left[\sum_{k=1}^{n_\varepsilon} \gamma_k \zeta_{k+1}\right] = \mathbb{E}\left[\sum_{k=1}^{n_\varepsilon} \gamma_k \, \mathbb{E}[\zeta_{k+1} \mid \mathcal{F}_k]\right] \leq V_* \, \mathbb{E}\left[\sum_{k=1}^{n_\varepsilon} \gamma_k\right] = V_* \, \mathbb{E}[\tau_\varepsilon]. \tag{6.22}$$

Our claim then follows by combining (6.20) and (6.22) with the bound (6.19). □

Theorem 6.4 should be contrasted to classic results on the Kurdyka–Łojasiewicz inequality where having a "bounded length" property is crucial in establishing trajectory convergence [6]. In our stochastic setting, it is not realistic to expect a bounded length (even on average), because, generically, the noise does not vanish in the neighborhood of a Nash equilibrium.[16] Instead, Theorem 6.4 should be interpreted as a measure of how the fluctuations due to noise and uncertainty affect the trajectories' average length; the authors are not aware of any similar results along these lines.

6.3 Sharp equilibria and fast convergence

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 (DA). Specifically, even with a rapidly decreasing step-size policy, a single realization of the error process ξ_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. [10] recently showed that logit-based learning (cf. Algorithm 2) achieves a quasi-linear convergence rate with high

[16] For a notable exception however, see Theorem 6.6 below.

probability if the equilibrium in question is strict. Specifically, Cohen et al. [10] showed that if x^* is a strict Nash equilibrium and X_n does not start too far from x^*, then, with high probability, $\|X_n - x^*\| = \mathcal{O}(-c\sum_{k=1}^{n}\gamma_k)$ for some positive constant $c > 0$ that depends only on the players' relative payoff differences.

Building on the variational characterization of strict Nash equilibria provided by Proposition 5.2, we consider below the following analogue for continuous games:

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

$$\langle v(x^*), z\rangle \leq 0 \quad \text{for all } z \in \mathrm{TC}(x^*), \tag{6.23}$$

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

Remark 6.1 The terminology "sharp" follows Polyak [37, Chapter 5.2], who introduced a similar notion for (unconstrained) convex programs. In particular, in the single-player case, it is easy to see that (6.23) implies that x^* is a *sharp maximum* of $u(x)$, i.e. $u(x^*) - u(x) \geq c\|x - x^*\|$ for some $c > 0$.

A first consequence of Definition 6.5 is that $v(x^*)$ lies in the topological interior of the polar cone $\mathrm{PC}(x^*)$ to $\mathcal{X}$ at x^* (for a schematic illustration, see Fig. 1); in turn, this implies that sharp equilibria can only occur at *corners* of $\mathcal{X}$. By continuity, this further implies that sharp equilibria are locally stable (cf. the proof of Theorem 6.6 below); hence, by Proposition 2.7, sharp equilibria are also isolated. Our next result shows that if players employ (DA) with surjective choice maps, then, with high probability, sharp equilibria are attained in a *finite* number of steps:

Theorem 6.6 *Fix a tolerance level $\delta > 0$ and suppose that* (DA) *is run with surjective choice maps and a sufficiently small step-size γ_n such that $\sum_{n=1}^{\infty}\gamma_n^2 < \infty$ and $\sum_{n=1}^{\infty}\gamma_n = \infty$. If x^* is sharp and* (DA) *is not initialized too far from x^*, we have*

$$\mathbb{P}(X_n \text{ reaches } x^* \text{ in a } \textit{finite} \text{ number of steps}) \geq 1 - \delta, \tag{6.24}$$

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

Proof As we noted above, $v(x^*)$ lies in the interior of the polar cone $\mathrm{PC}(x^*)$ to $\mathcal{X}$ at x^*.[17] 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. Therefore, by Theorem 4.11, there exists a neighborhood U of x^* such that 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 $\langle v(x), z\rangle \leq -c\|z\|$ for some $c > 0$ and for all $z \in \mathrm{TC}(x^*)$.[18] 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 $\langle v(X_n), z\rangle \leq -c\|z\|$ for all $n \geq n_0$. Thus, for all $z \in \mathrm{TC}(x^*)$ with $\|z\| = 1$, we have

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

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

$$\langle Y_{n+1}, z\rangle = \langle Y_{n_0}, z\rangle + \sum_{k=n_0}^{n} \gamma_k \langle v(X_k), z\rangle + \sum_{k=n_0}^{n} \gamma_k \langle \xi_{k+1}, z\rangle$$
$$\leq \|Y_{n_0}\|_* - c \sum_{k=n_0}^{n} \gamma_k + \sum_{k=n_0}^{n} \gamma_k \langle \xi_{k+1}, z\rangle. \tag{6.25}$$

By the law of large numbers for martingale difference [14, Theorem 2.18], we also have $\sum_{k=n_0}^{n} \gamma_k \xi_{k+1} / \sum_{k=n_0}^{n} \gamma_k \to 0$ (a.s.), so there exists some n^* such that $\|\sum_{k=n_0}^{n} \gamma_k \xi_{k+1}\|_* \leq (c/2) \sum_{k=n_0}^{n} \gamma_k$ for all $n \geq n^*$ (a.s.). We thus obtain

$$\langle Y_{n+1}, z\rangle \leq \|Y_{n_0}\|_* - c \sum_{k=n_0}^{n} \gamma_k + \frac{c}{2}\|z\| \sum_{k=n_0}^{n} \gamma_k \leq \|Y_{n_0}\|_* - \frac{c}{2} \sum_{k=n_0}^{n} \gamma_k, \tag{6.26}$$

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 1 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 $y^* + \mathrm{PC}(x^*)$ 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 globally stable case, recall that X_n converges to x^* with probability 1 from any initial condition (Theorem 4.7). The argument above shows that $X_n = x^*$ for all large n, so X_n converges to x^* in a finite number of steps (a.s.). □

Remark 6.2 Theorem 6.6 suggests that dual averaging with surjective choice maps leads to significantly faster convergence to sharp equilibria. In this way, it is consistent with an observation made by Mertikopoulos and Sandholm [26, Proposition 5.2] for the convergence of the continuous-time, deterministic dynamics (DA-c) in finite games.

7 Discussion

An important question in the implementation of dual averaging is the choice of regularizer, which in turn determines the players' choice maps $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$. From a qualitative point of view, this choice would not seem to matter much: the convergence results of Sects. 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 sharp 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 Ω/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 [44, p. 140].

Another key question in game-theoretic and online learning has to do with the information that is available to the players at each stage. If players perform a two-point sampling step in order to simulate an extra oracle call at an action profile different than the one employed, this extra information could be presumably leveraged in order to increase the speed of convergence to a Nash equilibrium. In an offline setting, this can be achieved by more sophisticated techniques relying on dual extrapolation [32] and/or mirror-prox methods [19]. Extending these extra-gradient approaches to online learning processes as above would be an interesting extension of the current work.

At the other end of the spectrum, if players only have access to their realized, in-game payoffs, they would need to reconstruct their individual payoff gradients via a suitable single-shot estimator [13,37]. We believe our convergence analysis can be extended to this case by properly controlling the "bias-variance" tradeoff of this estimator and using more refined stochastic approximation arguments. The very recent manuscript by Bervoets et al. [5] provides an encouraging first step in the case of (strictly) concave games with one-dimensional action sets; we intend to explore this direction in future work.

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

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

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

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

leading to the bound

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

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 [39, 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

$$h(x) \geq h(x^*) + \langle y^*, x - x^*\rangle \geq h(x^*) + \langle y^* + v, x - x^*\rangle. \tag{7.5}$$

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 (DA-c):

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

$$\frac{d}{dt}F(p, y(t)) = \langle v(x(t)), x(t) - p\rangle. \tag{7.6}$$

Proof By definition, we have

$$\begin{aligned} \frac{d}{dt}F(p, y(t)) &= \frac{d}{dt}\big[h(p) + h^*(y(t)) - \langle y(t), p\rangle\big] \\ &= \langle \dot{y}(t), \nabla h^*(y(t))\rangle - \langle \dot{y}(t), p\rangle = \langle v(x(t)), x(t) - p\rangle, \end{aligned} \tag{7.7}$$

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

Our last auxiliary result shows that, if the sequence of play generated by (DA) 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 (DA) is run with a step-size such that $\sum_{n=1}^{\infty}\gamma_n^2 < \infty$ and $\sum_{n=1}^{\infty}\gamma_n = \infty$. Assume further that $(X_n)_{n=1}^{\infty}$ 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 (DA) is run with perfect feedback ($\sigma = 0$), the above holds under the lighter assumption $\sum_{k=1}^{n}\gamma_k^2 \big/ \sum_{k=1}^{n}\gamma_k \to 0$.*

Proof of Lemma A.3 Let U be a neighborhood of $\mathcal{X}^*$ and assume to the contrary that, with positive probability, $X_n \notin U$ for all sufficiently large n. By starting the sequence at a later index if necessary, we may 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 $c > 0$ such that

$$\langle v(X_n), X_n - x^* \rangle \leq -c \quad \text{for all } x^* \in \mathcal{X}^* \text{ and for all } n. \tag{7.8}$$

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

$$\begin{aligned} F(x^*, Y_{n+1}) &= F(x^*, Y_n + \gamma_n \hat{v}_{n+1}) \\ &\leq F(x^*, Y_n) + \gamma_n \langle v(X_n) + \xi_{n+1}, X_n - x^* \rangle + \frac{1}{2K} \gamma_n^2 \|\hat{v}_{n+1}\|_*^2 \\ &\leq F(x^*, Y_n) - c\gamma_n + \gamma_n \psi_{n+1} + \frac{1}{2K} \gamma_n^2 \|\hat{v}_{n+1}\|_*^2, \end{aligned} \tag{7.9}$$

where we used Proposition 4.3 in the second line and we set $\psi_{n+1} = \langle \xi_{n+1}, X_n - x^* \rangle$ in the third. Telescoping (7.9) then gives

$$F(x^*, Y_{n+1}) \leq F(x^*, Y_1) - \tau_n \left[c - \frac{\sum_{k=1}^n \gamma_k \psi_{k+1}}{\tau_n} - \frac{1}{2K} \frac{\sum_{k=1}^n \gamma_k^2 \|\hat{v}_{k+1}\|_*^2}{\tau_n} \right], \tag{7.10}$$

where $\tau_n = \sum_{k=1}^n \gamma_k$.

Since $\mathbb{E}[\psi_{n+1} \mid \mathcal{F}_n] = \langle \mathbb{E}[\xi_{n+1} \mid \mathcal{F}_n], X_n - x^* \rangle = 0$ by (H1) and $\mathbb{E}[|\psi_{n+1}|^2 \mid \mathcal{F}_n] \leq \mathbb{E}[\|\xi_{n+1}\|_*^2 \|X_n - x^*\|^2 \mid \mathcal{F}_n] \leq \sigma^2 \|\mathcal{X}\|^2 < \infty$ by (H2), the law of large numbers for martingale difference yields $\tau_n^{-1} \sum_{k=1}^n \gamma_k \psi_{k+1} \to 0$ [14, Theorem 2.18]. Furthermore, letting $R_{n+1} = \sum_{k=1}^n \gamma_k^2 \|\hat{v}_{k+1}\|_*^2$, we also get

$$\mathbb{E}[R_{n+1}] \leq \sum_{k=1}^n \gamma_k^2 \, \mathbb{E}[\hat{v}_{k+1}]^2 \leq V_*^2 \sum_{k=1}^\infty \gamma_k^2 < \infty \quad \text{for all } n, \tag{7.11}$$

so Doob's martingale convergence theorem shows that R_n converges (a.s.) to some random, finite value [14, Theorem 2.5].

Combining the above, (7.10) gives $F(x^*, Y_n) \sim -a\tau_n \to -\infty$ (a.s.), a contradiction. Finally, if $\sigma = 0$, we also have $\psi_{n+1} = 0$ and $\|\hat{v}_{n+1}\|_*^2 = \|v(X_n)\|_*^2 \leq V_*^2$ for all n, so (7.10) yields $F(x^*, Y_n) \to -\infty$ provided that $\tau_n^{-1} \sum_{k=1}^n \gamma_k^2 \to 0$, a contradiction. In both cases, we conclude that X_n is recurrent, as claimed. □

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

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

$(a) \Longrightarrow (b)$ Suppose that $x^* = (\alpha_1^*, \dots, \alpha_N^*)$ 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

$$\langle v_i(x^*), z_i\rangle = \sum_{\alpha_i \neq \alpha_i^*} z_{i,\alpha_i}\left[u_i(\alpha_i^*; \alpha_{-i}^*) - u_i(\alpha_i; \alpha_{-i}^*)\right] < 0, \tag{7.12}$$

where we used the fact that z_i is tangent to $\mathcal{X}$ at x_i^*, so $\sum_{\alpha_i \in \mathcal{A}_i} z_{i\alpha_i} = 0$ and $z_{i\alpha_i} \geq 0$ for $\alpha_i \neq \alpha_i^*$, with at least one of these inequalities being strict when $z_i \neq 0$.
$(b) \Longrightarrow (c)$ 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.
$(c) \Longrightarrow (a)$ Assume that x^* is stable but not strict, so $u_{i\alpha_i}(x^*) = u_{i\beta_i}(x^*)$ for some $i \in \mathcal{N}$, and some $\alpha_i \in \mathrm{supp}(x_i^*)$, $\beta_i \in \mathcal{A}_i$. Then, if we take $x_i = x_i^* + \lambda(e_{i\beta_i} - e_{i\alpha_i})$ and $x_{-i} = x_{-i}^*$ with $\lambda > 0$ small enough, we get

$$\langle v(x), x - x^*\rangle = \langle v_i(x), x_i - x_i^*\rangle = \lambda u_{i\beta_i}(x^*) - \lambda u_{i\alpha_i}(x^*) = 0, \tag{7.13}$$

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

References

1. Alvarez, F., Bolte, J., Brahic, O.: Hessian Riemannian gradient flows in convex programming. SIAM J. Control Optim. **43**(2), 477–501 (2004)
2. Arora, S., Hazan, E., Kale, S.: The multiplicative weights update method: a meta-algorithm and applications. Theory Comput. **8**(1), 121–164 (2012)
3. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. Oper. Res. Lett. **31**(3), 167–175 (2003)
4. Benaïm, M.: Dynamics of stochastic approximation algorithms. In: Azéma, J., Émery, M., Ledoux, M., Yor, M. (eds.) Séminaire de Probabilités XXXIII, Lecture Notes in Mathematics, vol. 1709, pp. 1–68. Springer, Berlin (1999)
5. Bervoets, S., Bravo, M., Faure, M.: Learning and convergence to Nash in network games with continuous action set. Working paper (2016)
6. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. **362**(6), 3319–3363 (2010)
7. Bravo, M., Mertikopoulos, P.: On the robustness of learning in games with stochastically perturbed payoff observations. Games Econ. Behav. **103**(John Nash Memorial issue), 41–66 (2017)
8. Bubeck, S., Cesa-Bianchi, N.: Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Found. Trends Mach. Learn. **5**(1), 1–122 (2012)
9. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. **3**(3), 538–543 (1993)
10. Cohen, J., Héliou, A., Mertikopoulos, P.: Hedging under uncertainty: regret minimization meets exponentially fast convergence. In: SAGT '17: Proceedings of the 10th International Symposium on Algorithmic Game Theory (2017)
11. Coucheney, P., Gaujal, B., Mertikopoulos, P.: Penalty-regulated dynamics and robust learning procedures in games. Math. Oper. Res. **40**(3), 611–633 (2015)
12. Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. 4OR **5**(3), 173–210 (2007)
13. Flaxman, A.D., Kalai, A.T., McMahan, H.B.: Online convex optimization in the bandit setting: gradient descent without a gradient. In: SODA '05: Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 385–394 (2005)
14. Hall, P., Heyde, C.C.: Martingale Limit Theory and Its Application. Probability and Mathematical Statistics. Academic Press, New York (1980)
15. Hazan, E.: A survey: the convex optimization approach to regret minimization. In: Sra, S., Nowozin, S., Wright, S.J. (eds.) Optimization for Machine Learning, pp. 287–304. MIT Press, Cambridge (2012)

16. Hofbauer, J., Sandholm, W.H.: On the global convergence of stochastic fictitious play. Econometrica **70**(6), 2265–2294 (2002)
17. Hofbauer, J., Sandholm, W.H.: Stable games and their dynamics. J. Econ. Theory **144**(4), 1665–1693 (2009)
18. Hofbauer, J., Schuster, P., Sigmund, K.: A note on evolutionarily stable strategies and game dynamics. J. Theor. Biol. **81**(3), 609–612 (1979)
19. Juditsky, A., Nemirovski, A.S., Tauvel, C.: Solving variational inequalities with stochastic mirror-prox algorithm. Stoch. Syst. **1**(1), 17–58 (2011)
20. Kiwiel, K.C.: Free-steering relaxation methods for problems with strictly convex costs and linear constraints. Math. Oper. Res. **22**(2), 326–349 (1997)
21. Laraki, R., Mertikopoulos, P.: Higher order game dynamics. J. Econ. Theory **148**(6), 2666–2695 (2013)
22. Leslie, D.S., Collins, E.J.: Individual Q-learning in normal form games. SIAM J. Control Optim. **44**(2), 495–514 (2005)
23. Littlestone, N., Warmuth, M.K.: The weighted majority algorithm. Inf. Comput. **108**(2), 212–261 (1994)
24. Maynard Smith, J., Price, G.R.: The logic of animal conflict. Nature **246**, 15–18 (1973)
25. McKelvey, R.D., Palfrey, T.R.: Quantal response equilibria for normal form games. Games Econ. Behav. **10**(6), 6–38 (1995)
26. Mertikopoulos, P., Sandholm, W.H.: Learning in games via reinforcement and regularization. Math. Oper. Res. **41**(4), 1297–1324 (2016)
27. Mertikopoulos, P., Papadimitriou, C.H., Piliouras, G.: Cycles in adversarial regularized learning. In: SODA '18: Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (2018)
28. Monderer, D., Shapley, L.S.: Potential games. Games Econ. Behav. **14**(1), 124–143 (1996)
29. Nemirovski, A.S., Yudin, D.B.: Problem Complexity and Method Efficiency in Optimization. Wiley, New York (1983)
30. Nemirovski, A.S., Juditsky, A., Lan, G.G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. SIAM J. Optim. **19**(4), 1574–1609 (2009)
31. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization. Kluwer Academic Publishers, Dordrecht (2004)
32. Nesterov, Y.: Dual extrapolation and its applications to solving variational inequalities and related problems. Math. Program. **109**(2), 319–344 (2007)
33. Nesterov, Y.: Primal-dual subgradient methods for convex problems. Math. Program. **120**(1), 221–259 (2009)
34. Neyman, A.: Correlated equilibrium and potential games. Int. J. Game Theory **26**(2), 223–227 (1997)
35. Perkins, S., Leslie, D.S.: Asynchronous stochastic approximation with differential inclusions. Stoch. Syst. **2**(2), 409–446 (2012)
36. Perkins, S., Mertikopoulos, P., Leslie, D.S.: Mixed-strategy learning with continuous action sets. IEEE Trans. Autom. Control **62**(1), 379–384 (2017)
37. Polyak, B.T.: Introduction to Optimization. Optimization Software, New York (1987)
38. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
39. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. A Series of Comprehensive Studies in Mathematics, vol. 317. Springer, Berlin (1998)
40. Rosen, J.B.: Existence and uniqueness of equilibrium points for concave N-person games. Econometrica **33**(3), 520–534 (1965)
41. Sandholm, W.H.: Population games and deterministic evolutionary dynamics. In: Young, H.P., Zamir, S. (eds.) Handbook of Game Theory IV, pp. 703–778. Elsevier, Amsterdam (2015)
42. Scutari, G., Facchinei, F., Palomar, D.P., Pang, J.S.: Convex optimization, game theory, and variational inequality theory in multiuser communication systems. IEEE Signal Process. Mag. **27**(3), 35–49 (2010)
43. Shalev-Shwartz, S.: Online learning: theory, algorithms, and applications. Ph.D. thesis, Hebrew University of Jerusalem (2007)
44. Shalev-Shwartz, S.: Online learning and online convex optimization. Found. Trends Mach. Learn. **4**(2), 107–194 (2011)
45. Shalev-Shwartz, S., Singer, Y.: Convex repeated games and Fenchel duality. In: Advances in Neural Information Processing Systems, vol. 19, pp. 1265–1272. MIT Press, Cambridge (2007)
46. Shiryaev, A.N.: Probability, 2nd edn. Springer, Berlin (1995)
47. Sorin, S., Wan, C.: Finite composite games: equilibria and dynamics. J. Dyn. Games **3**(1), 101–120 (2016)

48. Viossat, Y., Zapechelnyuk, A.: No-regret dynamics and fictitious play. J. Econ. Theory **148**(2), 825–842 (2013)
49. Vovk, V.G.: Aggregating strategies. In: COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory, pp. 371–383 (1990)
50. Xiao, L.: Dual averaging methods for regularized stochastic learning and online optimization. J. Mach. Learn. Res. **11**, 2543–2596 (2010)
51. Zinkevich, M.: Online convex programming and generalized infinitesimal gradient ascent. In: ICML '03: Proceedings of the 20th International Conference on Machine Learning, pp. 928–936 (2003)