Mirror Descent Learning in Continuous Games

Zhengyuan Zhou, Panayotis Mertikopoulos, Aris L. Moustakas, Nicholas Bambos, and Peter Glynn

***Abstract*—Online Mirror Descent (OMD) is an important and widely used class of adaptive learning algorithms that enjoys good regret performance guarantees. It is therefore natural to study the evolution of the joint action in a multi-agent decision process (typically modeled as a repeated game) where every agent employs an OMD algorithm. This well-motivated question has received much attention in the literature that lies at the intersection between learning and games. However, much of the existing literature has been focused on the time average of the joint iterates. In this paper, we tackle a harder problem that is of practical utility, particularly in the online decision making setting: the convergence of the last iterate when all the agents make decisions according to OMD. We introduce an equilibrium stability notion called *variational stability* (VS) and show that in variationally stable games, the last iterate of OMD converges to the set of Nash equilibria. We also extend the OMD learning dynamics to a more general setting where the exact gradient is not available and show that the last iterate (now random) of OMD converges to the set of Nash equilibria almost surely.**

I. Introduction

Online decision making is a broad and powerful paradigm that has found widespread applications and has achieved great success (see [1] for a survey). The archetypal online decision process may be described as follows: At each instance $t = 1, 2, \dots$, a *player* (viewed here as an optimizing agent) selects an action x^t from some set $\mathcal{X}$ and obtains a reward $u^t(x^t)$ based on an a priori unknown payoff function $u^t \colon \mathcal{X} \to \mathbb{R}$. Subsequently, the player receives some feedback (such as the past history of the reward functions or some restricted information thereof), and selects a new action x^{t+1} with the goal of maximizing the obtained reward. Aggregating over the stages of the process, this is usually quantified by asking that the player's (external) *regret* $R^t \equiv \max_{x\in\mathcal{X}} \sum_{k=1}^{t} \left[u^k(x) - u^k(x^k)\right]$ grow sublinearly with t, a property known as "no regret".

Starting with the seminal work of [2], one of the most widely used learning algorithms for achieving no regret in such online decision problems is the *online mirror descent* (OMD) class of algorithms proposed by [3]. This class contains several closely related variants and, perhaps unsurprisingly, takes on different names such as dual averaging (DA) [4] or (lazy) online mirror descent [5] and so on. In a nutshell, the main idea of OMD is as follows: at each stage $t = 1, 2, \dots$, every player takes a step along (an estimate of) the individual gradient of their payoff function and the output is "mirrored" onto each player's action space by means of a "choice map" that is analogous to ordinary Euclidean projection – in fact, it is a natural generalization thereof. Note that the merit of OMD lies not only in its provable regret guarantees, but also in the parsimonious feedback required: a player need only have access to a single piece of gradient information of the previous iteration's reward function (i.e. evaluated at the previous decision point).

So far, the reward (or cost) function for a single-player online decision problem is kept at the most general and abstract level. A common instantiation for such an abstract reward function lies in repeated multi-player games, where each player's payoff function is determined by the actions of all the players via a fixed mechanism – the *stage game* (even though this mechanism may be unknown and/or opaque to the players). For any particular player, its reward function (as a function of his own action alone) will depend both on its own utility function and on the actions adopted by all the other players. Given the merits of no-regret algorithms (OMD being one of them), it is natural to expect that every player will adopt one in selecting their actions in the decision process. This leads to the following central question: what is the evolution of the joint action when every player adopts a no-regret learning algorithm? In particular, *if all players of a repeated game employ an updating rule that leads to no regret, do their actions converge to a Nash equilibrium of the one-shot stage game?*

In fact, these well-motivated questions have generated an extensive literature that lies at the intersection between learning and game theory, which has witnessed a surge in interest in the past decade or so. Most of this literature has focused on the convergence of the time average of the iterates $\bar{\mathbf{x}}^t = \sum_{k=1}^{t} \gamma_k \mathbf{x}^k / \sum_{k=1}^{t} \gamma^k$ (where γ^t is the step-size and $\mathbf{x}^t$ is the joint action at time t) as opposed to the last iterate $\mathbf{x}^t$ (i.e. actual sequence of joint action employed by the players). For instance, it is well-known that if each player employs a no-regret learning algorithm, the time average of the iterates converges to the Hannan set [6] (perhaps the coarsest equilibrium notion). If each player plays employs a no-internal-regret learning algorithm (a stronger regret notion), then the time average of the iterates converges to a finer equilibrium, called correlated equilibrium [6]. Convergence to a Nash equilibrium is, in the words of [6] "considerably more

Z. Zhou, N. Bambos and P. Glynn are with the Department of Electrical Engineering and Department of Management Science and Engineering, Stanford University, CA, 94305, USA.

P. Mertikopoulos is with Univ. Grenoble Alpes, CNRS, Grenoble INP, Inria, LIG, F-38000, Grenoble, France.

A. Moustakas is with Department of Physics, University of Athens and Institute of Accelerating Systems and Applications (IASA), Athens, Greece.

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. A. Moustakas was partially supported by the Huawei Innovation Research Program ULTRON.

A. Moustakas was partially supported by the Huawei Innovation Research Program ULTRON.

difficult", because Nash equilibrium is the finest equilibrium[1]: in general, such convergence results do not hold for Nash equilibria. However, it is also the most meaningful question since it is the most stable equilibrium (and hence one that has the most predictive power).

As such, a growing literature has been devoted to studying this problem, each focusing on special classes of games. Here again, the attention is almost exclusively focused on the convergence of the time average [7–9]. However, the convergence of the last iterate is also worth investigating for two reasons. First, the convergence of the last iterate is stronger and theoretically more appealing: it is easier for the average iterate to converge than for the last iterate to converge. Second, it is the convergence of the last iterate rather than that of the average, that is of principal interest and practical utility for an online repeated game setting considered here. This is the main question we tackle: our overarching objective is to analyze the last-iterate convergence properties of OMD, a wide class of adaptive learning algorithms, for continuous games.

Our Contributions

Our contributions are threefold. First, we introduce an equilibrium stability notion called *variational stability* (VS), which is formally similar to the influential notion of *evolutionary stability* introduced in [10]. Variational stability allows us to look at the general class of continuous games as opposed to a specific class of games (such as zero-sum or potential games). Further, variational stability is related to monotone operators in variational analysis [11] and can be seen as a generalization of operator monotonicity in the current game context and results in desirable structural properties of the game's Nash equilibria. In Section III we give two classes of games that satisfy this equilibrium notion (as well as a convenient sufficient condition that ensures variational stability). As an important example in engineering applications, convex potential games is one such special case. In addition, both the class of monotone games introduced in [12] and the broader class of pseudo-monotone games introduced in [13] are special cases satisfying the equilibrium notion introduced here. See Section III-C for a detailed discussion.

Second, we show that under variational stability, the last iterate of OMD converges to the set of Nash equilibria. In particular, when a unique Nash equilibrium exists, the last iterate of OMD converges to that unique Nash equilibrium. Our proof relies on designing a particular Lyapunov function, λ-*Fenchel coupling*, which serves as a "primal-dual divergence" measure between action and gradient variables that extends the well-known Bregman divergence. Thanks to its Lyapunov properties, the λ-Fenchel coupling provides a potent tool for proving convergence and we exploit it throughout. To the best of our knowledge, this is the first convergence result at this level of generality.

Third, we extend the OMD learning dynamics to a more general setting where the exact gradient is not available. This extension is of practical utility since on one hand, there can be noise associated with measuring/sensing the gradient in the underlying environment; and on the other hand, even if such noise is absent, a player's utility can be a random quantity fluctating from iteration to iteration. We consider the extended feedback model where players only have access to a first-order oracle providing unbiased, bounded-variance estimates of their payoff gradients at each step. Apart from this, players operate in a "black box" setting, without any knowledge of the game's structure or their payoff functions (or even that they are playing a game). Drawing tools and techniques from stochastic approximation, martingale limit theory and convex analysis. we establish that under variational stability, when a unique Nash equilibrium exists, the last iterate (now a random variable) of OMD converges to that unique Nash equilibrium almost surely. Further, when there are multiple Nash equilibria, the last iterate of OMD converges to the set of Nash equilibria almost surely.

Related work

The authors of [4, 14] already give several convergence results for dual averaging in (stochastic) convex programs and saddle-point problems, while [15] provides a thorough regret analysis for online optimization problems (with or without regularization). 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 $\bar{x}_n = \sum_{k=1}^{n} \gamma_k x_k / \sum_{k=1}^{n} \gamma_k$ From a mixed-strategy perspective, [16, 17] examined actor-critic algorithms that converge to a probability distribution that assigns most weight to equilibrium states (but still assigns positive probability to all pure strategies). At the pure strategy level, several authors have considered variational inequality (VI)-based approaches and Gauss–Seidel methods for solving generalized Nash equilibrium problems (GNEPs); for a survey, see [18] and [19]. The intersection of these works with the current paper is when the game satisfies a global monotonicity condition similar to the so-called diagonal strict concavity condition of [20]: in this case, VI methods converge to Nash equilibrium globally. That being said, 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, we emphasize the distinction between the OMD dynamics studied in this paper and the well-known best response dynamics [12, 21] commonly encountered in game theory. In best response dynamics, each player chooses its current action assuming all the other players will adopt their respective actions in the previous round. In other words, each player's action is the best response to all the players' previous actions. Consequently, it is easy to see that in adopting best response dynamics, the current joint action of all players only depends on the joint action in the previous iterate, whereas in OMD, all the past joint actions are incorporated in selecting the current action. In fact, precisely due to the

[1]In particular, a Nash equilibrium is a correlated equilibrium, which in turn is in the Hannan set

sole dependence of the previous joint action, best response dynamics are not very stable, particularly when there is noise in the environment. See [22, 23] for two recent studies of best response dynamics under stochastic environments/feedback on two different applications: they converge at best to a stationary distribution, where in the current setting OMD converges almost surely to a constant.

II. Online Mirror Descent Learning on Continuous Games

In this section, we present the learning-on-games model, which has two main components. First, a continuous game with concave payoffs that players repeatedly play. Second, the well-known Online Mirror Descent (OMD) learning dynamics that enjoys the no-regret performance guarantee (see [6, 24] for a precise statement).

A. Continuous Games with Concave Payoff

A continuous game is a multi-player game with continuous actions sets. Here we focus on the class of continuous games that have concave payoffs [2]

Definition 1. A continuous game $\mathcal{G}$ with concave payoff, or a concave game in short, is given by the tuple $\mathcal{G} = (\mathcal{N}, \mathcal{X} = \prod_{i=1}^{N} \mathcal{X}_i, \{u_i\}_{i=1}^{N})$, where $\mathcal{N}$ is the set of N players $\{1, 2, \ldots, N\}$, $\mathcal{X}$ is the joint action space with $\mathcal{X}_i$ being the action space for player i and $u_i : \mathcal{X} \to \mathbf{R}$ is the utility function for player i, such that the following hold:

1) Each $\mathcal{X}_i$ is a nonempty, compact and convex subset of some finite dimensional real vector space $\mathcal{V}_i$ (i.e. $\mathcal{V}_i = \mathbf{R}^{m_i}$ for some positive integer m_i).
2) For each $i \in \mathcal{N}$, u_i is continuous in $\mathbf{x}$ and concave in x_i. The latter means that $u_i(x_i, \mathbf{x}_{-i})$ is concave in x_i for every $\mathbf{x}_{-i} \in \prod_{j \neq i} \mathcal{X}_j$. Throughout the paper, we use $\mathbf{x}_{-i}$ to denote the joint action of all players but player i. Consequently, the joint action $\mathbf{x}$ will frequently be written as $(x_i, \mathbf{x}_{-i})$.
3) For each $i \in \mathcal{N}$, u_i is continuously differentiable in x_i and each individual gradient $v_i(\mathbf{x}) \triangleq \nabla_{x_i} u_i(\mathbf{x})$ is continuous in $\mathbf{x}$.

Remark 1. A note on the notation: we use the boldfaced letter $\mathbf{x}$ to denote the joint action of all players and x_i to denote the action of player i. It should be in kept in mind, however, that x_i is a vector itself: it is not boldfaced here in order to distinguish the individual action from the joint action.

Two central quantities in this paper are the gradient of the utility functions and Nash equilibrium, respectively, which we define next.

Definition 2. We denote by $\mathbf{v}(\mathbf{x})$ to be the collection of all individual gradients of the utility functions: $\mathbf{v}(\mathbf{x}) \triangleq (v_i(\mathbf{x}))_{i \in \mathcal{N}}$, where $v_i(\mathbf{x}) \triangleq \nabla_{x_i} u_i(\mathbf{x})$, as defined in Definition 1.

[2]Note that all the convergence to Nash equilibria results in this paper continue to hold even if we do not assume the continuous games have concave payoffs. We make this concave payoff assumption mainly because the no-regret guarantee of OMD is only guaranteed to hold when the payoff is concave.

Note that per the last assumption in the definition of a concave game (Definition 1), the gradient $\mathbf{v}(\mathbf{x})$ always exists and is a continuous function on the joint action space $\mathcal{X}$.

Definition 3. Given a concave game $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^{N} \mathcal{X}_i, \{u_i\}_{i=1}^{N})$, $\mathbf{x}^* \in \mathcal{X}$ is called a Nash equilibrium if for each $i \in \mathcal{N}$, $u_i(x_i^*, \mathbf{x}_{-i}^*) \geq u_i(x_i, \mathbf{x}_{-i}^*), \forall x_i \in \mathcal{X}_i$.

Remark 2. The gradients of the utility functions and a Nash equilibrium are related. In this context, we state an equivalent characterization of Nash equilibrium that is well-known in the literature (e.g. [25]) and that will be useful later: $\mathbf{x}^*$ is a Nash equilibrium of the game if and only if for every $i \in \mathcal{N}$, and every $\mathbf{x} \in \mathcal{X}$, $\langle v_i(\mathbf{x}^*), x_i - x_i^* \rangle \leq 0$, where $\langle \cdot, \cdot \rangle$ denotes the inner-product operation.

We close this subsection by citing a seminar result in [26]:

Theorem 1. *For any concave game* $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^{N} \mathcal{X}_i, \{u_i\}_{i=1}^{N})$, *there exists a Nash equilibrium* $\mathbf{x}^* \in \mathcal{X}$.

B. Online Mirror Descent

When the players play a repeated game with each stage game being the concave game defined in the previous subsection, it is an interesting question as to what learning dynamics the players would adopt (i.e. how each player would behave in such a repeated strategically interactive setting). In fact, from the perspective of each individual player, the adaptive selection of an action in such a setting conforms to the broad and elegant online convex optimization framework as given below (see [1] for a survey).

Online Convex Optimization
Input: A Bounded Convex Set $S \subset \mathbf{R}^d$
for $t = 1, 2, \ldots$
  Choose a vector $s_t \in S$
  Receive a convex loss function $f_t(\cdot)$ and incur loss $f_t(s_t)$

Note that our current repeated game setting provides a natural formation of the convex loss function: for a particular player i, $f_t(\cdot) = -u_i(\cdot, \mathbf{x}_{-i}^t)$. In this case, even though the individual utility function is fixed, the convex loss function is time-dependent because all the other players' actions change over time, which in turn changes the convex loss function at every iteration. A well-known class of learning dynamics in the online convex optimization/online learning literature that enjoys provably good performance is online mirror descent (OMD), which, when applied to the repeated game setting here results in Algorithm 1.

Several comments are in order here. First, the gradient step size α^t in Algorithm 1 can be any positive and non-increasing sequence that satisfies the standard not-summable-but-square-summable assumption:

Definition 4. A positive and non-increasing sequence $\{\alpha^t\}_{t=0}^{\infty}$ is called slowly vanishing if the following conditions are

Algorithm 1 Online Mirror Descent under Perfect Information

1: Each player i chooses an initial y_i^0.
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: **for** $i = 1, \ldots, N$ **do**
4: $x_i^t = \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i^t, x_i \rangle - h_i(x_i)\}$
5: $y_i^{t+1} = y_i^t + \alpha^t v_i(\mathbf{x}^t)$
6: **end for**
7: **end for**

satisfied:

$$\sum_{t=0}^{\infty} \alpha^t = \infty, \sum_{t=0}^{\infty} (\alpha^t)^2 < \infty.$$

Second, we are referring to Algorithm 1 as Online Mirror Descent under Perfect Information because here we assume the exact gradient $v_i(\mathbf{x}^t)$ is available at each iteration in the update (Line 5). Later, we shall generalize this to noisy gradient case and analyze the behavior of OMD under those more general, stochastic environments. Third, we emphasize that y_i^t is an auxiliary variable that accumulates gradient in a discounted way (discounted by the pre-determined sequence $\{\alpha^t\}_{t=1}^{\infty}$), while the chosen actions is given by x_i^t. Note that x_i^t is obtained via a lazy projection on y_i^t (by finding a x_i that best aligns with y_i^t subject to a penalty induced by h_i). Fourth, the penalty function $h_i(\cdot)$ needs to be a regularizer on $\mathcal{X}_i$:

Definition 5. Let $\mathcal{D}$ be a compact and convex subset of $\mathbf{R}^m$ (for some positive integer m). We say that $h_i : \mathcal{D} \to \mathbf{R}$ is a regularizer (with respect to some vector norm $\|\cdot\|$) if:

1) h_i is continuous.
2) h_i is strongly convex with respect to $\|\cdot\|$: there exists some $K > 0$ such that $\forall t \in [0,1], \forall \mathbf{d}, \mathbf{d}' \in \mathcal{D}$: $h_i(t\mathbf{d} + (1-t)\mathbf{d}') \leq t h_i(\mathbf{d}) + (1-t) h_i(\mathbf{d}') - \frac{1}{2} K t (1-t) \|\mathbf{d}' - \mathbf{d}\|^2$. In this case, we say that h_i is K-strongly convex (with respect to $\|\cdot\|$).

Consequently, the arg max in Step 4 of Algorithm 1 is well-defined since it is a maximization of a continuous function over a compact set (existence of a maximizer) and since the (continuous) function to be maximized over is strongly concave (uniqueness of the maximizer).

III. Variational Stability: A Key Criterion

In this section, we introduce the notion of variational stability, a key quantity that will relate the structural properties of the gradient function to the set of all Nash equilibria, and that will, ultimately, ensure the convergence of the online mirror descent.

A. Variational Stability

We start by defining variational stability and then relating it to an important concept in variational analysis: monotone operator.

Definition 6. Given a concave game $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$, a set $\mathcal{C} \subset \mathcal{X}$ is called variationally stable, if

$$\langle \mathbf{v}(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \triangleq \sum_{i=1}^{N} \langle v_i(\mathbf{x}), x_i - x_i^* \rangle \leq 0, \forall \mathbf{x} \in \mathcal{X}, \forall \mathbf{x}^* \in C,$$

with equality if and only if $\mathbf{x} \in \mathcal{C}$.

We emphasize that variationaly stability is related to and much weaker than an important concept in variational analysis. Specifically, $v(\cdot)$ is called a monotone operator [11] if the following holds:

$$\langle v(\mathbf{x}) - v(\tilde{\mathbf{x}}), \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq 0, \forall \mathbf{x}, \tilde{\mathbf{x}} \in \mathcal{X}, \tag{1}$$

with equality if and only if $\mathbf{x} = \tilde{\mathbf{x}}$. Let $\mathbf{x}^*$ be a Nash equilibrium (whose existence is guaranteed by Theorem 1). Per Remark 2, we have

$$\langle v(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle = \sum_{i=1}^{N} \langle v_i(\mathbf{x}^*), x_i - x_i^* \rangle \leq 0.$$

Consequently, by expanding Equation 1, it then follows that $\langle v(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \leq \langle v(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \leq 0$, where equality is achieved if and only if $\mathbf{x} = \mathbf{x}^*$. This suggests that when $v(\mathbf{x})$ is a monotone operator, there exists a unique Nash equilibrium and the singleton set of this unique Nash equilibrium is variationally stable. The converse is not true: when $v(\mathbf{x})$ is not a monotone operator, we can still have a variationally stable set $\mathcal{C}$ for the concave game.

Note also in the above definition, we referred to an element in $\mathcal{C}$ as $\mathbf{x}^*$, seemingly to suggest that such an element will be a Nash equilibrium. This is not a coincidence, as we explore in the next subsection.

B. Properties of Variational Stability

Here we study the structural properties of a variationally stable set. It turns out that a variationally stable set contains all Nash equilibria of the game (note that at least one Nash equilibrium exists per Theorem 1). Before proceeding, a word on the notation for the remainder of the paper: for convenience, we shall write $v_j(\mathbf{x})(x_j - x_j^*)$ to denote the inner product between $v_j(\mathbf{x})$ and $x_j - x_j^*$ in replacement of the more cumbersome notation $\langle v_j(\mathbf{x}), x_j - x_j^* \rangle$. Due to space limitation, we omit all the proofs in the subsection.

Lemma 1. *Give a concave game $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$. If $\mathcal{C}$ is a non-empty variationally stable set, then $\mathcal{C}$ is a closed and convex set of all Nash equilibria of the game.*

We can then define variationally stable games:

Definition 7. A concave game is called variationally stable if its set of Nash equilibria is a variationally stable set.

On the other hand, if $\mathbf{x}^*$ is the unique Nash equilibrium of the game (an important and useful special case), the set $\{\mathbf{x}^*\}$ is not necessarily a variational stable set. This means that in the singleton set case, variational stability is a stronger notion than that of the unique Nash equilibrium per Lemma 1. In general, however, the notion of variational stability is neither stronger nor weaker than that of the unique Nash equilibrium. The following lemma gives us a convenient sufficient condition ensuring that there exists a singleton set $\{\mathbf{x}^*\}$ that is variationally stable; in this case, to avoid notational clutter, we simply

say that $\mathbf{x}^*$ is variationally stable, although it should be kept in mind that variational stability always refers to a set.

Lemma 2. *Given a concave game $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$, where each u_i is twice continuously differentiable. For each $\mathbf{x} \in \mathcal{X}$, define the Hessian matrix $H(\mathbf{x})$ as follows:*

$$H_{ij}(\mathbf{x}) = \frac{1}{2} \nabla_{x_j} v_i(\mathbf{x}) + \frac{1}{2}(\nabla_{x_i} v_j(\mathbf{x}))^T. \tag{2}$$

If $H(\mathbf{x})$ is negative-definite for every $\mathbf{x} \in \mathcal{X}$, then the game admits a unique Nash equilibrium $\mathbf{x}^*$ that is variationally stable.

Remark 3. It is important to note that the Hessian matrix so defined is a block matrix: each $H_{ij}(\mathbf{x})$ is a matrix itself. Writing it in terms of the utility function, we have $H_{ij}(\mathbf{x}) = \frac{1}{2} \nabla_{x_j} \nabla_{x_i} u_i(\mathbf{x}) + \frac{1}{2}(\nabla_{x_i} \nabla_{x_j} u_j(\mathbf{x}))^T$. In particular, $H_{ij}(\mathbf{x})$ is a $m_i \times m_j$ matrix, where $\dim(x_i) = m_i, \dim(x_j) = m_j$. Also note that this sufficient condition is far more than necessary: it in fact implies $v(\cdot)$ is a monotone operator (see [12]) (and hence the game is variationally stable).

C. Examples

Here we give two important classes of games that satisfy the variational stability criterion. This is by no means a comprehensive list. Due to space limitation, we will only have a very limited discussion here.

1) **Potential Games** A game $\mathcal{G} = (\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$ is called a potential game [27] if there exists a potential function $V : \mathcal{X} \to \mathbf{R}$ such that $u_i(x_i, \mathbf{x}_{-i}) - u_i(\tilde{x}_i, \mathbf{x}_{-i}) = V(x_i, \mathbf{x}_{-i}) - V_i(\tilde{x}_i, \mathbf{x}_{-i}), \forall i \in \mathcal{N}, \forall \mathbf{x} \in \mathcal{X}, \forall \tilde{x}_i \in \mathcal{X}_i$. A potential game is called a convex potential game if the potential function $V(\cdot)$ is concave[3] Note that in a convex potential game, we have

$$H_{ij}(\mathbf{x}) = \frac{1}{2} \nabla_{x_j} v_i(\mathbf{x}) + \frac{1}{2}(\nabla_{x_i} v_j(\mathbf{x}))^T \tag{3}$$

$$= \frac{1}{2} \nabla_{x_j} \nabla_{x_i} V(\mathbf{x}) + \frac{1}{2}(\nabla_{x_i} \nabla_{x_j} V(\mathbf{x}))^T. \tag{4}$$

Consequently, $H^(\mathbf{x}) = \nabla^2 V$, which is negative semi-definite when V is concave. This implies that in a convex potential game, $\mathcal{C} = \arg\max_{\mathbf{x} \in \mathcal{X}} V(\mathbf{x})$ is variationally stable.

2) **Monotone Games and Pseudo-Monotone Games** Monotone games are introduced in [12] (called diagonally strict concave games there): it is a concave game satisfying $\langle v(\mathbf{x}) - v(\tilde{\mathbf{x}}), \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq 0, \forall \mathbf{x}, \tilde{\mathbf{x}} \in \mathcal{X}$. Namely, a monotone game is a concave game where the joint gradient is a monotone operator. Per the discussion in Section III-A, a monotone game is a variationally stable game.
The recent work [13] relaxed the monotone operator assumption and introduced a broader class of games called pseudo-monotone games. A pseudo-monotone game is a concave game satisfying: $\forall \mathbf{x}, \tilde{\mathbf{x}} \in \mathcal{X}$, if $\langle v(\tilde{\mathbf{x}}), \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq 0$, then $\langle v(\mathbf{x}), \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq 0$. To see that a pseudo-monotone game is variational stable, note that at a Nash equilibrium $\mathbf{x}^*$, we have $\langle v(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \leq 0$ per Remark 2. The definition of a pseudo-monotone game then immediately implies $\langle v(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \leq 0$, thereby establishing the conclusion.

[3]It is called convex potential game as opposed to concave potential game because in engineering, the utility is typically framed in terms of costs and convex costs correspond to concave utilities.

IV. Convergence of OMD to Nash Equilibria

In this section, we tackle the main problem of the paper and establish that the last iterate of OMD converges to Nash equilibria under variationaly stability.

A. Fenchel Coupling

We first construct a Lyapunov function called Fenchel coupling, that will play a indispensable role in establishing the convergence of the OMD dynamics. The Fenchel coupling can be viewed as measuring the distance between the primal decision variable $\mathbf{x}$ and the dual gradient variable $\mathbf{y}$ and is a generalization of Bregman divergence. As a reminder, we emphasize that for notational convenience, we denote the inner product by $x_i y_i$ (in replacement of $\langle x_i, y_i \rangle$).

Definition 8. Given a concave game $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$ and for each player i, let $h_i : \mathcal{X}_i \to \mathbf{R}$ be a regularizer with respect to the norm $\|\cdot\|_i$ that is K_i-strongly convex. Let $\mathcal{Y} = \prod_{i=1}^N \mathbf{R}^{m_i}$.

1) The convex conjugate function $h_i^* : \mathbf{R}^{m_i} \to \mathbf{R}$ of h_i is defined as:
$$h_i^*(y_i) = \max_{x_i \in \mathcal{X}_i} \{x_i y_i - h_i(x_i)\}, \forall y_i \in \mathbf{R}^{m_i}.$$
2) The choice function $C_i : \mathbf{R}^{m_i} \to \mathcal{X}_i$ associated with the regularizer h_i for player i is defined as:
$$C_i(y_i) = \arg\max_{x_i \in \mathcal{X}_i} \{x_i y_i - h_i(x_i)\}, \forall y_i \in \mathbf{R}^{m_i}.$$
3) The Fenchel coupling $F : \mathcal{X} \times \mathcal{Y} \to \mathbf{R}$ induced by the regularizers $\{h_i\}_{i=1}^N$ is defined as:
$$F(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N (h_i(x_i) - x_i y_i + h_i^*(y_i)), \forall \mathbf{x} \in \mathcal{X}, \forall \mathbf{y} \in \mathcal{Y}.$$

Remark 4. Two things worth noting is that first, although the domain of h_i is $\mathcal{X}_i \subset \mathbf{R}^{m_i}$, the domain of its conjugate h_i^* is $\mathbf{R}^{m_i}$. Second, the choice function C_i projects y_i from the gradient space to x_i in the decision space, and corresponds to Line 4 in Algorithm 1.

The two key properties of Fenchel that will be important in establishing the convergence of OMD are given next. The proof is long and tedious and is therefore omitted.

Lemma 3. *For each $i \in \{1, \ldots, N\}$, let $h_i : \mathcal{X}_i \to \mathbf{R}$ be a regularizer with respect to the norm $\|\cdot\|_i$ that is K_i-strongly convex. Then, $\forall \mathbf{x} \in \mathcal{X}, \forall \tilde{\mathbf{y}}, \mathbf{y} \in \mathcal{Y}$:*

1)

$$F(\mathbf{x}, \mathbf{y}) \geq \frac{1}{2} \sum_{i=1}^N K_i \|C_i(y_i) - x_i\|_i^2 \tag{5}$$

$$\geq \frac{1}{2} (\min_i K_i) \sum_{i=1}^N \|C_i(y_i) - x_i\|_i^2. \tag{6}$$

2)

$$F(\mathbf{x},\tilde{\mathbf{y}}) \leq F(\mathbf{x},\mathbf{y}) + \sum_{i=1}^{N}(\tilde{y}_i - y_i)(C_i(y_i) - x_i) + \tag{7}$$

$$\frac{1}{2}(\max_i \frac{1}{K_i}) \sum_{i=1}^{N}(\|\tilde{y}_i - y_i\|_i^*)^2, \tag{8}$$

where $\|\cdot\|_i^*$ *is the dual norm of* $\|\cdot\|_i$ *(i.e.* $\|y_i\|_i^* = \max_{\|x_i\|_i \leq 1} x_i y_i$*).*

Remark 5. Collecting each individual choice map into a vector, we obtain the aggregate choice map $C : \mathcal{Y} \to \mathcal{X}$, with $C(\mathbf{y}) = (C_1(y_1), \ldots, C_N(y_N))$. Since each space $\mathcal{X}_i$ is endowed with norm $\|\cdot\|_i$, we can define the induced aggregate norm $\|\cdot\|$ on the joint space $\mathcal{X}$ as follows: $\|\mathbf{x}\| = \sum_{i=1}^N \|x_i\|_i$, which can be easily verified to be a valid norm. We can also similarly define the aggregate dual norm: $\|\mathbf{y}\|^* = \sum_{i=1}^N \|y_i\|_i^*$. Henceforth, it shall be clear that the convergence in the joint space (e.g. $C(\mathbf{y}^t) \to \mathbf{x}$, $\mathbf{y}^t \to \mathbf{y}$) will be defined under the respective aggregate norm.

B. Convergence Analysis

We will primarily be focused on the case where the game admits a singleton variationaly stable set (and hence a necessarily unique Nash equilibrium), in which case the last iterate of OMD converges to the unique Nash equilibrium. We do so for two reasons: First, this is an important special case not only because many games arising in engineering applications have an unique Nash equilibrium, but also because it is not known whether the last iterate of OMD would converge to the unique Nash equilibrium even in this special case. Second, perhaps more importantly, the analysis for multiple Nash equilibria case is almost identical to the single Nash equilibrium case and admits a trivial generalization. Before proceeding, we identify an important class of choice maps that are regular in an intuitive sense:

Definition 9. The choice map $C(\cdot)$ is said to be Fenchel coupling conforming if $C(\mathbf{y}^t) \to \mathbf{x}$ implies $F(\mathbf{x},\mathbf{y}^t) \to 0$ as $t \to \infty$.

We are now ready to state our first main convergence result.

Theorem 2. *Given a concave game* $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$ *that admits* $\mathbf{x}^*$ *as the unique Nash equilibrium that is variationaly stable. Then if the following assumptions are satisfied:*

1) The step size sequence $\{\alpha^t\}_{t=0}^\infty$ *in Algorithm 1 is slowly vanishing;*

2) The choice map $C(\cdot)$ *is Fenchel coupling conforming;*

then the OMD iterate $\mathbf{x}^t$ *given in Algorithm 1 converges to* $\mathbf{x}^*$*, irrespective of the initial point* $\mathbf{x}^0$*.*

Remark 6. There are three main ingredients that together establish this theorem.

1) Let $B(\mathbf{x}^*,\epsilon) \triangleq \{\mathbf{x} \in \mathcal{X} \mid \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ be the open ball centered around $\mathbf{x}^*$ with radius ϵ, where the $\|\cdot\|$ is the aggregate norm induced by the individual norms $\{\|\cdot\|_i\}_{i=1}^N$. Then, for any $\epsilon > 0$ the iterate $\mathbf{x}^t$ will eventually enter $B(\mathbf{x}^*,\epsilon)$ and visit $B(\mathbf{x}^*,\epsilon)$ infinitely often, no matter what the initial point $\mathbf{x}^0$ is. Mathematically, the claim is that $\forall \epsilon > 0, \forall \mathbf{x}^0, |\{t \mid \mathbf{x}^t \in B(\mathbf{x}^*,\epsilon)\}| = \infty$.
2) Fix any $\delta > 0$ and consider the set $\tilde{B}(\mathbf{x}^*,\delta) \triangleq \{C(\mathbf{y}) \mid F(\mathbf{x}^*,\mathbf{y}) < \delta\}$. In other words, $\tilde{B}(\mathbf{x}^*,\delta)$ is some "neighborhood" of $\mathbf{x}^*$, which contains every $\mathbf{x}$ that is an image of some $\mathbf{y}$ (under the choice map $C(\cdot)$) that is within δ distance of $\mathbf{x}^*$ under the Fenchel coupling "metric". Although $F(\mathbf{x}^*,\mathbf{y})$ is not a metric, $\tilde{B}(\mathbf{x}^*,\delta)$ contains an open ball within it. Mathematically, the claim is that for any $\delta > 0$, $\exists \epsilon(\delta) > 0$ such that: $B(\mathbf{x}^*,\epsilon) \subset \tilde{B}(\mathbf{x}^*,\delta)$.
3) For any "neighborhood" $\tilde{B}(\mathbf{x}^*,\delta)$, after long enough iterations, if $\mathbf{x}^t$ ever enters $\tilde{B}(\mathbf{x}^*,\delta)$, it will be trapped inside $\tilde{B}(\mathbf{x}^*,\delta)$ thereafter. Mathematically, the claim is that for any $\delta > 0$, there exists a $T(\delta)$, such that for any $t \geq T(\delta)$, if $\mathbf{x}^t \in \tilde{B}(\mathbf{x}^*,\delta)$, then $\mathbf{x}^{\tilde{t}} \in \tilde{B}(\mathbf{x}^*,\delta), \forall \tilde{t} \geq t$.

Putting all three elements above together, we note that the significance of Claim 2 is that, since the iterate $\mathbf{x}^t$ will enter $B(\mathbf{x}^*,\epsilon)$ infinitely often (per Claim 1), $\mathbf{x}^t$ must enter $\tilde{B}(\mathbf{x}^*,\delta)$ infinitely often. It therefore follows that, per Claim 3, starting from iteration t, $\mathbf{x}^t$ will remain in $\tilde{B}(\mathbf{x}^*,\delta)$. Since this is true for any $\delta > 0$, we have $F(\mathbf{x}^*,\mathbf{y}^t) \to 0$ as $t \to \infty$. Per Statement 1 in Lemma 3, this leads to that $\|C(\mathbf{y}^t) - \mathbf{x}^*\| \to 0$ as $t \to \infty$, thereby establishing that $\mathbf{x}^t \to \mathbf{x}^*$ as $t \to 0$.

■

In fact, the result generalizes straightforwardly to a variationally stable set of Nash equilibria. The proof of the convergence to the set case is similar, provided that we redefine, in a standard way, every quantity that measures the distance between two points to the corresponding quantity that measures the distance between a point and a set (by taking the infimum over the distances between the point and a point in that set). We therefore have the following result.

Theorem 3. *Given a concave game* $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$ *that admits a variationaly stable set*[4] $\mathcal{C}$*. Then if the following assumptions are satisfied:*

1) The step size sequence $\{\alpha^t\}_{t=0}^\infty$ *in Algorithm 1 is slowly vanishing;*

2) The choice map $C(\cdot)$ *is Fenchel coupling conforming;*

then the OMD iterate $\mathbf{x}^t$ *given in Algorithm 1 converges to* $\mathcal{C}$*, irrespective of the initial point* $\mathbf{x}^0$*:* $\lim_{t\to\infty} \boldsymbol{dist}(\mathbf{x}^t,\mathcal{C}) = 0$*.*

Remark 7. In the multiple Nash equilibria case, we point out that Theorem 3 only says that the iterate converges to set of Nash equilibria (under the point-to-set distance metric). A priori, this does not imply that the iterate will converge to any given Nash equilibrium in that set. However, by a more refined analysis, one can show that the OMD iterate will indeed converge to some Nash equilibrium in the set of all Nash equilibria. We omit this discussion due to space limitation.

V. Online Mirror Descent under Noisy Feedback

The standard OMD as stated in Algorithm 1 is somewhat restricted in most applications. This is because in order to

[4]Recall that per Lemma 1, this variationaly stable set is necessarily a set of all Nash equilibria.

perform the update in Step 5 in Algorithm 1, player i needs to know the exact gradient v_i. This is not feasible in many cases for at least two reasonse. First, there can be noise associated with measuring/sensing the gradient in the underlying environment. Second, a player's utility $u_i(\mathbf{x})$ is typically the mean of a random quantity: $u_i(\mathbf{x}) = \mathbf{E}_\eta[f(\mathbf{x}, \eta)]$. Consequently, even if there is no measurement noise, the player only obtains a sample of the realized gradient $\nabla_{x_i} f(\mathbf{x}, \eta)$, which is stochastic.

A natural extension then is to generalize the OMD to handle such cases, where only a noisy estimate of the gradient (as opposed to the exact gradient) is needed. Algorithm 2 gives a formal description of this generalized version.

Algorithm 2 Online Mirror Descent under Noisy Feedback

Each player i chooses an initial Y_i^0.
for $t = 0, 1, 2, \ldots$ **do**
  for $i = 1, \ldots, N$ **do**
    $X_i^t = \arg\max_{X_i \in \mathcal{X}_i} \{\langle Y_i^t, X_i \rangle - h_i(X_i)\}$
    $Y_i^{t+1} = Y_i^t + \alpha^t \hat{v}_i(\mathbf{X}^t)$
  end for
end for

The main difference between Algorithm 2 and Algorithm 1 lies in Step 5, where a noisy estimate of the gradient is used. In addition, the iterates are capitalized to make explicit the fact that due to the noisy gradient used in Step 5, they are now random variables. Specifically, we have used the capital letters X_i^t and Y_i^t in Algorithm 2 because these iterates are now random variables as a result of the noisy gradients $\hat{v}_i$. Of course, in order for convergence to be guaranteed, $\hat{v}_i(\mathbf{X}^t)$ cannot be just any noisy perturbation of the gradient. Here we employ a rather standard model on the noisy gradient that is commonly seen in the optimization literature:

Assumption 1. Let $\mathcal{F}^t$ be the canonical filtration induced by the (random) iterates up to time t: $\mathbf{X}^1, \ldots, \mathbf{X}^t$. We assume:

1) The noisy gradients are conditionally unbiased:

$$\forall i \in \mathcal{N}, \forall t, \mathbf{E}[\hat{v}_i(\mathbf{X}^t) \mid \mathcal{F}^t] = v_i(\mathbf{X}^t), \text{a.s.}. \tag{9}$$

2) The noisy gradients are bounded in mean square:

$$\forall i \in \mathcal{N}, \forall t, \mathbf{E}[\|\hat{v}_i(\mathbf{X}^t)\|_2^2 \mid \mathcal{F}^t] \leq V, \text{a.s.}, \tag{10}$$

for some constant $V > 0$.

Finally, note that when there is only one player, Algorithm 1 exactly recovers the well-known optimization algorithm stochastic mirror descent [28, 29].

Our main result is that under the above-stated uncertainty model, OMD converges almost surely to the set of Nash equilibria. Again, we first start with the unique Nash equilibrium case.

Theorem 4. *Given a concave game $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$ that admits $\mathbf{x}^*$ as the unique Nash equilibrium that is variationaly stable. Then if the following assumptions are satisfied:*

1) Assumption 1 holds.
2) Each $v_i(\mathbf{x})$ is Lipschitz continuous in $\mathbf{x}$ on $\mathcal{X}$.
3) The step size sequence $\{\alpha^t\}_{t=0}^\infty$ in Algorithm 1 is slowly vanishing;
4) The choice map $C(\cdot)$ is Fenchel coupling conforming;

then the iterate $\mathbf{X}^t$ in the generalized OMD given in Algorithm 2 converges to $\mathbf{x}^*$ almost surely.

Proof Sketch: Due the stochastic gradient model in this case, the proof here will be very different from and more involved than that of Theorem 2. Due to space limitation, we will only provide a brief sketch that outlines the main ideas of the major steps.

1) Building on the proof to Theorem 2, and apply Martingale convergence theorems (both Doob's martingale convergence theorem and the law of large number for Martingale differences), we can establish that every neighborhood $B(\mathbf{x}^*, \epsilon)$ is recurrent: for every $\epsilon > 0$, $B(\mathbf{x}^*, \epsilon)$ will be visited infinitely often with probability 1.
2) We consider the continuous dynamics approximation of OMD:

$$\dot{\tilde{\mathbf{y}}} = v(\tilde{\mathbf{x}}), \tilde{\mathbf{x}} = C(\tilde{\mathbf{y}}), \tag{11}$$

and establish that the Fenchel coupling function can never increase and will decrease linearly for a certain interval of time before the continuous trajectory $\tilde{\mathbf{y}}(t)$ gets close to Nash equilibrium $\mathbf{x}^*$ (in distance measured by $F(\mathbf{x}^*, \tilde{\mathbf{y}}(t))$). Note that we have used $\tilde{\mathbf{y}}(t), \tilde{\mathbf{x}}(t)$ to denote the trajectories induced by the ODE given in Equation 11, in order to distinguish them from the discrete version. Further note that Assumption 2 ensures that a unique solution trajectory to this ODE exists.
3) We then consider an affine interpolation of the discrete trajectory $Y(t)$ generated by Algorithm 2 (i.e. connect consecutive iterates via a straight line). We show that via a path-by-path argument that this affine interpolation trajectory is an asymptotic pseudo-trajectory [30] of $\tilde{\mathbf{y}}(t)$:

$$\lim_{t \to \infty} \sup_{0 \leq h \leq T} \|Y(t+h) - \tilde{Y}(t+h)\|^* = 0, \forall T > 0, \textbf{a.s.},$$

where $\tilde{Y}(t+h)$ represents the solution trajectory to the ODE in Equation 11 at time $t+h$, given that it starts at $Y(t)$ at time t. This essentially means that these two trajectories, the affine interpolation of the discrete trajectory and the continuous trajectory induced by the OD, are close after long enough time.
4) Building on the previous point, one can then show that for time sufficiently large, Fenchel coupling is approximately the same whether one uses the affine interpolation of the discrete trajectory or the continuous one. But note that point 2) establishes that in the continuous case, the Fenchel coupling will not increase. Consequently, it implies that the affine interpolation trajectory will then stay in a certain neighborhood $B(\mathbf{x}^*, \epsilon)$. Note that this is true only after large enough time t, but point 1) ensures that the discrete trajectory visits $B(\mathbf{x}^*, \epsilon)$ infinitely often and will therefore be trapped inside starting from some large t_0 and onwards. Since this is true for any ϵ, the conclusion follows.

■

Again, the argument for the unique Nash equilibrium case generalizes straightforwardly to the multiple Nash equilibria case as follows:

Theorem 5. *Given a concave game* $(\mathcal{N}, \mathcal{X} = \prod_{i=1}^{N} \mathcal{X}_i, \{u_i\}_{i=1}^{N})$ *and let* $\mathcal{X}^*$ *be a variationally stable set (of all Nash equilibria). Then if the following assumptions are satisfied:*

1) *Assumption 1 holds.*
2) *Each* $v_i(\mathbf{x})$ *is Lipschitz continuous in* $\mathbf{x}$ *on* $\mathcal{X}$.
3) *The step size sequence* $\{\alpha^t\}_{t=0}^{\infty}$ *in Algorithm 1 is slowly vanishing;*
4) *The choice map* $C(\cdot)$ *is Fenchel coupling conforming;*

then the iterate $\mathbf{X}^t$ *in the generalized OMD given in Algorithm 2 converges to* $\mathcal{X}^*$ *almost surely:* $\lim_{t\to\infty} \boldsymbol{dist}(\mathbf{X}^t, \mathcal{X}^*) = 0, a.s.$.

VI. Conclusion and Future Work

We studied the problem of learning Nash equilibria via OMD, under both perfect gradient feedback and noisy gradient feedback. As we have briefly mentioned, concavity is not needed for convergence to hold: variational stability suffices. This raises the question if there exists a broader class of games (than variationally stable games) in which OMD still converges to Nash equilibria. Although we do not have a conclusive answer, we do conjecture that variational stability, due to its particularly light requirement on the attraction towards the equilibria, is the minimal assumption needed for achieving global convergence to Nash equilibria (irrespective of the initial condition). Finally, we note that noisy gradient feedback considered in this paper is but one type of imperfect feedback. Delay is another important type of imperfect feedback that frequently occurs (often in different ways) in the online learning setting [31, 32]. We leave that for future work.

References

[1] S. Shalev-Shwartz *et al.*, "Online learning and online convex optimization," *Foundations and Trends® in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2012.

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

[3] S. Shalev-Shwartz and Y. Singer, "Convex repeated games and Fenchel duality," in *Advances in Neural Information Processing Systems 19*. MIT Press, 2007, pp. 1265–1272.

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

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

[6] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge university press, 2006.

[7] S. Krichene, W. Krichene, R. Dong, and A. Bayen, "Convergence of heterogeneous distributed learning in stochastic routing games," in *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*. IEEE, 2015, pp. 480–487.

[8] M. Balandat, W. Krichene, C. Tomlin, and A. Bayen, "Minimizing regret on reflexive banach spaces and learning nash equilibria in continuous zero-sum games," *arXiv preprint arXiv:1606.01261*, 2016.

[9] K. Lam, W. Krichene, and A. Bayen, "On learning how players learn: estimation of learning dynamics in the routing game," in *Cyber-Physical Systems (ICCPS), 2016 ACM/IEEE 7th International Conference on*. IEEE, 2016, pp. 1–10.

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

[11] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, ser. A Series of Comprehensive Studies in Mathematics. Berlin: Springer-Verlag, 1998, vol. 317.

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

[13] M. Zhu and E. Frazzoli, "Distributed robust adaptive equilibrium computation for generalized convex games," *Automatica*, vol. 63, pp. 82–91, 2016.

[14] 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.

[15] L. Xiao, "Dual averaging methods for regularized stochastic learning and online optimization," *Journal of Machine Learning Research*, vol. 11, no. Oct, pp. 2543–2596, 2010.

[16] S. Perkins and D. S. Leslie, "Asynchronous stochastic approximation with differential inclusions," *Stochastic Systems*, vol. 2, no. 2, pp. 409–446, 2012.

[17] S. Perkins, P. Mertikopoulos, and D. S. Leslie, "Mixed-strategy learning with continuous action sets," vol. 62, no. 1, pp. 379–384, January 2017.

[18] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *4OR*, vol. 5, no. 3, pp. 173–210, September 2007.

[19] G. Scutari, F. Facchinei, D. P. Palomar, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory in multiuser communication systems," vol. 27, no. 3, pp. 35–49, May 2010.

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

[21] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge University Press, 2011. [Online]. Available: https://books.google.com/books?id=mvaUAwAAQBAJ

[22] Z. Zhou and N. Bambos, "Wireless communications games in fixed and random environments," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 1637–1642.

[23] Z. Zhou, N. Bambos, and P. Glynn, "Dynamics on linear influence network games under stochastic environments," in *International Conference on Decision and Game Theory for Security*. Springer International Publishing, 2016, pp. 114–126.

[24] A. Blum and Y. Mansour, "From external to internal regret," *Journal of Machine Learning Research*, vol. 8, pp. 1307–1324, 2007.

[25] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media, 2003.

[26] 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.

[27] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[28] A. Nedic and S. Lee, "On stochastic subgradient mirror-descent algorithm with weighted averaging," *SIAM Journal on Optimization*, vol. 24, no. 1, pp. 84–107, 2014.

[29] Z. Zhou, P. Mertikopoulos, N. Bambos, S. Boyd, and P. Glynn, "Stochastic mirror descent in variationally coherent optimization problems," in *Advances in Neural Information Processing Systems*, 2017.

[30] M. Benaïm, "Dynamics of stochastic approximation algorithms," in *Séminaire de Probabilités XXXIII*, ser. Lecture Notes in Mathematics, J. Azéma, M. Émery, M. Ledoux, and M. Yor, Eds. Springer Berlin Heidelberg, 1999, vol. 1709, pp. 1–68.

[31] P. Joulani, A. Gyorgy, and C. Szepesvári, "Online learning under delayed feedback," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 1453–1461.

[32] Z. Zhou, P. Mertikopoulos, N. Bambos, P. Glynn, and C. Tomlin, "Countering delayed feedback in multi-agent learning," in *Advances in Neural Information Processing Systems*, 2017.