Convergence to Nash equilibrium in continuous games with noisy first-order feedback

Panayotis Mertikopoulos and Mathias Staudigl

***Abstract*—This paper examines the convergence of a broad class of distributed learning dynamics for games with continuous action sets. The dynamics under study comprise a multi-agent generalization of Nesterov's dual averaging (DA) method, a primal-dual mirror descent method that has recently seen a major resurgence in the field of large-scale optimization and machine learning. To account for settings with high temporal variability and uncertainty, we adopt a continuous-time formulation of dual averaging and we investigate the dynamics' long-run behavior when players have either noiseless or noisy information on their payoff gradients. In both the deterministic and stochastic regimes, we establish sublinear rates of convergence of actual and averaged trajectories to Nash equilibrium under a variational stability condition.**

I. Introduction

In this paper, we consider online decision processes involving several optimizing agents who interact in continuous time and whose collective actions determine their rewards at each instance. Situations of this type arise naturally in wireless communications, data networks, and many other fields where decisions are taken in real time and carry an immediate impact on the agents' welfare. Due to the real-time character of these interactions, the feedback available to the agents is often subject to estimation errors, measurement noise and/or other stochastic disturbances. As a result, every agent has to contend not only with the endogenous variability caused by other agents, but also with the exogenous uncertainty surrounding the feedback to their decision process.

Regarding the agents' interaction model, we focus on a general class of non-cooperative games with a finite number of players and continuous action sets. At each instance, players are assumed to pick an action following a continuous-time variant of Nesterov's well-known "dual averaging" method [1], itself a primal-dual extension of the universal mirror descent scheme of [2]. This method is widely used in (offline) continuous optimization and control because it is optimal from the viewpoint of worst-case black-box complexity bounds [2]. Furthermore, in the context of a single player facing a time-varying environment (sometimes referred to as a "game against nature"), it is also known that dual averaging leads to "*no regret*", i.e. the player's average payoff over time matches asymptotically that of the best fixed action in hindsight [3, 4]. As such, online control processes based on dual averaging comprise natural candidates for learning in games with continuous action sets.

In this framework, the players' individual payoff functions are determined at each instance by the actions of all other players via the underlying one-shot game. The game itself may be opaque to the players (who might not even know that they are playing a game), but the additional structure it provides means that finer convergence criteria apply, chief among them being that of convergence to a *Nash equilibrium* (NE). Thus, given the desirable properties of dual averaging in single-agent optimization problems (both offline and online), our paper focuses on the following question: *if all players employ a no-regret control policy based on dual averaging, do their actions converge to a Nash equilibrium of the underlying game?*

A. Outline of results

We begin our discussion in Section II with the notion of *variational stability* (VS), an analogue of evolutionary stability (ES) for population games [5] which was recently introduced in [6]. In a certain sense (made precise below), variational stability is to games with a finite number of players and continuous action spaces what evolutionary stability is to games with a continuum of players and a finite number of actions.

The class of learning schemes under study is introduced in Section III. Based on a Lyapunov analysis we show that the resulting (deterministic) dynamics converge to stable equilibria from any initial condition. On the other hand, a major challenge arises if the players' gradient feedback is contaminated by noise and/or other exogenous disturbances: in this case, the convergence of dual averaging is destroyed, even in simple games with a single player and one-dimensional action sets. This leads to the second question that we seek to address in this paper: *is it possible to recover the equilibrium convergence properties of dual averaging in the presence of noise and uncertainty?*

We provide a positive answer to this question in Section IV where we prove (a.s) trajectory convergence in an ergodic sense, and we also estimate the rate of convergence.

B. Related work

This paper connects learning dynamics in concave games with advanced tools from mathematical programming [1, 2, 7]. The underlying mirror descent dynamics comprise a "universal" method [7, 8] for finding approximate solutions in large-scale optimization problems relying on first-order information

P. Mertikopoulos is with the French National Center for Scientific Research (CNRS) and with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, F-38000, Grenoble, France.

M. Staudigl is with Maastricht University, Department of Quantitative Economics, Maastricht, The Netherlands.

This work was supported by the French National Research Agency (ANR) under grant no. ANR–16–CE33–0004–01 (ORACLESS) and the Huawei HIRP FLAGSHIP project ULTRON.

only. Motivated by applications to adaptive control and network optimization, a recent stream of literature has focused on continuous-time versions of noisy mirror descent schemes formulated as stochastic differential equations [9]. We extend this literature by providing first results on the long-run behavior of dual averaging and mirror descent for distributed learning in continuous time subject to random perturbations. Our method also connects to recent investigations on mirror-prox algorithms for monotone variational inequalities [1, 7, 10]. The main focus in these papers is to estimate the rate of convergence of averaged trajectories to the set of solutions to the montone variational inequalities problem. We extend these results to a continuous-time setting with unbounded random perturbations taking the form of a Brownian motion, and we show that similar guarantees can be established in our framework.

II. Preliminaries

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 closed convex subset $\mathcal{X}_i$ of an n_i-dimensional normed space $\mathcal{V}_i$, and their reward is determined by their individual objective and the profile $x = (x_1, \dots, x_N) \equiv (x_i; x_{-i})$ of all players' actions. Specifically, writing $\mathcal{X} \equiv \prod_i \mathcal{X}_i$ for the game's *action space*, each player's *payoff* is determined by an associated *payoff function* $u_i \colon \mathcal{X} \to \mathbb{R}$. In terms of regularity, we assume that u_i is differentiable in x_i and we write

$$v_i(x) \equiv \nabla_{x_i} u_i(x_i; x_{-i}) \tag{1}$$

for the *individual gradient* of u_i at x; we also assume that u_i and v_i are both Lipschitz continuous in x and we write $v(x) = (v_i(x))_{i \in \mathcal{N}}$ for the ensemble thereof. A *continuous game* is then defined as a tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ with players, action sets and payoffs defined as above.

An important class of such games is when the players' payoff functions are individually concave, viz.

$$u_i(x_i; x_{-i}) \text{ is concave in } x_i \text{ for all } x_{-i} \in \textstyle\prod_{j \neq i} \mathcal{X}_j,\ i \in \mathcal{N}. \tag{2}$$

Following Rosen [11], we say that $\mathcal{G}$ is itself *concave* in this case. We present a motivating example below:

Example 1 (Contention-based medium access). Consider a set of wireless users $\mathcal{N} = \{1, \dots, N\}$ accessing a shared wireless channel. Successful communication occurs when a user is alone in the channel and a collision occurs otherwise. If each user $i \in \mathcal{N}$ accesses the channel with probability x_i, the *contention measure* of user i is defined as $q_i(x_{-i}) = 1 - \prod_{j \neq i}(1 - x_j)$, i.e. it is the probability of user i colliding with another user. In the well known contention-based medium access framework of [12], the payoff of user i is then given by

$$u_i(x) = R_i(x_i) - \tfrac{1}{N} x_i q_i(x_{-i}), \tag{3}$$

where $R_i(x_i)$ is a concave, nondecreasing function that represents the utility of user i when there are no other users in the channel. The resulting random access game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ is then easily seen to be concave in the sense of (2).

The fundamental solution concept in non-cooperative games is that of a *Nash equilibrium* (NE). Formally, $x^* \in \mathcal{X}$ is a Nash equilibrium of $\mathcal{G}$ if

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

Importantly, if x^* is a Nash equilibrium, we have the following concise characterization [6, 13]:

Proposition 1. *If $x^* \in \mathcal{X}$ is a Nash equilibrium of $\mathcal{G}$, then*

$$\langle v(x^*), x - x^* \rangle \leq 0 \quad \textit{for all } x \in \mathcal{X}. \tag{4}$$

The converse also holds if the game is concave.

By Proposition 1, if the game is concave, existence of Nash equilibria follows from standard results [14]. Using a similar variational characterization, Rosen [11] established the following sufficient condition for equilibrium uniqueness:

Proposition 2 ([11])**.** *Assume that $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ satisfies the payoff monotonicity condition*

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

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

Games satisfying (MC) are called (*strictly*) *monotone* and they enjoy properties similar to those of (strictly) concave functions [10]. Combining Proposition 1 and (MC), it follows that the (necessarily unique) Nash equilibrium of a monotone game satisfies the inequality

$$\langle v(x), x - x^* \rangle \leq \langle v(x^*), x - x^* \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}. \tag{5}$$

Motivated by this, we introduce below the following stability notion:

Definition 1. We say that $x^* \in \mathcal{X}$ is *variationally stable* (or simply *stable*) if

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

with equality if and only if $x = x^*$.

As we remarked in the introduction, variational stability is formally similar to the notion of evolutionary stability [15, 16] for population games (i.e. games with a continuum of players and a common, finite set of actions $\mathcal{A}$). In this sense, 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.

We should also note here that variational stability *does not* presuppose that x^* is Nash equilibrium of $\mathcal{G}$. Nonetheless, as shown in [6], this is indeed the case:

Proposition 3. *If x^* is variationally stable, it is the game's unique Nash equilibrium.*

As an example, it is easy to verify that the random access game $\mathcal{G}$ of Example 1 admits a unique NE which is variationally stable in the case of "diminishing returns", i.e. when $R_i''(x_i) < -1$ [12]. It is also easy to verify that (VS) is satisfied in concave potential games, so variational stability has a wide range of applications in game theory.

III. Dual averaging with perfect information

A. Preliminaries on dual averaging

Motivated by Nesterov's original approach for solving offline optimization problems and variational inequalities [1], we focus on the following multi-agent online learning scheme: At each instance $t \geq 0$, every player takes an infinitesimal step along the individual gradient of their objective function; the output is then "mirrored" on each player's feasible region $\mathcal{X}_i$ and the process of play continues. Formally, this process boils down to the continuous-time dynamics

$$\begin{aligned} \dot{y}_i &= v_i(x), \\ x_i &= Q_i(\eta_i y_i), \end{aligned} \tag{DA}$$

where:

1. $v_i(x)$ is the individual payoff gradient of player i.
2. y_i is a "dual" variable that aggregates gradient steps.
3. $Q_i(y_i)$ denotes the *mirror* (or *choice*) map that outpus the i-th player's action as a function of the dual vector y_i.
4. $\eta_i > 0$ is a player-specific *sensitivity* parameter.

Given that the dual variables y_i aggregate individual gradient steps, a first choice for Q_i would be the arg max correspondence $y_i \mapsto \arg\max_{x_i \in \mathcal{X}_i} \langle y_i, x_i \rangle$ whose output is most closely aligned with y_i. However, this assignment generically selects only extreme points of $\mathcal{X}_i$, so it is ill-suited for general, nonlinear problems. On that account, (DA) is typically run with "regularized" mirror maps of the form $y_i \mapsto \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}$ where the regularization term $h_i(x_i)$ satisfies the following:

Definition 2. A continuous function $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is a *regularizer* on $\mathcal{X}_i$ if it is *strongly convex*, i.e.

$$h_i(\lambda x_i + (1-\lambda) x_i') \leq \lambda h_i(x_i) + (1-\lambda) h_i(x_i') - \tfrac{1}{2} K_i \lambda (1-\lambda) \|x_i' - x_i\|^2, \tag{6}$$

for some $K_i > 0$ and all $x_i, x_i' \in \mathcal{X}_i$, $\lambda \in [0,1]$. The *mirror map* induced by h_i is then given by

$$Q_i(y_i) = \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}. \tag{7}$$

The archetypal mirror map is the *Euclidean projector*

$$\Pi_i(y_i) = \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - \tfrac{1}{2} {x_i}^2\} = \arg\min_{x_i \in \mathcal{X}_i} \|y_i - x_i\|^2. \tag{8}$$

For more examples (such as logit choice in the case of simplex-like action sets), the reader is referred to [1, 6, 7, 17].

Concerning the parameter η, we see that the "deflated" mirror map $Q_i(\eta_i y_i)$ selects points that are closer to the "prox-center" $p_i \equiv \arg\min h_i$ of $\mathcal{X}_i$ as $\eta_i \to 0$. Therefore, for small η_i, the generated sequence of play becomes less susceptible to changes in the scoring variables y_i – hence the name "sensitivity".

B. Convergence analysis

Our analysis of (DA) will be based on the so-called *Fenchel coupling* [6], defined here as

$$F_\eta(p, y) = \sum_{i \in \mathcal{N}} \frac{1}{\eta_i} [h_i(p_i) + h_i^*(\eta_i y_i) - \langle \eta_i y_i, x_i \rangle], \tag{9}$$

where p is a basepoint in $\mathcal{X}$ and

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

denotes the convex conjugate of h_i [18]. This "primal-dual" coupling collects all terms of Fenchel's inequality [18], so we have $F_\eta(p, y) \geq 0$ with equality if and only if $p_i = Q_i(\eta_i y_i)$. Moreover, $F(p, y)$ enjoys the key comparison property [6, Prop. 4.3]

$$F_\eta(p, y) \geq \sum_{i \in \mathcal{N}} \frac{K_i}{2\eta_i} \|Q(\eta_i y_i) - p_i\|^2, \tag{11}$$

so $x(t) \to p$ whenever $F(p, y(t)) \to 0$. Because of this key property, convergence to a target point $p \in \mathcal{X}$ can be checked by showing that $F_\eta(p, y(t)) \to 0$.

To state our deterministic convergence result for (DA), it will be convenient to introduce the *equilibrium gap*

$$\epsilon(x) = \langle v(x), x^* - x \rangle. \tag{12}$$

Obviously, if x^* is stable, we have $\epsilon(x) \geq 0$ with equality if and only if $x = x^*$; as such, $\epsilon(x)$ can be seen as a (game-dependent) measure of the distance between x and x^*. We then have:

Theorem 1. *If $\mathcal{G}$ admits a variationally stable state x^*, every solution $x(t)$ of* (DA) *converges to x^*. Moreover, the average equilibrium gap $\bar{\epsilon}(t) = t^{-1} \int_0^t \epsilon(x(s))\, ds$ of $x(t)$ vanishes as*

$$\bar{\epsilon}(t) \leq V_0 / t, \tag{13}$$

where $V_0 \geq 0$ depends only on the initialization of (DA).

Theorem 1 is a strong convergence result guaranteeing global trajectory convergence to Nash equilibrium and an $\mathcal{O}(1/t)$ decay rate for the merit function $\bar{\epsilon}(t)$. Our proof (cf. Appendix A) relies on the fact that the Fenchel coupling $F_\eta(x^*, y)$ is a strict Lyapunov function for (DA), i.e. $F_\eta(x^*, y(t))$ is decreasing whenever $x(t) \neq x^*$. Building on this, our aim in the rest of this paper will be to explore how the strong guarantees of (DA) are affected if the players' gradient input is contaminated by observation noise and/or other stochastic disturbances.

IV. Learning under uncertainty

To account for errors in the players' feedback process, we will focus on the disturbance model

$$\begin{aligned} dY_i &= v_i(X)\, dt + dZ_i, \\ X_i &= Q_i(\eta_i Y_i), \end{aligned} \tag{SDA}$$

where $Z_i(t)$ is a continuous Itô martingale of the general form

$$dZ_{i,k}(t) = \sum_{\ell=1}^{m_i} \sigma_{i,k\ell}(X(t), t)\, dW_{i,\ell}(t), \quad k = 1, \dots, n_i, \tag{14}$$

and:

1) $W_i = (W_{i,\ell})_{\ell=1}^{m_i}$ is an m_i-dimensional Wiener process with respect to some stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$.[1]
2) The $n_i \times m_i$ *volatility matrix* $\sigma_i \colon \mathcal{X}_i \times \mathbb{R}_+ \to \mathbb{R}^{n_i \times m_i}$ of $Z_i(t)$ is measurable, bounded, and Lipschitz continuous in the

[1] In particular, we do not assume here that $m_i = n_i$; more on this below.

first argument. Specifically, we make the following noise regularity assumption:

$$\begin{aligned} &\sup_{x,t}\|\sigma_i(x,t)\| < \infty, \\ &\|\sigma_i(x',t) - \sigma_i(x,t)\| = \mathcal{O}(\|x'-x\|). \end{aligned} \tag{NR}$$

for all $x, x' \in \mathcal{X}$ and all $t \geq 0$.

Clearly, the noise in (SDA) may depend on t and $X(t)$ in a fairly general way: for instance, the increments of $Z_i(t)$ need not be i.i.d. and different components of Z_i need not be independent either. Such correlations can be captured by the *quadratic covariation* $[Z_i, Z_i]$ of Z_i, given here by

$$d[Z_{i,k}, Z_{i,\ell}] = \sum_{r,s=1}^{m_i} \sigma_{i,kr}\sigma_{i,\ell s}\, dW_{i,r} \cdot dW_{i,s} = \Sigma_{i,k\ell}\, dt, \tag{15}$$

where $\Sigma_i = \sigma_i \sigma_i^\top$ [19]. As a consequence of (NR), we then have

$$\|\sigma(x,t)\|_F^2 \equiv \operatorname{tr}[\Sigma(x,t)] \leq \sigma_*^2 \quad \text{for some } \sigma_* > 0. \tag{16}$$

The bound σ_* essentially captures the intensity of the noise affecting the players' observations in (SDA); obviously, when $\sigma_* = 0$, we recover the noiseless dynamics (DA).

A first observation regarding (SDA) is that the induced sequence of play $X_i(t) = Q_i(\eta_i(t)Y_i(t))$ may fail to converge with probability 1. A simple example of this behavior is as follows: consider a single player with action space $\mathcal{X} = [-1, 1]$ and payoff function $u(x) = 1 - x^2/2$. Then, $v(x) = \nabla u(x) = -x$ for all $x \in [-1, 1]$, so (SDA) takes the form

$$\begin{aligned} dY &= -X\, dt + \sigma\, dW, \\ X &= [Y]_{-1}^{1}, \end{aligned} \tag{17}$$

where, for simplicity, we took $\eta = 1$, σ constant, and we used the shorthand $[x]_a^b$ for x if $x \in [a, b]$, a if $x \leq a$, and b if $x \geq b$.

In this case, the game's unique Nash equilibrium obviously corresponds to $X = Y = 0$. However, the dynamics (17) describe a truncated Ornstein–Uhlenbeck (OU) process [19], leading to the explicit solution formula

$$Y(t) = C_{t_0} e^{-t} + \sigma \int_{t_0}^{t} e^{-(t-s)}\, dW(s) \quad \text{for some } C_{t_0} \in \mathbb{R}, \tag{18}$$

valid whenever $Y(s) \in [-1, 1]$ for $s \in [t_0, t]$. Thanks to this expression, we conclude that (SDA) cannot converge to Nash equilibrium with positive probability in the presence of noise.

Despite the nonconvergence of (17) in general games, the induced sequence of play roughly stays within $\mathcal{O}(\sigma)$ of the game's Nash equilibrium for most of the time (and with high probability). Hence, it stands to reason that if the players employed a sufficiently small sensitivity parameter η, the primal process $X(t) = Q(\eta Y(t))$ would be concentrated even more closely around 0. This observation suggests that using a *vanishing* sensitivity parameter $\eta_i \equiv \eta_i(t)$ which decreases to 0 as $t \to \infty$ could be more beneficial in the face of uncertainty. With this in mind, we make the following assumption:

$$\eta_i(t) \text{ is Lipschitz, nonincreasing, and } \lim_{t\to\infty} t\eta_i(t) = \infty. \tag{19}$$

Under this assumption, we have:

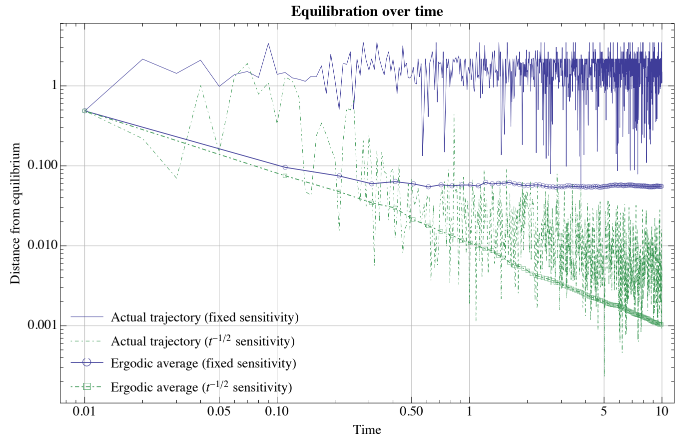

Fig. 1: Convergence to Nash equilibrium in the game of Example 1. When (DA) is run with a fixed sensitivity, neither the actual trajectory $X(t)$ (blue line) nor its time-average $\bar{X}(t) = t^{-1}\int_0^t X(s)\, ds$ converge (blue line with circles). If run with a $t^{-1/2}$ sensitivity schedule, $X(t)$ gets closer to the optimum (dashed green line) and its time-average converges following a power law (dashed green line with squares).

Theorem 2. *Suppose that $\mathcal{G}$ admits a variationally stable equilibrium x^*. Then, the long-term average equilibrium gap $\bar{\epsilon}(t) = t^{-1}\int_0^t \epsilon(X(s))\, ds$ vanishes as*

$$\bar{\epsilon}(t) \leq \sum_{i\in\mathcal{N}} \left[\frac{\Omega_i}{t\eta_i(t)} + \frac{\sigma_*^2}{2K_i t}\int_0^t \eta_i(s)\, ds\right] + \mathcal{O}(t^{-1/2}\log\log t), \tag{20}$$

and enjoys the mean bound

$$\mathbb{E}[\bar{\epsilon}(t)] \leq \sum_{i\in\mathcal{N}} \left[\frac{\Omega_i}{t\eta_i(t)} + \frac{\sigma_*^2}{2K_i t}\int_0^t \eta_i(s)\, ds\right], \tag{21}$$

where $\Omega_i = \max h_i - \min h_i$ is a positive constant.

Corollary 1. *If $\lim_{t\to\infty}\eta_i(t) = 0$ and $\mathcal{G}$ is monotone (in particular, if $\mathcal{G}$ admits a concave potential), the long-run average $\bar{X}(t) = t^{-1}\int_0^t X(s)\, ds$ of $X(t)$ converges to x^* (a.s.).*

Corollary 2. *If $\eta_i(t) \propto t^{-\beta}$ for sufficiently large t, we have*

$$\mathbb{E}[\bar{\epsilon}(t)] = \begin{cases} \mathcal{O}(t^{-\beta}) & \text{if } 0 < \beta < \frac{1}{2}, \\ \mathcal{O}(t^{-1/2}) & \text{if } \beta = \frac{1}{2}, \\ \mathcal{O}(t^{\beta-1}) & \text{if } \frac{1}{2} < \beta < 1, \end{cases} \tag{22}$$

In fact, if the growth of $v(x)$ near a stable state $x^* \in \mathcal{X}$ is suitably bounded, we can obtain even finer results for the rate of convergence of (DA). Specifically, if $\mathcal{G}$ grows as

$$\langle v(x), x^* - x\rangle \geq B\|x^* - x\|^\gamma \quad \text{for some } B > 0, \gamma \geq 1, \tag{23}$$

we have:

Proposition 4. *Suppose that (23) holds and (SDA) is run with $\eta_i(t) \propto t^{-1/2}$. Then, the long-run average $\bar{X}(t) = t^{-1}\int_0^t X(s)\, ds$ of $X(t)$ enjoys the (a.s.) convergence rate*

$$\|\bar{X}(t) - x^*\| = \widetilde{\mathcal{O}}(t^{-1/2\gamma}). \tag{24}$$

Proof: Jensen's inequality and (23) readily yield

$$\|\bar{X}(t) - x^*\|^\gamma \leq \frac{1}{t}\int_0^t \|X(s) - x^*\|^\gamma\, ds \leq B^{-1}\bar{\epsilon}(t), \tag{25}$$

so our claim follows from Theorem 2. ■

With this result, we are able to explicitly control the distance of the averaged trajectory to the underlying Nash equilibrium of the game. The precise rate of convergence is then obtained by exploiting the fine details of the game's payoff functions. We are not aware of similar results in the discrete-time literature.

V. Conclusions and perspectives

In this paper, we investigated the convergence of a class of distributed dual averaging schemes for games with continuous action sets. When players have access to perfect gradient information, dual averaging converges to variationally stable Nash equilibria from any initial condition. On the other hand, in the presence of feedback noise and uncertainty, trajectory convergence is destroyed. To rectify this, we introduced a variable sensitivity parameter which allows us to recover convergence to stable Nash equilibria in an ergodic sense, as well as providing an estimate of the rate of convergence to such states.

Two questions that arise are whether it is possible to obtain stronger convergence results (*i*) when the noise in the players' feedback vanishes over time (corresponding to the case where the players' feedback becomes more accurate as measurements accrue over time); and (*ii*) when the Nash equilibrium has a special structure (for instance, if it is interior or a corner point of $\mathcal{X}$). We leave these questions for future work.

Appendix A
Deterministic analysis

We begin with some basic properties of the Fenchel coupling (9) that will also be used in Appendix B. To that end, by the basic properties of convex conjugation and the fact that h_i is K_i-strongly convex, it follows (see e.g. [18, Theorem 23.5] or [6, Proposition 3.2]) that h_i^* is continuously differentiable and

$$Q_i(y_i) = \nabla h_i^*(y_i) \tag{26}$$

is $\frac{1}{K_i}$-Lipschitz under the *dual norm* $\|y\|_* = \sup_{x:\|x\|\leq 1}\langle y, x\rangle$ [3, Chap. 2]. Using this relation between h^* and Q, we obtain the following Lyapunov-like property of the Fenchel coupling:

Lemma 1. *Let $V(t) = F_\eta(x^*, y(t))$. Then, under (DA), we have*

$$\dot{V}(t) = \langle v(x(t)), x(t) - x^*\rangle. \tag{27}$$

Proof: By Eq. (9) we get:

$$\begin{aligned}\frac{dV}{dt} &= \sum_{i\in\mathcal{N}} \frac{1}{\eta_i}[\langle \eta_i\dot{y}_i, \nabla h_i^*(\eta_i y_i)\rangle - \langle \eta_i\dot{y}_i, x_i^*\rangle] \\ &= \sum_{i\in\mathcal{N}}\langle \dot{y}_i, Q_i(\eta_i y_i) - x_i^*\rangle = \langle v(x), x - x^*\rangle, \end{aligned} \tag{28}$$

as claimed. ■

Proof of Theorem 1: To begin with, (27) and (VS) yield

$$V(y(t)) - V(y(0)) = \int_0^t \langle v(x(s)), x(s) - x^*\rangle\, ds = -t\bar{\epsilon}(t), \tag{29}$$

and hence, letting $V_0 \equiv V(y(t))$, we have:

$$\bar{\epsilon}(t) = \frac{V(y(0)) - V(y(t))}{t} \leq \frac{V_0}{t}. \tag{30}$$

For the second part, let $\hat{x}$ be an ω-limit of $x(t)$ and assume that $\hat{x} \neq x^*$. Then, by continuity, there exists a neighborhood U of $\hat{x}$ in $\mathcal{X}$ such that $\langle v(x), x - x^*\rangle \leq -a$ for some $a > 0$. Furthermore, since $\hat{x}$ is an ω-limit of $x(t)$, there exists an increasing sequence of times $t_k \uparrow \infty$ such that $x(t_k) \in U$ for all k. Then, by the definition of Q, we have

$$\begin{aligned}\|x_i(t_k+\tau) - x_i(t_k)\| &= \|Q_i(\eta_i y_i(t_k+\tau)) - Q_i(\eta_i y_i(t_k))\| \\ &\leq \frac{\eta_i}{K_i}\|y_i(t_k+\tau) - y_i(t_k)\|_* \\ &\leq \frac{\eta_i}{K_i}\int_{t_k}^{t_k+\tau}\|v_i(x(s))\|_*\, ds \leq \frac{\eta\tau}{K_i}\max_{x_i\in\mathcal{X}_i}\|v_i(x)\|_*. \end{aligned} \tag{31}$$

Since (31) does not depend on k, there exists some sufficiently small $\delta > 0$ such that $x(t_k+\tau) \in U$ for all $\tau \in [0, \delta]$, $k \in \mathbb{N}$ (so we also have $\langle v(x(t_k+\tau)), x(t_k+\tau) - x^*\rangle \leq -a$). Combining this with the fact that $\langle v(x), x - x^*\rangle \leq 0$ for all $x \in \mathcal{X}$, we get

$$\begin{aligned}V(y(t_k+\delta)) &\leq V(y(0)) + \sum_{\ell=1}^{k}\int_{t_\ell}^{t_\ell+\delta}\langle v(x(s)), x(s) - x^*\rangle\, ds \\ &\leq V(y(0)) - ak\delta, \end{aligned} \tag{32}$$

showing that $\liminf_{t\to\infty} V(y(t)) = -\infty$, a contradiction. Since $x(t)$ admits at least one ω-limit in $\mathcal{X}$, we get $x(t) \to x^*$. ■

Appendix B
Stochastic analysis

We first show that the Fenchel coupling $V = F_\eta(x^*, Y(t))$ satisfies a noisy version of Lemma 1:

Lemma 2. *Let $x^* \in \mathcal{X}$. Then, for all $t \geq 0$, we have*

$$V(Y(t)) \leq V(Y(0))$$

$$+ \int_0^t \langle v(X(s)), X(s) - x^*\rangle\, ds \tag{33a}$$

$$- \sum_{i\in\mathcal{N}}\int_0^t \frac{\dot{\eta}_i(s)}{\eta_i(s)^2}[h_i(x_i^*) - h_i(X_i(s))]\, ds \tag{33b}$$

$$+ \sum_{i\in\mathcal{N}}\frac{1}{2K_i}\int_0^t \eta_i(s)\ \mathrm{tr}[\Sigma_i(X(s), s)]_*\, ds \tag{33c}$$

$$+ \sum_{i\in\mathcal{N}}\sum_{k=1}^{n_i}\int_0^t (X_{i,k}(s) - x_{i,k}^*)\, dZ_{i,k}(s). \tag{33d}$$

Proof: The proof of the lemma follows from the (weak) Itô's lemma proved in [20, Lemma C.2]. ■

Our final result is a growth estimate for Itô martingales with bounded volatility, proved in [21]:

Lemma 3. *Let $W(t)$ be a Wiener process in $\mathbb{R}^m$ and let $\zeta(t)$ be a bounded, continuous process in $\mathbb{R}^m$. Then, for every function $f\colon [0,\infty) \to (0,\infty)$, we have*

$$f(t) + \int_0^t \zeta(s)\cdot dW(s) \sim f(t) \quad \text{as } t\to\infty \text{ (a.s.)}, \tag{34}$$

whenever $\lim_{t\to\infty}(t\log\log t)^{-1/2} f(t) = +\infty$.

With all this at hand, we are finally in a position to prove Theorem 2:

Proof of Theorem 2: After rearranging, Lemma 2 yields

$$\int_0^t \langle v(X(s)), x^* - X(s)\rangle \, ds \tag{35a}$$

$$\leq V(0) - V(t) \tag{35b}$$

$$- \sum_{i\in\mathcal{N}} \int_0^t \frac{\dot{\eta}_i(s)}{\eta_i^2(s)} \left[h_i(x_i^*) - h_i(X_i(s))\right] ds \tag{35c}$$

$$+ \sum_{i\in\mathcal{N}} \frac{1}{2K_i} \int_0^t \eta_i(s) \operatorname{tr}[\Sigma_i(X(s), s)] \, ds \tag{35d}$$

$$+ \sum_{i\in\mathcal{N}} \sum_{k=1}^{n_i} \int_0^t (X_{i,k}(s) - x_{i,k}^*) \, dZ_{i,k}(s) \tag{35e}$$

We now proceed to bound each term of (35):

a) Since $V \geq 0$ for all t, (35b) is bounded from above by V_0.

b) For (35c), let $\Omega_i = \max_{\mathcal{X}_i} h_i - \min_{\mathcal{X}_i} h_i$. Then, we have $h_i(x_i^*) - h_i(X_i(s)) \leq \Omega_i$, so, with η_i nonincreasing, we get

$$(35c) \leq - \sum_{i\in\mathcal{N}} \Omega_i \int_0^t \frac{\dot{\eta}_i(s)}{\eta_i^2(s)} \, ds = \sum_{i\in\mathcal{N}} \left[\frac{\Omega_i}{\eta_i(t)} - \frac{\Omega_i}{\eta_i(0)}\right] \tag{36}$$

because $\lim_{t\to\infty} t\eta(t) = \infty$ by assumption (recall also that $\dot{\eta} \leq 0$).

c) For (35d), the definition of σ_*^2 gives immediately

$$(35d) \leq \sum_{i\in\mathcal{N}} \frac{\sigma_*^2}{2K_i} \int_0^t \eta_i(s) \, ds. \tag{37}$$

d) Finally, for (35e), let $\psi_i(t) = \int_0^t \sum_{k=1}^{n_i} (X_{i,k}(s) - x_{i,k}^*) \, dZ_{i,k}(s)$ and set $\rho_i = [\psi_i, \psi_i]$ for the quadratic variation of ψ_i. Then:

$$\begin{aligned} d[\psi_i, \psi_i] = d\psi_i \cdot d\psi_i &= \sum_{k,\ell=1}^{n_i} \Sigma_{i,k\ell} (X_{i,k} - x_{i,\ell}^*)(X_{i,\ell} - x_{i,\ell}^*) \, dt \\ &\leq \sigma_*^2 \|X_i(s) - x_i^*\|^2 \, dt, \end{aligned} \tag{38}$$

so $\rho_i(t) \leq R\sigma_*^2 \|\mathcal{X}_i\|^2 t$ for some norm-dependent constant $R > 0$. Then, by a standard time-change argument [19, Problem 3.4.7], there exists a one-dimensional Wiener process $\widetilde{W}_i(t)$ with induced filtration $\widetilde{\mathcal{F}}_s = \mathcal{F}_{\tau_{\rho_i}(s)}$ and such that $\widetilde{W}_i(\rho_i(t)) = \psi_i(t)$ for all $t \geq 0$. By the law of the iterated logarithm [19], we then obtain

$$\begin{aligned} &\limsup_{t\to\infty} \frac{\widetilde{W}_i(\rho_i(t))}{\sqrt{2Mt \log\log(Mt)}} \\ &\quad \leq \limsup_{t\to\infty} \frac{\widetilde{W}_i(\rho_i(t))}{\sqrt{2\rho_i(t) \log\log \rho_i(t)}} = 1 \quad \text{(a.s.)}, \end{aligned} \tag{39}$$

where $M = \sigma_*^2 R \sum_{i\in\mathcal{N}} \|\mathcal{X}_i\|^2$. Thus, with probability 1, we have $\psi_i(t) = \mathcal{O}(\sqrt{t \log\log t})$.

Combining all of the above and dividing by t, we then get

$$\begin{aligned} \bar{\epsilon}(t) &= \frac{1}{t} \int_0^t \langle v(X(s)), x^* - X(s)\rangle \, ds \\ &\leq \sum_{i\in\mathcal{N}} \left[\frac{\Omega_i}{t\eta_i(t)} + \frac{\sigma_*^2}{2tK_i} \int_0^t \eta_i(s) \, ds\right] + \mathcal{O}(t^{-1/2} \sqrt{\log\log t}), \end{aligned}$$

where we have absorbed all $\mathcal{O}(1/t)$ terms in the logarithmic term $\mathcal{O}(\sqrt{t^{-1} \log\log t})$. ∎

References

[1] Y. Nesterov, "Primal-dual subgradient methods for convex problems," *Mathematical Programming*, vol. 120, no. 1, pp. 221–259, 2009.

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

[3] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[4] J. Kwon and P. Mertikopoulos, "A continuous-time approach to online optimization," *Journal of Dynamics and Games*, vol. 4, pp. 125–148, April 2017.

[5] J. Maynard Smith, "Game theory and the evolution of fighting," in *On Evolution*, pp. 8–28, Edinburgh: Edinburgh University Press, 1972.

[6] P. Mertikopoulos, "Learning in games with continuous action sets and unknown payoff functions." https://arxiv.org/abs/1608.07310, 2016.

[7] A. S. Nemirovski, A. Juditsky, G. G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[8] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.

[9] M. Raginsky and J. Bouvrie, "Continuous-time stochastic mirror descent on a network: Variance reduction, consensus, convergence," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 6793–6800, IEEE, 2012.

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

[11] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica: Journal of the Econometric Society*, pp. 520–534, 1965.

[12] T. Cui, L. Chen, and S. H. Low, "A game-theoretic framework for medium access control," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, 2008.

[13] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[14] G. Debreu, "A social equilibrium existence theorem," *Proceedings of the National Academy of Sciences of the U.S.A.*, vol. 38, pp. 886–893, 1952.

[15] J. Maynard Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, pp. 15–18, 1973.

[16] J. Hofbauer, P. Schuster, and K. Sigmund, "A note on evolutionarily stable strategies and game dynamics," *Journal of Theoretical Biology*, vol. 81, pp. 609–612, 1979.

[17] P. Mertikopoulos and W. H. Sandholm, "Learning in games via reinforcement and regularization," *Mathematics of Operations Research*, vol. 41, pp. 1297–1324, November 2016.

[18] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.

[19] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. Berlin: Springer-Verlag, 1998.

[20] P. Mertikopoulos and M. Staudigl, "On the convergence of gradient-like flows with noisy gradient input." http://arxiv.org/abs/1611.06730, 2016.

[21] M. Bravo and P. Mertikopoulos, "On the robustness of learning in games with stochastically perturbed payoff observations," *Games and Economic Behavior*, vol. 103, John Nash Memorial issue, pp. 41–66, May 2017.