Second-Order Mirror Descent: Convergence in Games Beyond Averaging and Discounting

Bolin Gao and Lacra Pavel

Abstract— In this paper, we propose a second-order extension of the continuous-time game-theoretic mirror descent (MD) dynamics, referred to as MD2, which provably converges to mere (but not necessarily strict) variationally stable states (VSS) without using common auxiliary techniques such as time-averaging or discounting. We show that MD2 enjoys no-regret as well as an exponential rate of convergence towards strong VSS upon a slight modification. MD2 can also be used to derive many novel continuous-time primal-space dynamics. We then use stochastic approximation techniques to provide a convergence guarantee of discrete-time MD2 with noisy observations towards interior mere VSS. Selected simulations are provided to illustrate our results.

I. INTRODUCTION

A central problem of multi-agent online learning is the design of adaptive policies which a set of subscribing agents (or players) can utilize to arrive at a desired collective outcome. These adaptive policies can typically be viewed as the following iterative process: an *information processing step*, whereby game-relevant information is made available to the player and then collected for processing, followed by a *decision step*, whereby the processed information is converted into the next strategy. A unified mathematical codification of this two-step process that emerged in recent years is referred to as *mirror descent*, first studied by [3] and subsequently by [4]–[11] in the context of optimization.

For *games*, mirror descent is often analyzed as a set of ordinary differential equations (ODEs), referred to as the continuous-time mirror descent dynamics (MD), which can be seen as the multi-agent extension of [3, p. 87]. MD is also referred to as dual averaging (DA) in continuous games [1], [19] and Follow-the-Regularized-Leader (FTRL) or exponential learning in finite games [18], [20], [21]. The core appeal of MD lies in its flexibility in adapting to a wide variety of problem settings as well as its rich theoretical properties. In particular, MD is known to converge to the Nash equilibrium (NE) in terms of its generated strategies under a trio of *strict* assumptions:

  (i) the game is *strictly monotone*, i.e., the pseudo-gradient of the game is a strictly monotone operator [50]. The class of strictly monotone games captures games that admit strictly concave potential functions, as well as saddle-point problems with strictly convex, strictly concave saddle functions, a special case of diagonally strict concavity of [51].
 (ii) the NE is *strictly variationally stable* (in the sense of [1]). The class of games with a strictly variationally stable NE (also known as *strictly variationally stable games* [17]) contains the class of strictly monotone games, as well as strictly coherent saddle point problems [46].
(iii) the NE is *strict* [14], which coincides with a locally strictly variationally stable NE in finite games [1].

Despite these theoretical guarantees, the applicability of MD and its variants remains limited, as many games found in practice do not conveniently exhibit these strict properties. In other words, MD does not converge in many games, most notably in zero-sum (ZS) finite games that feature a unique interior NE [21]. [12] showed that no two-player ZS finite game is strictly monotone. Furthermore, strictness is also intimately related to the uniqueness of the game equilibrium. For instance, a strictly monotone game admits a unique NE [50]. In practice, however, many games exhibit a convex set of equilibria as opposed to a unique one. This means strictness imposes serious constraints on the problem parameters or types. Moreover, many dynamics or algorithms that are directly derived from MD (through discretization or otherwise) share similar limitations in that some type of strictness needs to be assumed to ensure convergence, e.g., [1], [2], [16]–[20], [22].

To achieve convergence beyond strictness in a continuous-time setup, the existing literature typically relies on two primary approaches: *averaging* and *discounting*. The first method utilizes the time-averaged (*ergodic*) strategies generated by MD as opposed to the actual strategies [18]. However, this approach has two drawbacks. Firstly, the players need to combine their time-averaged strategies in order to recover the NE, which is unrealistic in non-cooperative settings. Secondly, time-averaging could fail outside of ZS games [57]. Another method is via a *discounting* procedure [15], [27], [28], which is conceptually related to the *weight decay* method of [41]. Discounting can be seen as a regularization technique that makes use of a strictly convex function to offset poor game properties, such as the lack of strict monotonicity. However, in general, discounting cannot yield exact convergence [15], [27], [28]. Since *averaging* and *discounting* represent two of the most widely studied method for improving the convergence properties of continuous-time MD, yet at the same time both have been met with challenges, therefore it is natural for us to set our sights on alternative approaches.

Contributions: In this work, in a game-theoretic setup, we propose a second-order variant of the continuous-time MD (the first-order of which was studied in [4], [5], [18]–[21]), which we refer to as MD2. *Second-order* means that the set of ODEs is second-order in *time*, thus MD2 can be seen as a dual-space formulation of the heavy-ball method [32] and is closely related to the class of nth order discounted game dynamics of [27], adapted here in a more general game setting. Like MD, we show that MD2 satisfies the basic assumption of *no-regret*. However, unlike MD, MD2 converges without using averaging even when the NE is not strictly variationally stable. Moreover, unlike the discounted MD [28], which partially overcomes the convergence issue of MD at the cost of inexact convergence, MD2 converges *exactly*. Exact convergence via a primal-space, second-order pseudo-gradient-type dynamics has been achieved in (merely) monotone games [26]. We provide a dual-space generalization of this result, beyond the monotone game setting, into the case where the equilibria are *merely variationally stable*. Furthermore, MD2 can exactly recover the unconstrained, primal-space dynamics of [26] in the full-information case. Finally, we use our continuous-time convergence results to guide the proof of convergence of a *discrete-time* MD2. In gist, we show that higher-order dynamics can converge towards Nash solutions with more general stability properties.

Related Works: This paper incorporates several ideas from the variational stability, higher-order dynamics, and stochastic (semi-

bandit) Nash-seeking literature. The concept of a variationally stable state (VSS) has roots in evolutionary game theory, and it is motivated by and analogous to that of an evolutionarily stable state (ESS) [37]. Namely, an ESS is a locally strict VSS for games with one or more populations of agents, also known as population games (similarly, a global ESS or GESS is a globally strict VSS) [12]. Many evolutionary game dynamics such as the replicator and projection dynamics have been shown to converge towards NEs of strictly stable games, which are GESS [2], [12]. In contrast to ESS, VSS is defined for the more general class of continuous games [1]. The literature on strict VSS seeking without structural property such as strict monotonicity is relatively recent [1], [16], [17], [19], [20]. [19] showed that DA converges towards a globally strict VSS. [16], [17] studied online gradient descent (discrete MD with projection map) where the feedback is received asynchronously between agents.

In contrast to a strict VSS, the class of *mere VSS* is motivated by the more general notion of a neurally stable state (NSS) and global NSS (GNSS) [38]. Comparatively speaking, the set of literature that deals with convergence towards a mere VSS is fewer, especially in the absence of structural assumptions such as the game being merely monotone (also known as stable game [12]). In a population game setup, [12] showed that the best response, integrable excess payoff and impartial pairwise comparison dynamics converge globally to the set of NEs in (merely) stable games, which are GNSS. With a few exceptions such as [46], the current set of literature on the (non-ergodic) convergence towards a mere VSS outside of a finite/population game setup typically also assumes monotonicity of the game [24], [26], [58]. In contrast to the above references, in this work we generalize the convergence behavior of MD to non-strict NE (i.e., a mere VSS) without any structural assumption, through the use of higher-order augmentation.

Higher-order dynamics was pioneered by Polyak in [32], whose algorithm later became widely known as the *heavy-ball method* (HB) [32]. HB was recognized as an analogue of a second-order ODE upon its inception and has since been heavily investigated in the optimization literature [39] and arguably forms the backbone of *deep learning*, where non-convexity and local minima dominate [40]. Inspired by [32], the authors of [34] studied a second-order dynamics for continuous concave games, which can be seen as a second-order extension of the projection dynamics of [36] and showed that their dynamics converge whenever there exists a Lipschitz and bounded potential function. In a finite game setup, [33] studied nth-order variant of exponential learning [18], whereby the payoff is processed via successive integration, and showed that these higher-order learning schemes can achieve faster rate of elimination of iteratively dominated solutions and convergence towards strict NE. An important connection between higher-order dynamics and the stability of a NE was established by [35], which showed that the addition of an *anticipatory* process to first-order gradient-play and replicator dynamics can result in local convergence towards an interior (non-strict) NE, despite the dynamics themselves being *uncoupled*. For continuous-time equilibrium seeking via a higher-order augmentation in the dual-space, the closest work related to ours is the second-order *discounted* exponential learning scheme in [27], which overcomes the problem with non-convergence towards interior NE of [33] at the cost of inexact convergence. It was found empirically in [27] that such higher-order augmentation improves the convergence property of its first-order counterpart and converges in non-strictly monotone games where the first-order fails. We improve [27] by generalizing these results from finite games to continuous games and go beyond a monotone game setup while simultaneously achieving exact convergence. Exact convergence (also known as *last iterate convergence*) via a primal-space, second-order pseudo-gradient dynamics was achieved in merely monotone games [26], which we also generalize, by stripping away monotonicity assumptions and showing that MD2 encapsulates the dynamics of [26] as a special case.

In the discrete-time, semi-bandit setup, whereby each player receives a (possibly) noise-corrupted version of the pseudo-gradient, the closest work related to ours is [1]. [1] studied the first-order MD, known as DA therein, and showed that under imperfect gradient setup, DA converges to a globally variationally stable NE under diminishing step-sizes and its ergodic average converges to the set of NE in 2-player ZS games. Several variations to DA have been studied in the semi-bandit literature, whereby their convergence is towards either strict (strong) VSS or strict NE. For instance, in finite games, [23] studied a Hedge-variant of exponential weight algorithm (HEDGE) and provided convergence when the NE is strongly VS with respect to the L^1 norm. In a similar setup, [15] studied a discounted variant of HEDGE for potential games. A similar analysis was also performed by [22] for FTRL in finite games. [25] provided several primal-space algorithms for merely monotone games. [46] studied the optimistic MD algorithm, which converges in coherent games, i.e., two-player games that possess mere VSS. Compared to optimistic MD, discrete-time MD2 only requires one mirror projection instead of two prox projections. Furthermore, optimistic MD cannot converge in the presence of noise [46] unless the game is strictly coherent (a stronger assumption), whereas the discrete-time MD2 can converge under noisy conditions.

Organization: We provide the background materials in Section II. Section III reviews several basic properties of the first-order MD. We discuss the convergence properties of MD2 in Section IV and provide the rate of convergence as well as regret minimization guarantee in Section V. We derive associated primal-space dynamics in Section VI. Section VII discusses MD2 in discrete-time with noisy observations. Simulations are presented in Section VIII. Section IX presents our conclusions. For readability, all proofs are relegated to Section X. A table of main notations used in this paper (Table 1) is provided in the Appendix.

II. BACKGROUND

The following material for convexity and duality are drawn from [48], [62]. Those for game theory are drawn from [49], [50], [59].

A. Sets and Vectors

Given a convex set $\mathcal{C} \subseteq \mathbb{R}^n$, the (relative) interior of the set is denoted as $(\text{rint}(\mathcal{C}))$ $\text{int}(\mathcal{C})$. $\text{rint}(\mathcal{C})$ coincides with $\text{int}(\mathcal{C})$ whenever $\text{int}(\mathcal{C})$ is non-empty. The closure of a set $\mathcal{C}$ is denoted as $\text{cl}(\mathcal{C})$. We denote the non-negative orthant of $\mathbb{R}^n$ as $\mathbb{R}^n_{\geq 0}$, and the positive orthant as $\mathbb{R}^n_{>0}$. A column vector in $\mathbb{R}^n$ is denoted as $x = (x_1, \ldots, x_n) = \left[x_1, \ldots, x_n \right]^\top$. $\mathbf{1}$ and $\mathbf{0}$ denote the column vector all ones and all zeros. $I_{n \times n}$ and $O_{n \times n}$ denote the $n \times n$ identity and zero matrices. Subscript is omitted when the dimensionality is unambiguous. Suppose n is a natural number, then $[n] := \{1, \ldots, n\}$.

B. Convex Functions and Duality

Let $\mathcal{M} = \mathbb{R}^n$ be endowed with norm $\|\cdot\|$ and dot product $\langle \cdot, \cdot \rangle$. An extended real-valued function is a function f that maps from $\mathcal{M}$ to $[-\infty, \infty]$. The (effective) domain of f is $\text{dom}(f) = \{x \in \mathcal{M} : f(x) < \infty\}$. A function $f : \mathcal{M} \to [-\infty, \infty]$ is proper if $f(x) \neq -\infty, \forall x$ and there exists at least one $x \in \mathcal{M}$ such that $f(x) < \infty$; f is closed if its epigraph is closed. A function $f : \mathcal{M} \to [-\infty, \infty]$ is supercoercive if $\lim_{\|x\| \to \infty} f(x)/\|x\| \to \infty$. The indicator function over $\mathcal{C}$ is denoted by $\delta_{\mathcal{C}}$. The normal cone of $\mathcal{C}$ is defined as, $N_{\mathcal{C}}(x) = \{v \in \mathbb{R}^n | v^\top (y - x) \leq 0, \forall y \in \mathcal{C}\}$. The tangent cone of a non-empty

convex set $\mathcal{C}$ at $x \in \mathcal{C}$ is $T_{\mathcal{C}}(x) = \overline{\bigcup_{\lambda \in \mathbb{R}_{>0}} \lambda(\mathcal{C} - x)}$ and equals $\emptyset$ for all $x \notin \mathcal{C}$. Id denotes the identity function.

Recall that $\pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|y - x\|_2^2$ is the Euclidean projection of x onto $\mathcal{C}$, where we refer to $\pi_{\mathcal{C}}$ as the Euclidean projector. Let $\partial f(x)$ denote a the set of subgradients of f at x and $\nabla f(x)$ the gradient of f at x. For differentiation on the boundary of a closed set $\mathcal{C}$, in lieu of the subgradient, we can also assume f is defined and differentiable on an open set containing $\mathcal{C}$. Given f, the function $f^{\star} : \mathcal{M}^{\star} \to [-\infty, \infty]$ defined by $f^{\star}(z) = \sup_{x \in \mathcal{M}} [x^{\top} z - f(x)]$, is called the conjugate function of f, where $\mathcal{M}^{\star}$ is the dual space of $\mathcal{M}$, endowed with the dual norm $\|\cdot\|_{\star}$. $f^{\star}$ is closed and convex if f is proper. By the conjugate subgradient theorem [48], suppose $f : \mathbb{R}^n \to (-\infty, +\infty]$ is proper, closed and convex and $f^{\star}$ is its Fenchel conjugate, then for any $x, z \in \mathbb{R}^n$, $x^{\top} z = f(x) + f^{\star}(z) \iff z \in \partial f(x) \iff x \in \partial f^{\star}(z)$. The Bregman divergence of a proper, closed, convex function f is $D_f : \text{dom}(f) \times \text{dom}(\partial f) \to \mathbb{R}$, $D_f(x, y) = f(x) - f(y) - g^{\top}(x - y), g \in \partial f(y)$. Given a vector-valued function F, the Jacobian of F is denoted as $\mathbf{J}_F$.

C. N-Player Concave Games

Let $\mathcal{G} = (\mathcal{N}, \{\Omega^p\}_{p \in \mathcal{N}}, \{\mathcal{U}^p\}_{p \in \mathcal{N}})$ be a game, where $\mathcal{N} = \{1, \dots, N\}$ is the set of players, $\Omega^p \subseteq \mathbb{R}^{n_p}$ is the set of player p's strategies (actions). We denote the strategy (action) set of player p's opponents as $\Omega^{-p} \subseteq \prod_{q \in \mathcal{N}, q \neq p} \mathbb{R}^{n_q}$. We denote the set of all the players strategies as $\Omega = \prod_{p \in \mathcal{N}} \Omega^p \subseteq \prod_{p \in \mathcal{N}} \mathbb{R}^{n_p} = \mathbb{R}^n, n = \sum_{p \in \mathcal{N}} n_p$. We refer to $\mathcal{U}^p : \Omega \to \mathbb{R}, x \mapsto \mathcal{U}^p(x)$ as player p's real-valued payoff function, where $x = (x^p)_{p \in \mathcal{N}} \in \Omega$ is the action profile of all players, and $x^p \in \Omega^p$ is the action of player p. We also denote x as $x = (x^p; x^{-p})$ where $x^{-p} \in \Omega^{-p}$ is the action profile of all players except p. For differentiability purposes, we make the implicit assumption that there exists some open set, on which $\mathcal{U}^p$ is defined and continuously differentiable, such that it contains Ω.

Assumption 1. For all $p \in \mathcal{N}$, Ω^p is a non-empty, closed, convex, subset of $\mathbb{R}^{n_p}$. $\mathcal{U}^p(x^p; x^{-p})$ is (jointly) continuous in $x = (x^p; x^{-p})$. $\mathcal{U}^p(x^p; x^{-p})$ is concave and continuously differentiable ($\mathcal{C}^1$) in x^p for all $x^{-p} \in \Omega^{-p}$.

Under Assumption 1, we refer to $\mathcal{G}$ as a *(continuous) concave game*. Given $x^{-p} \in \Omega^p$, each agent $p \in \mathcal{N}$ aims to find the solution of the following optimization problem,

$$\underset{x^p}{\text{maximize}} \quad \mathcal{U}^p(x^p; x^{-p}) \quad \text{subject to} \quad x^p \in \Omega^p. \quad (1)$$

A profile $x^{\star} = (x^{p\star})_{p \in \mathcal{N}}$ is a Nash equilibrium (NE) if,

$$\mathcal{U}^p(x^{p\star}; x^{-p\star}) \geq \mathcal{U}^p(x^p; x^{-p\star}), \forall x^p \in \Omega^p, \forall p \in \mathcal{N}. \quad (2)$$

At a NE, no player can increase its payoff by unilateral deviation.

Remark 1. Under Assumption 1, existence of a NE is guaranteed for bounded Ω^p [49, Theorem 4.4]. When Ω^p is closed but not bounded, the existence of a NE is guaranteed under the additional assumption that $-\mathcal{U}^p$ is coercive in x^p, that is, $\lim_{\|x^p\| \to \infty} -\mathcal{U}^p(x^p; x^{-p}) = +\infty$, for all $x^{-p} \in \Omega^{-p}, p \in \mathcal{N}$ [49, Corollary 4.2]. A NE is said to be *strict* when the inequality in (2) is strict [14].

A useful characterization of a NE of a concave game $\mathcal{G}$ is given in terms of the *pseudo-gradient* [51], which is defined as,

$$U : \Omega \to \mathbb{R}^n, U(x) = (U^p(x))_{p \in \mathcal{N}}, \quad (3)$$

where $U^p(x) = \nabla_{x^p} \mathcal{U}^p(x^p; x^{-p})$ is the *partial-gradient*. By [50, Proposition 1.4.2], $x^{\star} \in \Omega$ is a NE iff,

$$(x - x^{\star})^{\top} U(x^{\star}) \leq 0, \forall x \in \Omega. \quad (4)$$

Equivalently, $x^{\star}$ is a solution of the Stampacchia variational inequality VI$(\Omega, -U)$, [50]. Following [2], we say that a NE is globally strict if the inequality of (4) is held strictly for all $x^{\star} \neq x$.

D. Monotonicity

A general class of games in which many dynamics are guaranteed to converge is the class of monotone games, also known as stable games [12], [13] or dissipative games [44]. We contrast some known definitions associated with monotone games in the literature.

Definition 1. The game $\mathcal{G}$ is,
  (i) η-strongly monotone if $(U(x) - U(x'))^{\top}(x - x') \leq -\eta \|x - x'\|_2^2$, $\forall x, x' \in \Omega$, for some $\eta > 0$.
  (ii) strictly monotone if $(U(x) - U(x'))^{\top}(x - x') < 0, \forall x \in \Omega \backslash \{x'\}$, with equality if and only if $x = x'$.
  (iii) (merely) monotone if $(U(x) - U(x'))^{\top}(x - x') \leq 0, \forall x, x' \in \Omega$.
  (iv) μ-weakly monotone if $(U(x) - U(x'))^{\top}(x - x') \leq \mu \|x - x'\|_2^2$, $\forall x, x' \in \Omega$, for some $\mu > 0$.

Strongly and strictly monotone games have been extensively investigated in the literature, at least dating from [51]. Merely monotone games have been studied in [12], [24]–[28], [58]. Weakly monotone games were considered in [27], [30], [43]. When U is $\mathcal{C}^1$, there exists a natural characterization in terms of definiteness of its Jacobian [50, p. 155, Prop. 2.3.2]. Note that strictly monotone games can have at most one NE, whereas NEs can form a non-singleton convex set in merely monotone games [50].

E. Variational Stability

Although many classical examples of games satisfy monotonicity properties [12], these conditions may not hold in more complex scenarios. A recent line of research has started to relax monotonicity notions from a game to that of an equilibrium via the notion of a *variationally stable* (VS) equilibrium or state [1].

Definition 2. A Nash equilibrium $x^{\star} \in \Omega$ is,
  (i) η-strongly VS if $U(x)^{\top}(x - x^{\star}) \leq -\eta \|x - x^{\star}\|_2^2, \forall x \in \Omega$, for some $\eta > 0$.
  (ii) strictly VS if $U(x)^{\top}(x - x^{\star}) \leq 0, \forall x \in \Omega$, with equality if and only if $x = x^{\star}$.
  (iii) (merely) VS if $U(x)^{\top}(x - x^{\star}) \leq 0, \forall x \in \Omega$.
  (iv) μ-weakly VS if $U(x)^{\top}(x - x^{\star}) \leq \mu \|x - x^{\star}\|_2^2, \forall x \in \Omega$, for some $\mu > 0$.

If a condition (i) - (iv) hold on $\mathcal{D} \subset \Omega$, then the associated definition is said to hold *locally*. Otherwise the definition is said to hold globally.

Remark 2. We refer to $x^{\star}$ as a globally strong/strict/mere/weak *variationally stable state* (VSS) if it satisfies one of the corresponding VS notions in Definition 2 on all of Ω; $x^{\star}$ is a locally strong/strict/mere/weak VSS otherwise. For practical reasons, we will informally refer μ-weak VSS with a small μ as a *nearly* mere VSS, i.e., a weak VSS within a small distance away from becoming a mere VSS.

Globally mere VSS are the solutions to Minty variational inequality [50]. In a population game context, the set of globally mere VSS is called GNSS and (the unique) globally strict VSS is called GESS [2], [12], [13]. Strict VSS was extended to a local/global set-wise definition in [1]. [16] introduced a slight variation of strict VSS called λ-VS. [47] studied a local version of strict VSS under the name *locally asymptotically stable differential NE*, which requires twice continuous-differentiability of the payoff functions. Strong VSS (Definition 2(i)) was also studied in [1]. [23] studied a variant of the strong VSS with L^1 norm.

It can be seen that any NE of the strongly/strictly/merely/weak monotone game is a strong/strict/mere/weak VSS. In particular, the

non-empty set of NE coincides with the set of mere VSS for merely monotone concave games [2, Theorem 2]. When $\text{int}(\Omega) \neq \emptyset$ and $x^\star$ is an interior NE, i.e., $U(x^\star) = \mathbf{0}$, the condition for VS recovers the condition associated with monotone at $x' = x^\star$. Hence VS can be seen as a type of point-wise monotonicity and it is known to have a second-order characterization similar to that for monotone games, but specifically at the NE, see [1].

Recall that a NE $x^\star$ is *interior* if it lies in the interior of Ω, that is, $x^\star \in \text{rint}(\Omega)$. Throughout this work, we make the following assumption.

Assumption 2. $\mathcal{G}$ admits an interior mere VSS.

F. Second-Order Characterization of VSS

In practice, outside of monotone games, it is often difficult to verify that a NE is a particular type of VSS directly through Definition 2. A more common approach to characterize the VSS conditions of a solution is by looking at the *symmetric game Jacobian* whenever the pseudo-gradient U is $\mathcal{C}^1$, which is given by,

$$\mathfrak{J}_U(x) := \frac{1}{2}(\mathbf{J}_U(x) + \mathbf{J}_U^\top(x)), x \in \Omega. \quad (5)$$

The matrix $\mathfrak{J}_U(x) \in \mathbb{R}^{n \times n}$ is symmetric and thus guaranteed to have real eigenvalues, which makes it amenable to analysis.

We now provide several sufficient conditions for verifying whether a VSS $x^\star$ is strict, mere or weak. This is performed by checking the definiteness of $\mathfrak{J}_U(x)$ or $\mathfrak{J}_U(x^\star)$. Our condition for strict VSS is the same as that in [1]. Before proceeding, we say that the (symmetric) game Jacobian $\mathfrak{J}_U(x) \in \mathbb{R}^{n \times n}$ (5) is negative definite on $T_\Omega(x)$ if,

$$y^\top \mathfrak{J}_U(x) y < 0, \forall x \in \Omega, \forall y \in T_\Omega(x), y \neq \mathbf{0}, \quad (6)$$

and negative semi-definite on $T_\Omega(x)$ if the preceding inequality is non-strict. We use the shorthand notations $\mathfrak{J}_U(x) \prec O$ and $\mathfrak{J}_U(x) \preceq O$ respectively. For $\mu > 0$, $\mathfrak{J}_U(x) - \mu I \prec O$ is written as $\mathfrak{J}_U(x) \preceq \mu I$. Similar conventions for when $x = x^\star$.

Proposition 1. *Let $x^\star \in \Omega$ be a NE of $\mathcal{G}$. Suppose U is continuously differentiable, and,*

(i)   *$\mathfrak{J}_U(x) \prec O$ on $T_\Omega(x), \forall x \in \Omega$, then $x^\star$ is globally strictly VS and isolated.*

(ii) *$\mathfrak{J}_U(x) \preceq O$ on $T_\Omega(x), \forall x \in \Omega$, then $x^\star$ is globally merely VS.*

(iii) *$\mathfrak{J}_U(x) \preceq \mu I$ on $T_\Omega(x), \forall x \in \Omega$, then $x^\star$ is globally μ-weakly VS.*

Suppose instead,

(i')   *$\mathfrak{J}_U(x^\star) \prec O$ on $T_\Omega(x^\star)$, then $x^\star$ is locally strictly VS and isolated.*

(ii') *$\mathfrak{J}_U(x^\star) \preceq O$ on $T_\Omega(x^\star)$, then $x^\star$ is locally merely VS.*

(iii') *$\mathfrak{J}_U(x^\star) \preceq \mu I$ on $T_\Omega(x^\star)$, then $x^\star$ is locally μ-weakly VS.*

Remark 3. These conditions can be verified by calculating the maximum eigenvalue of $\mathfrak{J}_U(x)$ (resp. $\mathfrak{J}_U(x^\star)$), where $\lambda_{\max}(\mathfrak{J}_U(x))$ (resp. $\lambda_{\max}(\mathfrak{J}_U(x^\star))$) denotes the (real) eigenvalue with the largest magnitude associated with an eigenvector in $T_\Omega(x)$ (resp. $T_\Omega(x^\star)$). When $\mathbf{J}_U(x)$ (resp. $\mathbf{J}_U(x^\star)$) is symmetric, then analysis can be performed directly on $\mathbf{J}_U$ without resorting to calculating $\mathfrak{J}_U$.

Remark 4. Note that, even if $x^\star$ is shown to be μ-weak through Proposition 1(iii) or (iii'), it does not preclude the possibility that $x^\star$ could in fact be strict or mere due to the looseness of the bound in the proof of Proposition 1. Furthermore, unlike strict VSS, mere VSS need not be isolated. The next examples illustrate various notions of variational stability.

Example 1. (**Monotone game with a mere VSS**) Every NE of a merely monotone game is globally mere VSS. Perhaps the simplest example of a merely monotone game is the so-called bilinear

saddle point problem, whereby, suppose we have a saddle function, $f(x^1, x^2) = x^1 x^2$, $x^p \in \mathbb{R}$ for which we want to minimize in x^1 and maximize in x^2. We can cast this saddle point problem as a game by utilizing two payoff functions $\mathcal{U}^1(x^1; x^2) = -x^1 x^2$ and $\mathcal{U}^2(x^2; x^1) = x^1 x^2$. This game has a pseudo-gradient of $U(x) = (-x^2, x^1)$ and a game Jacobian $\mathfrak{J}_U(x) = O$. By Proposition 1(ii), the unique interior NE $x^\star = (0, 0)$ is the globally mere VSS.

Example 2. (**Non-monotone potential game with a mere VSS**) This example shows that the mere monotonicity enjoyed by Example 1 can be destroyed by the addition of another player, even when all the player's payoff functions remain linear in its own argument. Considered a three-player game, whereby,

$$\mathcal{U}^p(x^p; x^{-p}) = -x^p \prod_{q \neq p} x^q, \quad p \in \{1, 2, 3\}, \quad (7)$$

and each $x^p \in [-1, 1]$, i.e., $\Omega = [-1, 1]^3$. For this game, the pseudo-gradient is $U(x) = -(x^2 x^3, x^1 x^3, x^1 x^2)$ which shows that there exists an interior NE at $x^\star = (0, 0, 0)$. The Jacobian, which is symmetric, is,

$$\mathbf{J}_U(x) = -\begin{bmatrix} 0 & x^3 & x^2 \\ x^3 & 0 & x^1 \\ x^2 & x^1 & 0 \end{bmatrix},$$

hence this game is non-monotone in general and in fact possesses a non-convex and non-concave *potential function* $P(x) = -x^1 x^2 x^3$, i.e., $\nabla P = U$. Since $\mathbf{J}_U(x^\star) = O$ is negative semi-definite, hence by Proposition 1(ii'), $x^\star$ is locally merely VS. Note that if we restrict the strategy set to $\Omega = [0, 1]^3$ instead, every NE $x^\star \in \{x \in \Omega | x^p = x^q = 0, p, q \in \mathcal{N}, p \neq q\}$ is a globally mere VSS, as $U(x)^\top(x - x^\star) = -3x^1 x^2 x^3 \leq 0, \forall x \in [0, 1]^3$.

Example 3. (**RPS with non-negative payoff for ties**) Consider a two-player Rock-Paper-Scissors (RPS) game with $\mathcal{A}$ and $\mathcal{B}$ being the payoff matrices for player 1 and 2 respectively,

$$\mathcal{A} = \begin{bmatrix} \varsigma & -\ell & w \\ w & \varsigma & -\ell \\ -\ell & w & \varsigma \end{bmatrix}, \mathcal{B} = \mathcal{A}^\top, \quad (8)$$

where $\ell, w \geq 0$ are the values associated with a loss or a win and $\varsigma \in \mathbb{R}$ is the payoff of a tie. The strategy set associated with this game is the simplex $\Omega^p = \{x^p \in \mathbb{R}^{n_p} | \sum_{i=1}^{n_p} x_i^p = 1, x_i^p \geq 0\}$. The pseudo-gradient (as known as payoff vector [12], [13]) and the game Jacobian are as follows,

$$U(x) = \begin{bmatrix} O & \mathcal{A} \\ \mathcal{A} & O \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \quad \mathfrak{J}_U(x) = \frac{1}{2} \begin{bmatrix} O & \mathcal{A} + \mathcal{A}^\top \\ \mathcal{A} + \mathcal{A}^\top & O \end{bmatrix}. \quad (9)$$

To simplify our analysis, consider an example where $l = 0$. In this game, $x^\star = (x^{p\star})_{p \in \mathcal{N}}, x^{p\star} = (1/3, 1/3, 1/3)$ is the unique interior NE for any $\varsigma \in [0, w), w > 0$. The eigenvalues of $\mathfrak{J}_U(x)$ on $T_\Omega(x)$ are $\{-(\varsigma + w), \pm(\varsigma - w/2), \pm(\varsigma - w/2)\}$, hence this game is μ-weakly merely monotone with $\mu = \varsigma - w/2$ whenever $\varsigma \in (w/2, w)$. When $\varsigma = w/2$, the game is merely monotone (Definition 1). Taken together, this means for $\varsigma \in (w/2, w)$, $x^\star$ is a unique, locally μ-weak VSS where $\mu = \varsigma - w/2$, whereas for $\varsigma = w/2$, $x^\star$ is a unique, globally mere VSS. Finally, note that when $\ell = w = \varsigma = 0$ (and hence both $\mathcal{A}, \mathcal{B}$ are zero matrices), every strategy is globally merely VS and forms a convex set (the strategy set itself).

III. REVIEW OF FIRST-ORDER MD DYNAMICS

We now describe the dynamic process that a group of agents utilizes in order to arrive at one of the equilibrium notions discussed in previous sections. One such general model is the mirror descent (MD) dynamics. Intuitively, the family of MD dynamics ascribes an abstract behavior model to each player, which states that, upon receiving the partial-gradient of its payoff function, each player processes

the partial-gradient information (typically via an aggregation), then converts the processed information into the next strategy. This process can be described by the following set of ODEs [18]–[21],

$$\dot{z}^p = \gamma \nabla_{x^p} \mathcal{U}^p(x^p; x^{-p}) = \gamma U^p(x), \quad x^p = C_\epsilon^p(z^p), \quad \text{(MD)}$$

where z^p is referred to as score or dual variable, $\gamma > 0$ is a rate parameter, and $C_\epsilon^p : \mathbb{R}^{n_p} \to \Omega^p$ is referred to as the *mirror map*,

$$C_\epsilon^p(z^p) = \text{argmax}_{y^p \in \Omega^p} \left[y^{p\top} z^p - \epsilon \vartheta^p(y^p) \right], \epsilon > 0. \quad (10)$$

Here, $\vartheta^p : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{\infty\}$ is assumed to be a closed, proper and (at least) strictly convex function, referred to as a *regularizer*, where $\text{dom}(\vartheta^p)$ is assumed be a non-empty, closed and convex set, which agrees with the strategy set Ω^p, $\epsilon > 0$ is referred to as the regularization constant. The regularizer often satisfies one of the following distinct assumptions that ensure the existence of a unique solution of MD and (10):

Assumption 3. $\vartheta^p : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{\infty\}$ is closed, proper, convex, with $\text{dom}(\vartheta^p) = \Omega^p$ non-empty, closed and convex. In addition, ϑ^p is, (i) Legendre (i.e., strictly convex, steep, $\text{int}(\text{dom}(\vartheta^p)) \neq \emptyset$) and supercoercive, or (ii) ρ-strongly convex, $\rho > 0$.

We note that for ϑ^p to be *steep*, it means that $\|\nabla \vartheta^p(x_k^p)\|_\star \to +\infty$, $\|\cdot\|_\star$ is the dual norm, whenever $\{x_k^p\}_{k=1}^\infty$ is a sequence in $\text{rint}(\text{dom}(\vartheta^p)) = \text{rint}(\Omega^p)$ converging to a point in the (relative) boundary. Let $\psi_\epsilon^{p\star}$ be the convex conjugate of ψ_ϵ^p, $\psi_\epsilon^p := \epsilon \vartheta^p$. Then by Lemma 3 and Lemma 2 (in the Appendix), under Assumption 3, $C_\epsilon^p = \nabla \psi_\epsilon^{p\star}$. When ϑ^p is steep, C_ϵ^p maps from $\mathbb{R}^{n_p}$ to all values in Ω^p except those at the boundary; C_ϵ^p could map to boundary points otherwise. We refer to C_ϵ^p as the *mirror map induced by ψ_ϵ^p* and *Legendre* or *strongly-induced* (by ψ_ϵ^p) when ϑ^p satisfies Assumption 2(i) or (ii). Note that the notions of Legendre and strongly convex need not be dichotomous. It is possible for a C_ϵ^p to be both Legendre and strongly-induced. The following examples capture the three types of mirror maps that are extensively used in the literature [4], [5], [15], [17]–[21], [27], [28].

Example 4. Let $\Omega^p = \mathbb{R}^{n_p}$, $\vartheta^p(x^p) = \frac{1}{2\epsilon}\|x^p\|_2^2$, hence it satisfies Assumption 3(i), (ii). In this case, $C_\epsilon^p = Id$ and it is both Legendre and strongly-induced.

Example 5. Let $\Omega^p = \{x^p \in \mathbb{R}^{n_p} | \sum_{i=1}^{n_p} x_i^p = 1, x_i^p \geq 0\}$ (the unit simplex), $\vartheta^p(x^p) = \sum_{i=1}^{n_p} x_i^p \log(x_i^p)$, where we assume $0 \log(0) = 0$. ϑ^p can be shown to be 1-strongly convex in $\|\cdot\|_2$ (or $\|\cdot\|_1$), steep, hence it satisfies Assumption 3(ii). $C_\epsilon^p(z^p) = (\exp(\epsilon^{-1} z_i^p)(\sum_{j=1}^{n_p} \exp(\epsilon^{-1} z_j^p))^{-1})_{i \in [n_p]}$ is the *softmax* function.

Example 6. Let $\Omega^p \subset \mathbb{R}^{n_p}$ be a non-empty, convex, compact set, assume $\vartheta^p(x^p) = \frac{1}{2}\|x^p\|_2^2$, hence ϑ^p satisfies Assumption 3(ii) and is non-steep. Here, $C_\epsilon^p = \pi_{\Omega^p}$ is the Euclidean projection onto Ω^p.

We can represent MD in a more compact stacked notation,

$$\dot{z} = \gamma U(x), \quad x = C_\epsilon(z), \quad (11)$$

where $x = (x^p)_{p \in \mathcal{N}}, z = (z^p)_{p \in \mathcal{N}}, C_\epsilon = (C_\epsilon^p)_{p \in \mathcal{N}}, U = (U^p)_{p \in \mathcal{N}}$. Observe that at rest, MD (11) satisfies,

$$\mathbf{0} = U(\bar{x}), \quad \bar{x} = C_\epsilon(\bar{z}), \quad \bar{z} \in C_\epsilon^{-1}(\bar{x}), \quad (12)$$

where $\bar{x} = (\bar{x}^p)_{p \in \mathcal{N}}, \bar{z} = (\bar{z}^p)_{p \in \mathcal{N}}$ are the rest points of MD and $C_\epsilon^{-1} = ((C_\epsilon^p)^{-1})_{p \in \mathcal{N}} = (\partial \psi_\epsilon^p)_{p \in \mathcal{N}}$ is the inverse (or the pre-image) of C_ϵ. The rest condition (12) implies that $\bar{x}$ is an interior NE. Hence if a trajectory $z(t)$ of MD comes to a rest, $x(t) = C_\epsilon(z(t))$ converges to an interior NE. We note that the uniqueness of $\bar{x}$ does not imply the uniqueness of $\bar{z}$ unless C_ϵ is Legendre-induced (for which C_ϵ is one-to-one; this follows from Legendre theorem [52]). Hence, in general, the convergence of MD is to a set and not to

an equilibrium and $z(t)$ may continue to evolve even after $x(t)$ has reached an equilibrium. Note that rest points are not the only game relevant solutions for which MD may converge to. As pointed out in [18], there are non-rest points that occur on the boundary of Ω that are asymptotically stable under MD. The following lemma broadly summarizes the key convergence properties of MD (refer to [18], [19] for proofs).

Lemma 1. *Let $\mathcal{G}$ be a concave game with a globally strict VSS $x^\star$. Suppose that all players choose strategies according to MD (11). Let $x(t) = (x^p(t))_{p \in \mathcal{N}} = C_\epsilon(z(t))$ be generated by (11) and $C_\epsilon = (C_\epsilon^p)_{p \in \mathcal{N}}$ be induced by $\psi_\epsilon^p := \epsilon \vartheta^p$. For any $\gamma, \epsilon > 0$ and any $x(0) = C_\epsilon(z(0)) \in \Omega, z(0) \in \mathbb{R}^n$,*

(i) *suppose ϑ^p satisfies Assumption 3(i), $\forall p$, then $x(t) = C_\epsilon(z(t))$ converges to $x^\star$, whenever $x^\star$ is interior.*

(ii) *suppose Ω^p is compact and ϑ^p satisfies Assumption 3(ii), $\forall p$, then $x(t) = C_\epsilon(z(t))$ converges to $x^\star$.*

In general, MD does not converge beyond strict VSS, i.e., a mere VSS. This poses considerable limitations in practice, for instance, mere VSS are commonly found in ZS games. [12], [13] showed that every ZS finite game is merely (but not strictly) monotone, hence all of its NEs are mere (non-strict) VSS. A standard method to overcome non-convergence to the NE in ZS games is to calculate a time-averaged (*ergodic*) trajectory $x_{\text{avg}}(t) = t^{-1} \int_0^t x(\tau) d\tau$ in tandem with MD, that is,

$$\dot{z} = \gamma U(x), \quad x = C_\epsilon(z), \quad x_{\text{avg}}(t) = t^{-1} \int_0^t x(\tau) d\tau, \quad \text{(MDA)}$$

for which the time-averaged strategy x_{avg} has been shown in many contexts to converge, e.g., [18]. The main critique of using MDA is that the actual strategies do not arrive at the NE in the long-run, thereby making it unsuitable for on-line equilibrium seeking in the absence of a central planner or coordination between players. Furthermore, averaging may fail to converge outside of ZS games [57], which makes this approach vulnerable to parameter perturbation.

Another method for overcoming non-convergence is through *discounting*, which was studied in [28],

$$\dot{z} = \gamma(-z + U(x)), \quad x = C_\epsilon(z), \quad \text{(DMD)}$$

where DMD stands for *discounted* mirror descent. Compared to MD, an extra $-z$ term is inserted in the $\dot{z}$ system, which translates into an exponential weighted decay (or discounting) term in the closed-form solution $z(t)$. DMD has a connection with the so-called *weight decay* method in the machine learning literature [41], as $z \in C_\epsilon^{-1}(x)$ can be shown to be equivalent to a (usually non-Euclidean) regularization term, which directly interacts with the monotonicity property of U. While it is known that DMD could converge exactly in subclasses of finite games that exhibit symmetric interior NEs, whereby C_ϵ is also chosen to enforce symmetry (see [27]), in general, it cannot converge exactly to a NE, which also means that it cannot converge exactly to a VSS. This directs our attention to alternatives methods, such as higher-order augmentation of game dynamics [27], [33]–[35].

IV. SECOND-ORDER MIRROR DESCENT DYNAMICS

We now propose the *second-order mirror descent*, which in terms of each player p appears as,

$$\begin{cases} \dot{z}^p = \gamma(U^p(x) - \alpha(x^p - \xi^p)), & \dot{\xi}^p = \beta(x^p - \xi^p), \\ x^p = C_\epsilon^p(z^p), \end{cases} \quad \text{(MD2)}$$

and in stacked notation,

$$\dot{z} = \gamma(U(x) - \alpha(x - \xi)), \quad \dot{\xi} = \beta(x - \xi), \quad x = C_\epsilon(z), \quad (13)$$

where $x = (x^p)_{p \in \mathcal{N}}$, $z = (z^p)_{p \in \mathcal{N}}$, $\xi = (\xi^p)_{p \in \mathcal{N}}$, $C_\epsilon = (C_\epsilon^p)_{p \in \mathcal{N}}$, $U = (U^p)_{p \in \mathcal{N}}$. The rest point condition for MD2 is,

$$\mathbf{0} = U(\overline{x}), \quad \mathbf{0} = \overline{x} - \overline{\xi}, \quad \overline{x} = C_\epsilon(\overline{z}), \quad \overline{z} \in C_\epsilon^{-1}(\overline{x}), \quad (14)$$

which coincides with that of MD, i.e., $\overline{x} = x^\star$ is an interior NE.

Remark 5. (**Intuitive Learning Interpretation of MD2**) We can explicitly write $\xi^p(t) = (\xi_i^p(t))_{i \in [n_p]}$ as,

$$\xi_i^p(t) = e^{-\beta t}\xi_i^p(0) + \beta \int_0^t e^{-\beta(t-\tau)}x_i^p(\tau)\mathrm{d}\tau, \quad (15)$$

which represents an "exponential weighting" of the strategies. Hence, we refer to ξ as the *primal aggregate*, and z can be referred as the *dual aggregate*. We note that ξ^p resides in the unconstrained (ambient) space $\mathbb{R}^{n_p}$ containing the action set. Note that, whenever $\xi^p(0)$ is initialized in Ω^p, then $\xi^p(t) \in \Omega^p$ for all times. Furthermore, noting that we can re-write $\dot{z}^p$ subsystem in MD2 as $\dot{z}^p = \gamma(U^p(x) - \alpha\beta^{-1}\dot{\xi}^p)$. By assuming $z_i^p(0) = 0, \forall i, p$,

$$z_i^p(t) = \gamma(\int_0^t U_i^p(x(\tau))\mathrm{d}\tau - \alpha\beta^{-1}\int_0^t \dot{\xi}_i^p(\tau)\mathrm{d}\tau) \quad (16)$$
$$= \gamma(\int_0^t U_i^p(x)\mathrm{d}\tau - \alpha e^{-\beta t}(\beta^{-1}(1-e^{\beta t})\xi_i^p(0) + \int_0^t e^{\beta \tau}x_i^p\mathrm{d}\tau)).$$

It can be seen, when t is small, the effects of $\xi_i^p(0)$ and x_i^p are at their largest, so the presence of $\dot{\xi}$ term takes into consideration the uncertainty during the initial periods of play. Hence, the incorporation of $\dot{\xi}$ has an *exploratory* effect on the play, which can result in strategies being played more conservatively, preventing the players from immediately reaching a deadlock.

Remark 6. (**Comparison with Existing Dynamics**) The closest dynamics related to MD2 is the higher-order exponentially discounted dynamics (H-EXPD-RL) for finite games [27]. Specifically, when the higher-order augmentation is taken to be a high-pass filter, we obtain,

$$\begin{cases} \dot{z} = \gamma(U(x) - z - \alpha(x+\xi)), & \dot{\xi} = \beta(-x-\xi), \\ x = C_\epsilon(z). \end{cases} \quad (17)$$

(17) can be seen as the second-order extension of DMD. In [27] it was observed that (17) is robust to a greater degree of parameter perturbation in monotone games as compared to DMD.

Let's now consider MD2 as a purely second-order ODE in the dual-space. To simplify our calculations, we assume $x(t) = C_\epsilon(z(t))$ is differentiable. By using MD2 and noting that $\dot{z} = \gamma(U(x)-\alpha\beta^{-1}\dot{\xi})$, taking a time derivative of $\dot{z}$ and making the assumption that U and C_ϵ are $\mathcal{C}^1$, we have, $\ddot{z} = \gamma(\mathbf{J}_U(x)\dot{x} - \alpha\beta^{-1}\ddot{\xi})$, which, along with $\ddot{\xi} = \beta(\dot{x}-\dot{\xi})$, $x = C_\epsilon(z)$, $\dot{x} = \mathbf{J}_{C_\epsilon(z)}\dot{z}$, $\dot{\xi} = \beta\alpha^{-1}(U(x)-\gamma^{-1}\dot{z})$, we obtain,

$$\ddot{z} = \gamma\left[\mathbf{J}_U(C_\epsilon(z))\,\mathbf{J}_{C_\epsilon}(z) - \alpha\,\mathbf{J}_{C_\epsilon}(z) - \beta\gamma^{-1}I\right]\dot{z} + \gamma\beta U(x), \quad (18)$$

or by using the chain-rule for Jacobian,

$$\begin{cases} \ddot{z} = \gamma\left[\mathbf{J}_{U \circ C_\epsilon}(z) - \alpha\,\mathbf{J}_{C_\epsilon}(z) - \beta\gamma^{-1}I\right]\dot{z} + \gamma\beta U(x), \\ x = C_\epsilon(z), \end{cases} \quad (19)$$

where $U \circ C_\epsilon := U(C_\epsilon)$. To the best of our knowledge, most of the existing second-order continuous gradient-type dynamics (see [39] and references therein) *cannot* recover MD2 due to the presence of the Jacobian of C_ϵ. This is because in the unconstrained, primal setting (as studied in [26], [39]), C_ϵ is the identity and hence, $\mathbf{J}_{C_\epsilon}(z) = I, \forall z$ and its contribution is lumped together with $\beta\gamma^{-1}I$.

We proceed to demonstrate that the primal aggregate ξ contributes to the convergence of MD2 beyond strict VSS and such convergence depends on both the property of the regularizer as well as the topological properties of the underlying strategy set.

Theorem 1. *Let $\mathcal{G}$ be a concave game and assume that every*

interior NE is globally merely VS. Suppose that all players choose strategies according to MD2 (13). Let $x = (x^p(t))_{p \in \mathcal{N}} = C_\epsilon(z(t))$ be generated by (13) and $C_\epsilon = (C_\epsilon^p)_{p \in \mathcal{N}}$ be induced by $\psi_\epsilon^p := \epsilon\vartheta^p$. Let $x^\star$ denote an interior mere VSS (possibly from a set of them). For any $\alpha, \beta, \gamma, \epsilon > 0$, $x(0) = C_\epsilon(z(0)) \in \Omega$, $z(0), \xi(0) \in \mathbb{R}^n$,

(i) *suppose ϑ^p is twice-continuously differentiable ($\mathcal{C}^2$) and satisfies Assumption 3(i), $\forall p$, then $x(t) = C_\epsilon(z(t))$ converges to $x^\star$.*

(ii) *assume Ω^p is compact and ϑ^p satisfies Assumption 3(ii), $\forall p$, then $x(t) = C_\epsilon(z(t))$ converges to $x^\star$.*

We now sharpen Theorem 1 by considering several other classes of games in which MD2 is guaranteed to converge. From [2], [50],

Definition 3. *The game $\mathcal{G}$ is,*

(i) *pseudo-monotone (resp. strictly pseudo-monotone), if $U(x')^\top(x-x') \le 0 \implies U(x)^\top(x-x') \le 0, \forall x, x' \in \Omega$ (resp. $U(x)^\top(x-x') < 0$, with equality iff $x = x'$.)*

(ii) *quasi-monotone (resp. strictly quasi-monotone), $U(x')^\top(x-x') < 0 \implies U(x)^\top(x-x') \le 0, \forall x, x' \in \Omega$ (resp. $U(x)^\top(x-x') < 0$, with equality iff $x = x'$).*

Observe that when x' in the above definitions is NE $x^\star$ of a pseudo-monotone game (or a globally *strict NE* of a quasi-monotone game), it automatically implies the associated NE is globally mere VS. Moreover, $x^\star$ is globally strictly VS when the above games are strict. Since MD converges to globally strict VSS [1], therefore it converges to NE in strictly pseudo-monotone games and strict NE in strictly quasi-monotone games. Our next corollary, which immediately follows from Theorem 1, partially generalizes these results to the non-strict setting under MD2.

Corollary 1. *Let $\mathcal{G}$ be a concave game and assume that every NE is interior. Suppose that all players choose strategies according to MD2 (13). Let $x = (x^p(t))_{p \in \mathcal{N}} = C_\epsilon(z(t))$ be generated by (13) and $C_\epsilon = (C_\epsilon^p)_{p \in \mathcal{N}}$ be induced by $\psi_\epsilon^p := \epsilon\vartheta^p$, ϑ^p is $\mathcal{C}^2$ and satisfies Assumption 3(i), $\forall p$, then for any $\alpha, \beta, \gamma, \epsilon > 0$ and any $x(0) = C_\epsilon(z(0))$,*

(i) *$x(t)$ converges to an interior NE $x^\star$ whenever $\mathcal{G}$ is merely monotone or pseudo-monotone, and,*

(ii) *$x(t)$ converges to an interior strict NE $x^\star$ if $\mathcal{G}$ is quasi-monotone, whenever $x^\star$ exists.*

The same conclusions hold whenever ϑ^p satisfies Assumption 3(ii) and Ω^p is compact $\forall p$.

Remark 7. We note that the existence of a NE in pseudo-monotone games is guaranteed under our continuous game assumption [50, Theorem 2.3.5], possibly requiring $-\mathcal{U}^p$ to be coercive, see Remark 1, and the non-empty set of NEs coincides with the set of globally mere VSS [2, Theorem 2]. However, [2] pointed out that the existence of a globally merely VS for quasi-monotone games is not guaranteed in general. Instead, a globally merely VS exists under the stronger property of *properly* quasi-monotone [2, Theorem 3].

V. ADDITIONAL CONVERGENCE PROPERTIES OF MD2

In this section, we investigate two additional properties of MD2, namely, that of rate of convergence and regret minimization. To account for the geometry of the problem, we provide a *relative* extension to the strong VSS.

Definition 4. *Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be any differentiable, convex function with domain $\mathrm{dom}(h) = \Omega$. Then $x^\star \in \Omega$ is η-relatively strongly VS (with respect to h) if for all $x \in \Omega$, $U(x)^\top(x-x^\star) \le -\eta(D_h(x, x^\star) + D_h(x^\star, x))$, for some $\eta > 0$, where D_h is the Bregman divergence of h.*

We note that Definition 4 is analogous to that of a *relatively strongly monotone* game, which was previously introduced in [29]. Following [29, Theorem 4.4], it can be shown that MD converges to a relatively strongly VSS in $\mathcal{O}(e^{-\gamma\eta\epsilon^{-1}t})$ given a strongly-induced C_ϵ that is adapted to the geometry of this VSS. We now wish to provide a similar result for the convergence of MD2 towards a relatively strong VSS. However, exponential convergence does not follow from Theorem 1. Instead, we propose an augmented version of MD2 that exhibits exponential convergence,

$$\begin{cases} \dot{z} &= (U(x) - \gamma\alpha(x-\xi)), \qquad \dot{\xi} = \beta(x-\xi), \\ x &= C_\epsilon(z), \qquad\qquad\qquad\quad \dot{\gamma} = -\eta\epsilon^{-1}\gamma, \end{cases} \quad \text{(MD2}\gamma\text{)}$$

where $\gamma(0) > 0$. Observe that MD2γ is equivalent to the non-autonomous system,

$$\begin{cases} \dot{z} &= (U(x) - e^{-\eta\epsilon^{-1}t}\gamma(0)\alpha(x-\xi)), \quad \dot{\xi} = \beta(x-\xi), \\ x &= C_\epsilon(z), \end{cases} \quad \text{(20)}$$

whose rest point condition coincides with that of MD2.

Remark 8. From (20), MD2γ can be seen as MD with a *vanishing perturbation* $g(t,z,\xi) = e^{-\eta\epsilon^{-1}t}\alpha(0)(x-\xi)$, which allows for the following simple learning interpretation: when the players are aware that the game being played has a η-strong VSS, they no longer bother with exploring the strategy space during the initial stages and instead discard the extra information represented by $x-\xi$ exponentially fast.

Theorem 2. *Let $\mathcal{G}$ be a concave game with a unique interior η-strongly VSS relative to $h(x) = \sum_{p\in\mathcal{N}}\vartheta^p(x^p)$, denoted by $x^\star$, where $x = (x^p(t))_{p\in\mathcal{N}} = C_\epsilon(z(t))$ is generated by MD2γ and $C_\epsilon = (C_\epsilon^p)_{p\in\mathcal{N}}$ is induced by $\psi_\epsilon^p := \epsilon\vartheta^p$, ϑ^p satisfies Assumption 3(ii). Then for any $\alpha,\beta,\epsilon,\eta,\gamma_0 = \gamma(0) > 0$ and any $x_0 = (x^p(0))_{p\in\mathcal{N}} = C_\epsilon(z(0)), z(0), \xi_0 = \xi(0) \in \mathbb{R}^n$, x converges to $x^\star$ with the rate,*

$$D_h(x^\star, x) \leq e^{-\eta\epsilon^{-1}t}(\frac{\alpha\gamma_0}{2\beta\epsilon}\|\xi_0 - x^\star\|_2^2 + D_h(x^\star, x_0)). \quad \text{(21)}$$

Furthermore, since ϑ^p satisfies Assumption 3(ii), i.e., ϑ^p is ρ-strongly convex, therefore,

$$\|x^\star - x\|_2^2 \leq 2\rho^{-1}e^{-\eta\epsilon^{-1}t}(\frac{\alpha\gamma_0}{2\beta\epsilon}\|\xi_0 - x^\star\|_2^2 + D_h(x^\star, x_0)). \quad \text{(22)}$$

Remark 9. The above results confirm our intuition: the upper-bound rate of convergence diminishes faster when an equilibrium is more strongly VSS ($\eta\uparrow$), when C_ϵ is induced by an even more strongly convex regularizer ($\rho\uparrow$), when the regularization parameter goes down ($\epsilon\downarrow$), so that C_ϵ approximates a *best-response map*, or when the extra information $x-\xi$ diminishes more rapidly ($\beta\uparrow$).

Next, we turn our attention towards the question of regret minimization. Let us define the *time-averaged (external/static) regret function* of player $p \in \mathcal{N}$ as, $\mathcal{R}^p : [0,\infty) \to \mathbb{R}$, $\mathcal{R}^p(0) = 0$,

$$\mathcal{R}^p(t) = \max_{y^p\in\Omega^p}\frac{1}{t}\int_0^t \mathcal{U}^p(y^p; x^{-p}(\tau)) - \mathcal{U}^p(x(\tau))\mathrm{d}\tau. \quad \text{(23)}$$

Intuitively, the regret value at time t represents the time-averaged sum total of the payoff difference between the actual strategy $x(t) = (x^p(t); x^{-p}(t))$ versus the best strategy $(y^p; x^{-p}(t))$ that the player p could have played in hindsight, given any opponents' strategy $x^{-p}(t) \in \Omega^{-p}$. Player p's dynamics that generate $x^p(t)$ is said to achieve *no-regret* with respect to (23) if $\limsup_{t\to\infty}\mathcal{R}^p(t) \leq 0$.

While MD was shown to achieve no-regret in finite games, it cannot converge in ZS finite games with an interior mixed equilibrium [21]. In contrast to [20], [21], our next result, coupled with Theorem 1, shows that MD2 achieves both no-regret as well as exact convergence whenever the (interior) equilibrium is a mere VSS.

Theorem 3. *Let $\mathcal{G}$ be a concave game, where Ω^p is compact for all $p \in \mathcal{N}$. Then for every continuous trajectory of $x^{-p}(t)$ of opponents of player p, each player that uses MD2 achieves no-regret independently from the rest of the players.*

VI. CONSTRUCTION OF SECOND-ORDER PRIMAL-SPACE DYNAMICS

In this section, we show that MD2 can either be used to create new primal-space dynamics (that is, dynamics that solely evolve on Ω) or recover existing ones. To simplify our presentation and to avoid the technicality of non-differentiable mirror maps C_ϵ (for which examples of induced primal-space dynamics can still be generated through a technical treatment, see the so-called *projection dynamics* discussed in [18]), we impose the following general assumption,

Assumption 4. $\vartheta^p : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{\infty\}$ is steep and induces a $\mathcal{C}^1$ mirror map $C_\epsilon = (C_\epsilon^p)_{p\in\mathcal{N}} = (\nabla\psi_\epsilon^{p\star})_{p\in\mathcal{N}}$.

Under this assumption, taking the time-derivative of $x = C_\epsilon(z)$ shows that MD2 can be written as a pair of differential inclusions,

$$\dot{\xi} \in \beta(x-\xi), \quad \dot{x} \in \gamma\,\mathbf{J}_{C_\epsilon}(z)(U(x) - \alpha(x-\xi)), \quad \text{(24)}$$

where $\gamma > 0, \alpha,\beta \geq 0$ and the Jacobian of C_ϵ is,

$$\mathbf{J}_{C_\epsilon}(z) := \mathrm{blkdiag}(\mathbf{J}_{C_\epsilon^p}(z^p)) = \begin{bmatrix} \mathbf{J}_{C_\epsilon^1}(z^1) & & \\ & \ddots & \\ & & \mathbf{J}_{C_\epsilon^N}(z^N) \end{bmatrix} \in \mathbb{R}^{n\times n},$$

and each $\mathbf{J}_{C_\epsilon^p}(z^p) = \mathbf{J}_{\nabla\psi_\epsilon^{p\star}}(z^p) = \nabla^2\psi_\epsilon^{p\star}(z^p) \in \mathbb{R}^{n_p\times n_p}$. We note that the inclusions in (24) come from $z \in C_\epsilon^{-1}(x)$. We can further express (24) in each p as,

$$\dot{\xi}^p \in \beta(x^p - \xi^p), \quad \dot{x}^p \in \gamma\nabla^2\psi_\epsilon^{p\star}(z^p)(U^p(x) - \alpha(x^p - \xi^p)). \quad \text{(25)}$$

In the following, we offer two general ways of re-writing (25) as a set of ODEs in the primal-space.

General Case: Our next result shows that, in the general case, finding the primal-space dynamics associated with MD2 generally amounts to evaluating the operator $\nabla^2\psi_\epsilon^{p\star} \circ \nabla\psi_\epsilon^p := \nabla^2\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p)$.

Proposition 2. *Suppose Assumption 4 holds, then (25) reduces to,*

$$\begin{cases} \dot{\xi}^p &= \beta(x^p - \xi^p) \\ \dot{x}^p &= \gamma\nabla^2\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p))(U^p(x) - \alpha(x^p - \xi^p)). \end{cases} \quad \text{(26)}$$

Example 7. (**Simplex**) Let $\Omega^p = \{x^p \in \mathbb{R}_{\geq 0}^{n_p} | \|x^p\|_1 = 1\}$ and consider $\vartheta^p(x^p) = \sum_{i=1}^{n_p} x_i^p \log(x_i^p)$, $\psi_\epsilon^p(x^p) = \epsilon\vartheta^p(x^p)$ and $\psi_\epsilon^{p\star}(z^p) = \epsilon\log(\sum_{i=1}^{n_p}\exp(\epsilon^{-1}z_i^p))$. It can be shown that, $\nabla^2\psi_\epsilon^{p\star}(z^p) = \epsilon^{-1}(\mathrm{diag}(C_\epsilon^p(z^p)) - C_\epsilon^p(z^p)C_\epsilon^p(z^p)^\top)$ where $C_\epsilon^p(z^p) = (\exp(\epsilon^{-1}z_i^p)(\sum_{j=1}^{n_p}\exp(\epsilon^{-1}z_j^p))^{-1})_{i\in[n_p]}$ is the *soft-max* function (Example 5). Let $x^p \in \mathrm{rint}(\Omega^p)$. We can show $\nabla\psi_\epsilon^p(x^p) = \epsilon(\log(x^p) + \mathbf{1})$ (log is applied component-wise) and $C_\epsilon^p(\nabla\psi_\epsilon^p(x^p)) = x^p$. Taken together, we have, $\nabla^2\psi_\epsilon^{p\star} \circ \nabla\psi_\epsilon^p = \nabla^2\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p)) = \epsilon^{-1}(\mathrm{diag}(x^p) - x^p x^{p\top})$. By applying (26) and defining $\mathcal{H}(x^p) := \mathrm{diag}(x^p) - x^p x^{p\top}$, we obtain (27) as the induced primal-space dynamics of MD2 on the simplex,

$$\dot{\xi}^p = \beta(x^p - \xi^p), \quad \dot{x}^p = \gamma\epsilon^{-1}\mathcal{H}(x^p)(U^p(x) - \alpha(x^p - \xi^p)), \quad \text{(27)}$$

We note that (27) represents a version of second-order *replicator dynamics* (RD), which is different from the ones introduced in [33], [35]. In particular, when $\alpha = \beta = 0$, we recover the (multi-population) replicator dynamics [2], [12], [13], [18], [21], [44].

We can also reduce (27) to a *single population* model ($p = 1$),

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3291953

8

which takes the following form:

$$\begin{cases} \dot{\xi}_i &= \beta(x_i - \xi_i), \\ \dot{x}_i &= x_i(U_i(x) - \frac{\alpha}{\beta}\dot{\xi}_i - \sum_{j \in \mathcal{A}} x_j(U_j(x) - \frac{\alpha}{\beta}\dot{\xi}_j)), \end{cases} \quad (28)$$

where $\mathcal{A} = \{1, \ldots, n\}$ represents n subpopulations of non-atomic players, and U_i represents the fitness of subpopulation i. We note that structurally (28) shares similarity with the anticipatory replicator dynamics studied in [35], which can be written as,

$$\begin{cases} \dot{\xi}_i &= \beta(x_i - \xi_i), \\ \dot{x}_i &= x_i(U_i(x + \alpha\dot{\xi}) - \sum_{j \in \mathcal{A}} x_j(U_j(x + \alpha\dot{\xi}))). \end{cases} \quad (29)$$

Note that the variable ξ_j in (28) does not directly interacts with the pseudo-gradient/payoff vector $U(x)$, as is the case for the anticipatory RD.

Legendre and Supercoercive: A more structured scenario is when $\psi_\epsilon^p := \epsilon\vartheta^p$ is Legendre and supercoercive for all p (Assumption 3(i)). Since ψ_ϵ^p is finite on $\text{int}(\Omega^p)$, coercive, strictly convex and $\mathcal{C}^2$, then applying [62, Example 11.9, p. 480], $\psi_\epsilon^{p\star}$ shares the same properties and we have that $[\nabla^2\psi_\epsilon^p(x^p)]^{-1} = \nabla^2\psi_\epsilon^{p\star}(z^p) = (\nabla^2\psi_\epsilon^{p\star} \circ \nabla\psi_\epsilon^p)(x^p), \forall x^p = \nabla\psi_\epsilon^{p\star}(z^p) \in \text{int}(\Omega^p)$, which allows us to write (25) or (26) as,

$$\begin{cases} \dot{\xi}^p &= \beta(x^p - \xi^p), \\ \dot{x}^p &= \gamma[\nabla^2\psi_\epsilon^p(x^p)]^{-1}(U^p(x) - \alpha(x^p - \xi^p)), \end{cases} \quad (\text{NG2})$$

which we refer to as the *second-order natural gradient descent* (NG2). NG2 is so named because in the optimization setup ($p = 1$), for $U = -\nabla f$, where $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is some objective function and $\alpha, \beta = 0$, is analogous to the so-called *natural gradient descent* [31], $\dot{x} = -[\nabla^2\psi_\epsilon(x)]^{-1}\nabla f(x)$.

Example 8. (**Unconstrained**) Let $\Omega^p = \mathbb{R}^{n_p}$, $\vartheta^p(x^p) = \frac{1}{2}\|x^p\|_2^2$, then $\nabla^2\psi_\epsilon^p(x^p) = \epsilon I$ where $\psi_\epsilon^p = \epsilon\vartheta^p$. By comparing with NG2, we obtain,

$$\dot{\xi}^p = \beta(x^p - \xi^p), \quad \dot{x}^p = \gamma\epsilon^{-1}(U^p(x) - \alpha(x^p - \xi^p)), \quad (30)$$

which recovers the one recently introduced in [26]. (30) can be further shown via discretization to recover algorithms such as optimistic gradient-descent/ascent, Polyak's heavy-ball method, among others (see [26]). While (30) is by far the most standard choice for unconstrained action sets, more than one type of dynamics can reside on the same action sets. To witness, let $\Omega^p = \mathbb{R}^{n_p}$, $\vartheta^p(x^p) = \sum_{i=1}^{n_p} x_i^p \text{arcsinh}(x_i^p/\wp) - \sqrt{x_i^{p2} + \wp^2}, \wp > 0$, which represents an interpolation between entropic and Euclidean terms [61]. ϑ^p is Legendre and supercoercive (its convex conjugate is $\sum_{i=1}^{n_p} \wp\cosh(x_i^p)$, which is also Legendre). Let $\psi_\epsilon^p = \epsilon\vartheta^p$, then $\nabla\psi_\epsilon^p = \epsilon\sum_{i=1}^{n_p} \text{arcsinh}(x_i^p/\wp)$ and $\nabla^2\psi_\epsilon^p = \epsilon\left[\text{diag}(\left[\sqrt{x_1^{p2} + \wp^2}, \ldots, \sqrt{x_{n^p}^{p2} + \wp^2}\right]^\top)\right]^{-1}$. By comparing with NG2, we obtain,

$$\dot{\xi}^p = \beta(x^p - \xi^p), \quad \dot{x}^p = \gamma\epsilon^{-1}\mathcal{S}(x^p)(U^p(x) - \alpha(x^p - \xi^p)). \quad (31)$$

where $\mathcal{S}(x^p) := \text{diag}(\left[\sqrt{x_1^{p2} + \wp^2}, \ldots, \sqrt{x_{n^p}^{p2} + \wp^2}\right]^\top)$, which translates into a version of (30) with a strategy-dependent coefficient.

Example 9. (**Orthant**) Let $\Omega^p = \mathbb{R}_{\geq 0}^{n_p}$ and consider $\vartheta^p(x^p) = \sum_{j=1}^{n_p} x_i^p \log(x_i^p) - x_i^p$ with $0\log(0) = 0$. We can show $\nabla^2\psi_\epsilon^p(x^p) = \epsilon[\text{diag}(x^p)]^{-1}$ where $\psi_\epsilon^p = \epsilon\vartheta^p$. Using NG2, we immediately arrive at,

$$\dot{\xi}^p = \beta(x^p - \xi^p) \quad \dot{x}^p = \gamma\epsilon^{-1}\text{diag}(x^p)(U^p(x) - \alpha(x^p - \xi^p)). \quad (32)$$

We note that the first-order variant ($\alpha = \beta = 0$) of (32) was recently studied by [60] as the *mean-dynamics* to a payoff-based learning dynamics in continuous games with non-negative orthant action sets. Suppose instead $\vartheta^p(x^p) = -\sum_{i=1}^n \ln(x_i^p)$ (the log-barrier function), which leads to $\nabla^2\psi_\epsilon^p(x^p) = \epsilon[\text{diag}(x^{p2})]^{-1}$ and,

$$\dot{\xi}^p = \beta(x^p - \xi^p) \quad \dot{x}^p = \gamma\epsilon^{-1}\text{diag}(x^{p2})(U^p(x) - \alpha(x^p - \xi^p)). \quad (33)$$

We stress however that the log-barrier does not fall under our definition of a regularizer, as $\text{dom}(\vartheta^p)$ is a proper subset of Ω^p. (33) with $\alpha = \beta = 0$ has been previously studied in the context of optimization, see [63].

VII. DISCRETE-TIME SECOND-ORDER DUAL AVERAGING WITH NOISY OBSERVATIONS

So far we have considered a continuous-time setup whereby each player is able to acquire a partial-gradient U^p at each time instance. In practical scenarios where the games are played in discrete-time, the acquired pseudo-gradient information could be corrupted due to a multitude of reasons, such as a noisy communication channel. This leads us to consider the so-called "noise-corrupted" pseudo-gradient scenario (also known as "semi-bandit" learning). In this setting, each player receives a realization of a so-called noise-corrupted version of the true pseudo-gradient,

$$\widehat{U}_{k+1}^p := U^p(x_k) + \zeta_{k+1}^p,$$

where ζ_k^p is some noise process. A special case of the semi-bandit scenario is when the payoff is obtained as the expectation of the true payoff, i.e., $\mathcal{U}^p(x_k^p; x_k^{-p}) = \mathbb{E}[\mathcal{U}^p(x_k^p; x_k^{-p}, v_k^p)]$, v_k^p some random vector, where $\mathbb{E}$ denotes the expectation operator. Here $\widehat{U}_k^p$ is an estimate of the expected partial-gradient $\nabla_{x^p}\mathbb{E}[\mathcal{U}^p(x_k^p; x_k^{-p}, v_k^p)]$.

Let's consider the convergence of the discrete-time MD2 with noisy observations which we refer to as second-order *dual averaging* or DA2,

$$\begin{cases} X_k^p &= C_\epsilon^p(Z_k^p) \\ Z_{k+1}^p &= Z_k^p + \gamma t_{k+1}(\widehat{U}_{k+1}^p - \alpha(X_k^p - \Xi_k^p)) \\ \Xi_{k+1}^p &= \Xi_k^p + \tau_{k+1}\beta(X_k^p - \Xi_k^p), \end{cases} \quad (\text{DA2})$$

where X_k^p, Z_k^p, Ξ_k^p are the stochastic counter-part of (resp.) x^p, z^p, ξ^p at time k. $\{t_k\}_{k \in \mathbb{N}}$, $\{\tau_k\}_{k \in \mathbb{N}}$ denote deterministic sequences of non-increasing step-sizes, assumed to be common for all players. $\alpha, \beta, \gamma > 0$ are the auxiliary parameters as from before. When $\alpha = \beta = 0$, we refer to the resulting expression as the dual averaging with noisy observations,

$$\begin{cases} X_k^p &= C_\epsilon^p(Z_k^p) \\ Z_{k+1}^p &= Z_k^p + \gamma t_{k+1}\widehat{U}_{k+1}^p, \end{cases} \quad (\text{DA})$$

which coincides with the dual averaging scheme studied in [1] for $\gamma = 1$.

To analyze the convergence behavior of DA2, we impose the following set of regularity assumptions [54], [55].

Assumption 5. ($\ell^2 \backslash \ell^1$-**Summability and Diminishing Step-sizes**)

$$\begin{array}{lll} \sum_{k \in \mathbb{N}} t_k = \infty & \sum_{k \in \mathbb{N}} t_k^2 < \infty & \lim_{k \to \infty} t_k = 0, \\ \sum_{k \in \mathbb{N}} \tau_k = \infty & \sum_{k \in \mathbb{N}} \tau_k^2 < \infty & \lim_{k \to \infty} \tau_k = 0. \end{array} \quad (34)$$

Assumption 6. (L^2-**Bounded Martingale Difference Noise**) Assume $\{\zeta_k\}_{k \in \mathbb{N}}, \zeta_k = (\zeta_k^p)_{p \in \mathcal{N}}$ is a L^2-bounded marginal difference process adapted to the filtration $\{\mathcal{F}_k\}_{k \in \mathbb{N}}$: each ζ_k is a random vector that is measurable with respect to $\mathcal{F}_k$ for each k, where each $\mathcal{F}_k$ is the σ-field, i.e., $\mathcal{F}_k = \sigma(\Xi_0, Z_0, \zeta_0, \ldots, \zeta_k)$ and $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$. In particular, $\{\zeta_k\}_{k \in \mathbb{N}}$ satisfies,

$$\mathbb{E}[\zeta_{k+1}|\mathcal{F}_k] = 0, \forall k \in \mathbb{N} \quad a.s., \quad (35)$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3291953

9

and for some $\sigma \geq 0$,

$$\mathbb{E}[\|\zeta_{k+1}\|_\star^2 | \mathcal{F}_k] \leq \sigma^2, \forall k \in \mathbb{N} \quad a.s. \tag{36}$$

Assumption 7. (**Bounded Iterates**)

$$\sup_k \|Z_k^p\| < \infty \quad \sup_k \|\Xi_k^p\| < \infty, \forall p. \tag{37}$$

Finally, we impose "global integrability" on MD2, i.e., MD2 has a complete vector field. To do so, we need to convert MD2 into a first-order system by defining: $\omega^p := (\xi^p, z^p)$, which generates the following stacked system on $\mathbb{R}^{2n} \cong \mathbb{R}^n \times \mathbb{R}^n$,

$$
\begin{aligned}
\dot{\omega}^p = \begin{bmatrix} \dot{\xi}^p \\ \dot{z}^p \end{bmatrix} &= \begin{bmatrix} \beta(C_\epsilon^p(z^p) - \xi^p) \\ \gamma(U^p(C_\epsilon(z)) - \alpha(C_\epsilon^p(z^p) - \xi^p)) \end{bmatrix}, \\
&= \begin{bmatrix} \beta C_\epsilon^p(z^p) \\ \gamma(U^p(C_\epsilon(z)) - \alpha C_\epsilon^p(z^p)) \end{bmatrix} + \begin{bmatrix} -\beta\xi^p \\ \gamma\alpha\xi^p \end{bmatrix},
\end{aligned}
\tag{38}
$$

or equivalently, $\dot{\omega}^p = F^p(\omega) = g^p(z) + A^p(\xi^p)$, where, $A^p(\xi^p) = \begin{bmatrix} -\beta\xi^p \\ \gamma\alpha\xi^p \end{bmatrix}$ and $g^p(z) = \begin{bmatrix} \beta C_\epsilon^p(z^p) \\ \gamma(U^p(C_\epsilon(z)) - \alpha C_\epsilon^p(z^p)) \end{bmatrix}$.

We then proceed to impose the following assumption on the overall system,

$$\dot{\omega} = F(\omega), \omega = (\omega^p)_{p \in \mathcal{N}}, F = (F^p)_{p \in \mathcal{N}}. \tag{39}$$

Assumption 8. (**Global Integrability**) The vector field $F : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ of (39) is continuous globally integrable, that is, for every initial condition $(\xi(0), z(0)) \in \mathbb{R}^{2n}$, the unique solution of (39) is defined for all $t \in \mathbb{R}$.

Remark 10. We note that global integrability of MD2 is satisfied whenever the following holds,

(i) Suppose C_ϵ^p is bounded on $\mathbb{R}^{n_p}$ and U^p is bounded continuous locally Lipschitz on $\text{ran}(C_\epsilon^p)$ for all $p \in \mathcal{N}$, then F (39) is bounded locally Lipschitz and hence continuous globally integrable. This follows from [54].

(ii) Suppose U^p is continuous locally Lipschitz and both C_ϵ^p and $U^p \circ C_\epsilon$ are *sublinear* for all p, that is,

$$\limsup_{\|z\|_2 \to \infty} \frac{\|C_\epsilon^p(z)\|_2}{\|z\|_2} < \infty, \quad \limsup_{\|z\|_2 \to \infty} \frac{\|U^p \circ C_\epsilon(z)\|_2}{\|z\|_2} < \infty, \tag{40}$$

then F is sublinear and hence continuous and globally integrable. This follows from [56].

Under Assumption 5 – 8, DA2 can be shown to track the continuous trajectories generated by MD2 via stochastic approximation arguments [54], [55].

Theorem 4. *Let $\mathcal{G}$ be a concave game and assume that every interior NE is globally merely VS. Suppose that all players choose strategies according to DA2 and $C_\epsilon = (C_\epsilon^p)_{p \in \mathcal{N}}$ is induced by $\psi_\epsilon^p := \epsilon \vartheta^p$, where ϑ^p is $\mathcal{C}^2$ and satisfies Assumption 3(i). Suppose that the following assumptions hold,*

- *Assumption 5 ($\ell^2 \backslash \ell^1$ Summability and diminishing step-sizes)*
- *Assumption 6 (L^2-bounded Martingale difference noise),*
- *Assumption 7 (Bounded iterates)*
- *Assumption 8 (Global integrability)*

then X_k converges to an interior mere VSS of $\mathcal{G}$ almost surely. The same conclusions hold whenever ϑ^p satisfies Assumption 3(ii) and Ω^p is compact $\forall p$.

Remark 11. The closest result to ours is [1, Theorem 4.7], but for DA (i.e., $\alpha, \beta = 0$). Assumption 5, Assumption 6 are identical to theirs and Assumption 8 encapsulates the Lipschitz continuous requirement for U. The major departure is Assumption 7, which does not hold under certain circumstances, such as convergence towards (boundary) strict VSS in finite games. Hence our result only deals with interior

mere VSS in general. To account for these boundary VSS possibly involves an extension of the inductive shadowing argument used by [1] which we leave for future work.

Remark 12. In addition to DA [1], there are several other algorithms that converge in similar settings. The most notable example is the *mirror-prox/optimistic mirror descent* algorithm, which converges to mere VSS in perfect-gradient feedback setting [46]. However, the authors of [46] noted that convergence in null-coherent saddle point problems (a game with a mere VSS) fails in the presence of noise. This is a key advantage of DA2 over mirror-prox. Another algorithm is the *stochastic iterative Tikhonov method* of [25], which converges in the semi-bandit setting. However, [25] requires the game to be strictly monotone, whereas DA2 does not require monotonicity.

VIII. SIMULATIONS

In this section, we consider three illustrative examples. The first example is the RPS game with a non-negative payoff for ties as in Example 3. We provide convergence behavior of MD and MD2 as well as MDA towards mere and weak VSS. We then provide the convergence behavior of DA2. The second example concerns a wireless power control game previously studied in [58]. We show that DA2 converges in this game under different noise assumptions and study the effect of the number of players. Finally, we provide an example involving a generative adversarial network (GAN), which admits a locally mere VSS. We show that DA2 also converges in this game under different noise assumptions. In lieu of exact estimation of the basin of attraction for *locally mere* VSS, which is difficult, we deal with all such cases through appropriate initialization.

Example 10. (**RPS with non-negative payoff for ties**) Consider the RPS game as discussed in Example 3. We set the game parameters to be $l = 0, w = 2$ and vary the tie payoff parameter $\varsigma \in [1, w) = [1, 2)$ to induce different VS properties on the NE $x^\star = (x^p)_{p \in \mathcal{N}}, x^p = (1/3, 1/3, 1/3)$. We simulate both MD and MD2 at initial conditions $z(0) = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix}^\top, \xi(0) = \mathbf{0}. \gamma, \beta, \alpha$ are kept as 1.

We begin by contrasting the continuous-time MD and MD2. For $\varsigma = 1$, this game is merely monotone, $x^\star$ is globally merely VS, MD2 converges by Theorem 1 while MD diverges (Figure 1). For $\varsigma = 1.1$, the game is 0.1-weakly monotone, but since $x^\star$ is only 0.1-weak VSS therefore it can be considered a nearly mere VSS (see Remark 2). In this case, MD2 still converges while MD approaches a heteroclinic orbit (Figure 2).
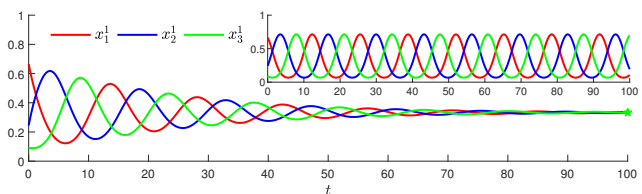

Fig. 1: $\varsigma = 1$, $\mathcal{G}$ merely monotone, $x^\star$ is merely VS. MD2 converges (Embedded: MD, cycling).

It is useful to also examine the advantage of MD2 over time-averaged MD (MDA). While MDA does converge for $\varsigma = 1$ towards the mere VSS, when the equilibrium becomes just slightly weak ($\varsigma = 1.1$), it no longer converges and instead approaches a Shapley triangle [57] (Figure 3). This shows time-averaging is in general not a panacea to non-convergence.

Next, we perform a set of experiments for the globally merely VS case ($\varsigma = 1$) using DA2, where we assume the same initial condition as before. For each of the following simulations, we separately perturb the pseudo-gradient U^p (9) with zero-mean Gaussian noise
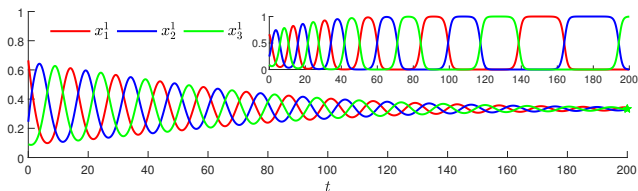
Fig. 2: $\varsigma = 1.1$, $\mathcal{G}$ 0.1-weakly monotone, $x^\star$ is 0.1-weakly VS. MD2 converges (Embedded: MD, cycling).
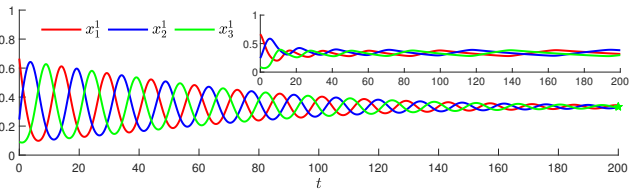

Fig. 3: $\varsigma = 1.1$, $\mathcal{G}$ 0.1-weakly monotone, $x^\star$ is 0.1-weakly VS. MD2 converges (Embedded: MDA, cycling).

with the same variance $\sigma_\zeta^2 > 0$ across all players. Such a perturbed gradient setting could describe a game involving multiple users interacting over a fully connected network with noisy communication channels. Figure 4 shows that DA2 easily converge in the low variance regimes $\sigma_\zeta^2 = 1$. As we employ larger variance, e.g., $\sigma_\zeta^2 = 10$, the standard step-sizes no longer leads to convergence: a larger t_k will amplify the additive noise. Instead, we utilize step-size sequences of the form $\frac{c}{k^\iota}$ and search over both $\iota \in (0,1), c > 0$ for optimal sets of parameters. The simulation with tuned step-size is shown in Figure 5. We see that despite the larger noise injected into U^p, the strategies $X = (X^p)_{p \in \mathcal{N}}$ still converge toward the vicinity of the mere VSS.
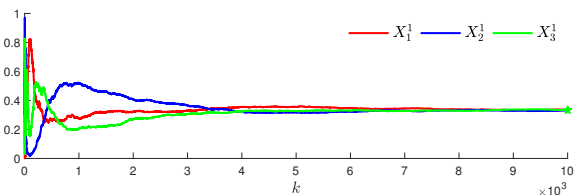

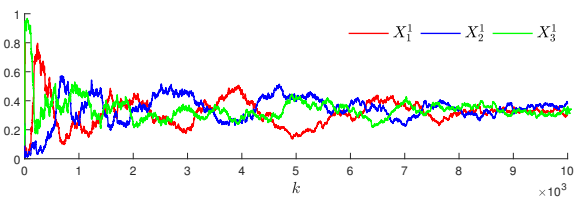Fig. 4: $\varsigma = 1$, $\sigma_\zeta^2 = 1$, $t_k = 4/k$, $\tau_k = 1/k$.

Fig. 5: $\varsigma = 1$, $\sigma_\zeta^2 = 10$, $t_k = 0.23/k^{0.48}$, $\tau_k = 0.34/k^{0.88}$.

Example 11. (**Wireless power control game with a non-concave potential**) Consider the wireless power control game in [58], where $\mathcal{N} = \{1, \ldots, N\}$ network users decide on *intensities* $x^p \in \mathbb{R}$ of power flow to send over a wireless network. These intensities are converted to the transmitted power through an exponential function, i.e., $\exp(x^p) \in \mathbb{R}_{>0}$. The payoff function for each user $p \in \mathcal{N}$ is modeled as,

$$\mathcal{U}^p(x^p; x^{-p}) = \log\left(1 + \frac{a^p \exp(x^p)}{1 + \sum_{p \neq r} a^r \exp(x^r)}\right) - \mathcal{K}^p(x^p) \quad (41)$$

where $a^p \in (0,1], \forall p$ and $\mathcal{K}^p(x^p) = b^p \log(1 + \exp(x^p)) - c^p x^p$ is the cost of user p for transmission, $b^p > 0, c^p \geq 0$. The partial-gradient U^p is calculated to be,

$$U^p(x) = \frac{a^p \exp(x^p)}{1 + \sum_{r \in \mathcal{N}} a^r \exp(x^r)} - \frac{b^p \exp(x^p)}{1 + \exp(x^p)} + c^p. \quad (42)$$

The second-order partial derivative of $\mathcal{U}^p$ can be calculated to be,

$$\frac{\partial^2 \mathcal{U}^p(x^p; x^{-p})}{\partial x^q \partial x^p} = \begin{cases} \dfrac{-a^p a^q \exp(x^p + x^q)}{(1 + \sum_{r \in \mathcal{N}} a^r \exp(x^r))^2} & p \neq q \\ \dfrac{a^p \exp(x^p)(1 + \sum_{r \neq p} a^r \exp(x^r))}{(1 + \sum_{r \in \mathcal{N}} a^r \exp(x^r))^2} & \\ \quad - \dfrac{b^p \exp(x^p)}{(1 + \exp(x^p))^2} & p = q \end{cases}$$
$$(43)$$

which generates a symmetric Jacobian $\mathbf{J}_U(x), \forall x$,

$$\mathbf{J}_U(x)_{p,p} = \frac{a^p \exp(x^p)(1 + \sum_{r \neq p} a^r \exp(x^r))}{(1 + \sum_{r \in \mathcal{N}} a^r \exp(x^r))^2} - \frac{b^p \exp(x^p)}{(1 + \exp(x^p))^2} \quad (44)$$

and

$$\mathbf{J}_U(x)_{p,q} = \frac{-a^p a^q \exp(x^p + x^q)}{(1 + \sum_{r \in \mathcal{N}} a^r \exp(x^r))^2}, p \neq q, \quad (45)$$

hence the game is a potential game with a non-concave potential function P such that $\nabla P = U$.

Consider an example with $N = 2$ and problem parameters, $a = (a^1, a^2) = (1,1), b = (b^1, b^2) = (4,4), c = (c^1, c^2) = (3,3)$. For all players, $Z^p(0)$ is sampled uniformly from $[0, 10]$, $\Xi^p(0) = 0$. The NE of this game can be found at $x^\star = (1.8663, 1.8663)$. The Jacobian at the NE is $\mathbf{J}(x^\star) = \text{diag}(\begin{bmatrix} -4.308 & 0 \end{bmatrix}^\top)$, hence by Proposition 1, the NE is a locally mere VSS. By Theorem 4, DA2 converges towards this NE. For each of the following experiments, we simulate for $T = 10^4$ steps and report the strategies at termination. We restrict Ω^p to be a large but compact set $[-1000, 1000]^2$ and set C_ϵ^p to be the projection operator, $C_\epsilon^p(Z^p) = \text{argmin}_{y^p \in \Omega^p} \|\epsilon^{-1} Z^p - y^p\|_2^2$ with $\epsilon = 1$. Unless specified otherwise, all other parameters γ, α, β are kept as 1. All additive noise are zero-mean Gaussian with the same variance σ_ζ^2.

We start our experiment with a small variance $\sigma_\zeta^2 = 0.1$ and set the step-sizes to be $t_k = 0.39/k^{0.26}, \tau_k = 0.12/k^{0.64}$. Figure 6 shows that DA2 converges to the NE. We increase the variance to $\sigma_\zeta^2 = 10$ and Figure 7 shows similar observation.
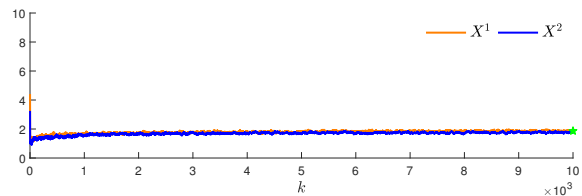

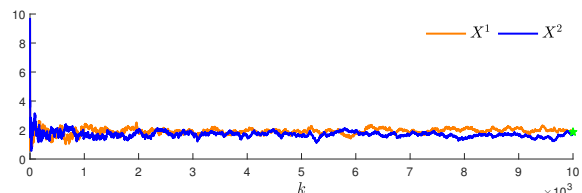Fig. 6: $\sigma_\zeta^2 = 0.1$, $t_k = 0.39/k^{0.26}$, $\tau_k = 0.12/k^{0.64}$

Fig. 7: $\sigma_\zeta^2 = 10$, $t_k = 0.38/k^{0.47}$, $\tau_k = 0.13/k^{0.71}$

Next, we investigate some possible computational issues can arise

when we increase the number of players in this game. Consider an example with $N = 10$ players. The game parameters are:

$$a = (12, 4, 2, 5, 20, 20, 12, 3, 1, 2),$$
$$b = (15, 20, 20, 14, 21, 20, 16, 19, 17, 17),$$
$$c = (13, 12, 14, 8, 10, 19, 10, 12, 15, 1).$$

A variance of $\sigma_\zeta^2 = 1$ is applied for the following experiments. We employ step-size sequences $t_k = 0.4/k^{0.033}, \tau_k = 0.78/k^{0.001}$ across all players. The NE of this game is located at $x^\star = (1.87, 0.41, 0.85, 0.28, -0.10, 16, 0.51, 0.54, 2.01, -2.77)$ (rounded to two decimal places) and was confirmed to be a *nearly* mere VSS (Remark 2). We note that an exact mere VSS is difficult to produce for this game when there are a large number of players.
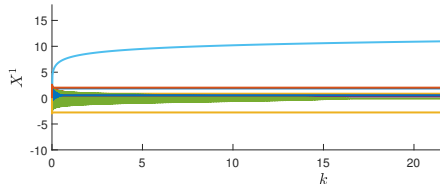

Fig. 8: $\sigma_\zeta^2 = 1$, $t_k = 0.40/k^{0.033}, \tau_k = 0.78/k^{0.001}$ (X^2 omitted).

Figure 8 shows that the strategy X^1 converges to $(1.87, 0.41, 0.85, 0.28, -0.10, 11.29, 0.51, 0.54, 2.01, -2.77)$ which is exactly the same as $x^\star$ in all entries except for the sixth entry (16.51 vs 11.29), which is likely due to the small rate parameters employed in the step-sizes. This highlights the difficulty of employing the same step-size sequences across all players. We adjust all the parameters β, α and step-sizes $ck^{-\iota}, \iota \in (0,1), c > 0$ on a per-player basis, which resulted in a much closer convergence to the NE as opposed to applying them across all players uniformly (Figure 9).
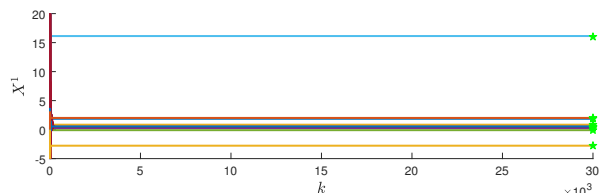

Fig. 9: $\sigma_\zeta^2 = 1$, per-player tuned parameters (X^2 omitted).

Example 12. (**Learning to Generate a Gaussian**)

Let $Z \sim P(z)$ and $X \sim Q(x)$ be two random variables. We wish to construct a model $G_\theta : \mathbb{R}^r \to \mathbb{R}^m, r, m \geq 1$, with an unknown, continuous parameter θ such that $G_\theta(Z)$ recovers the statistics (mean, variance, etc.) of X. Following [53], G_θ can be constructed through solving,

$$\min_\theta \max_w \mathbb{E}_X(D_w(X)) - \mathbb{E}_Z(D_w(G_\theta(Z))), \qquad (\star)$$

where the model $D_w : \mathbb{R}^m \to \mathbb{R}$ is parametrized by an unknown continuous parameter w. This problem can be thought of as a game between the owners of the models G_θ and D_w whereby θ, w are their respective strategies. The owner of G_θ (or the *designer*) varies θ and uses $G_\theta(Z)$ to estimate the statistics of X, whereas the owner of D_w (or the *tester*) varies w in an attempt to incur the largest penalty possible for the discrepancy between $G_\theta(Z)$ and X.

We construct a more complicated variant of the examples in [28], [53], whereby we assume that the G_θ owner wishes to learn the mean and variance of a one-dimensional Gaussian at the same time: let $Z \sim \mathcal{N}(0,1)$, and $X \sim \mathcal{N}(v, \sigma^2)$, where $v \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$

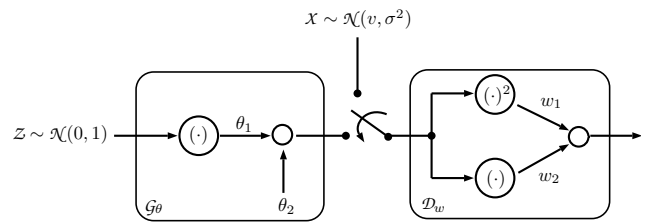Fig. 10: Diagram representation of G_θ and D_w models

are the mean and standard deviation of a Gaussian distribution. Let $D_w(X) = w_1 X^2 + w_2 X, w = (w_1, w_2) \in \mathbb{R} \times \mathbb{R}$ and $G_\theta(Z) = \theta_1 Z + \theta_2, \theta = (\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}$, then a brief calculation shows that

$$(\star) = w_1((\sigma^2 + v^2) - (\theta_1^2 + \theta_2^2)) + w_2(v - \theta_2).$$

Let $x^1 = \theta, x^2 = w$, we obtain a two-player ZS game,

$$\mathcal{U}^1(x^1; x^2) = x_1^2(\sigma^2 + v^2 - \textstyle\sum_{i=1}^2(x_i^1)^2) + x_2^2(v - x_2^1) \quad (46)$$

and $\mathcal{U}^2(x^2; x^1) = -\mathcal{U}^1(x^1; x^2)$ where the action sets are, $\Omega^1 = \mathbb{R} \times \mathbb{R} = \Omega^2$. The pseudo-gradient is, $U(x) = (-2x_1^2 x_1^1, -2x_1^2 x_2^1 - x_2^2, \sum_{i=1}^2(x_i^1)^2 - (\sigma^2 + v^2), x_2^1 - v)$, which implies an interior NE at $x^\star = (x^{1^\star}, x^{2^\star}) = ((\sigma, v), (0, 0))$. The Jacobian of U is shown to be,

$$\mathbf{J}_U(x) = \begin{bmatrix} -2x_1^2 & 0 & -2x_1^1 & 0 \\ 0 & -2x_1^2 & -2x_2^1 & -1 \\ 2x_1^1 & 2x_2^1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \qquad (47)$$

Since $y^\top \mathfrak{J}_U(x) y = \frac{1}{2} y^\top (\mathbf{J}_U(x) + \mathbf{J}_U^\top(x)) y = -2((y_1)^2 + (y_2)^2) x_1^2$ and x_1^2 is not restricted to be positive, therefore, this game is not monotone [50, Prop. 2.3.2]. However, at the interior NE, $x^{2^\star} = (x_1^{2^\star}, x_2^{2^\star}) = (0,0) \implies x_1^2 = 0$, and hence $y^\top \mathfrak{J}_U(x^\star) y = \frac{1}{2} y^\top (\mathbf{J}_U(x^\star) + \mathbf{J}_U^\top(x^\star)) y = -2((y_1)^2 + (y_2)^2)(0) = 0$. By Proposition 1(ii'), this implies that $x^\star$ is locally merely VS.

Consider an example with $v = 10, \sigma = 5$. We perform two simulations of the discrete-time MD2 with or without perturbation on the pseudo-gradient. We assume $Z(0)$ is uniformly sampled as follows: $Z^1(0) \sim \text{unif}([10, 50] \times [20, 40])$, $Z^2(0) \sim \text{unif}([0, 5] \times [0, 5])$, and $\Xi(0) = \mathbf{0}$. We restrict Ω^p to be a compact set $[-1000, 1000]^2$ and set C_ϵ^p to be the projection operator as in Example 11. We set $\lambda = \beta = \alpha = 1$. Each of the following simulations is ran for $T = 4 \times 10^4$ steps, with initial conditions randomly sampled as stated above. Note that our simulations has experimentally shown to accommodate larger sets of initial conditions than the ones provided above.

First, we consider the noiseless case, i.e., $\sigma_\zeta^2 = 0$. From Figure 11, we see that X^1 converges to $X^{1\star}$ (highlighted using a green star).
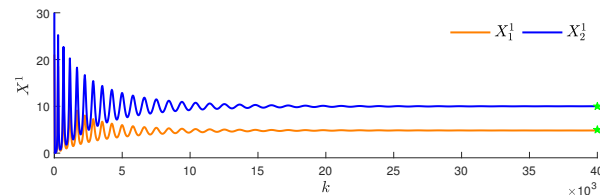

Fig. 11: $\sigma_\zeta^2 = 0$, $t_k = \dfrac{0.0108}{k^{0.3971}}, \tau_k = \dfrac{0.0485}{k^{0.7877}}$ (X^2 omitted).

Next, we consider the harder case where U^p is subjected to zero-mean Gaussian noise with a very large variance $\sigma_\zeta^2 = 30^2$ for each p. The trajectories of X^1 are shown in Figure 12. Our previous observation continues to hold.
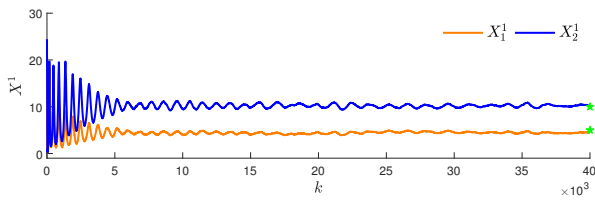
Fig. 12: $\sigma_\zeta^2 = 30^2$, $t_k = \dfrac{0.0108}{k^{0.3971}}$, $\tau_k = \dfrac{0.0485}{k^{0.7877}}$ (X^2 omitted).

IX. CONCLUSIONS

In this paper, we have shown that MD2 overcomes many short-comings of the first-order MD as well as other members of this class of dynamics, such as MD with time-averaging or discounting. To summarize our findings: first, we showed that the trajectories of MD2 converge exactly to an interior mere (and not necessarily strictly) VSS (Theorem 1). As such, it converges to interior mere NEs of monotone and pseudo-monotone games, and interior strict NE of quasi-monotone games (Corollary 1). Through simulation, we also found that MD2 is robust around interior mere VSS (Remark 2, Example 10). Next, we provided a modification to MD2 based on vanishing perturbation which results in an exponential rate of convergence towards a strongly stable VSS (Theorem 2). We then showed MD2 can achieve no-regret while converging beyond a strict VSS, such as to an interior/fully-mixed NEs of zero-sum finite games (Theorem 3). This result improves our current understandings between no-regret learning and convergence [20], [21]. We then provided several ways of deriving primal-space dynamics from MD2. Finally, we moved to the scenario where the game evolves in discrete-time while the pseudo-gradient is subjected to noise perturbation. Using stochastic approximation techniques, we were able to relate the limiting behavior of MD2 with that of its discrete-time, noisy counter-part, and showed that the discrete MD2 with noisy observations or second-order dual averaging (DA2) can converge to the interior mere VSS under standard regularity assumptions (Theorem 4).

There are several outstanding issues that could be better understood. First, our work did not provide a thorough convergence proof for local VSS, which are dealt with appropriate initialization. Moreover, we did not address convergence towards boundary point for MD2, as our proofs relied on an *interiority condition* associated with interior NE (14). One possible approach for dealing with these boundary solutions is to adopt the approach of [18], by restricting ourselves to the context of finite games, where these boundary solutions are quite relevant. Another basic issue is that it is still not entirely clear to us at this time how MD2 relates to other existing continuous-time dynamics for optimization or games. Aside from the cases that we know of, e.g., [26], it is possible that there are other existing algorithms that appear as specific instances of MD2.

Our work opens up an extensive array of directions on the interface between dynamical systems and games. As a starting point, in the continuous-time, an interesting direction would be to explain the difference between MD and MD2 in zero-sum games through a volume compressibility perspective as in [15], [22]. We can also try to craft third or even higher-order versions of MD based on the technique in [27]. In the discrete-time case, we can further reduce the information, to that of the *zero-order* information case (or full-bandit feedback). Moreover, we can tackle the case whereby the game itself has parameters that are time-varying. Finally, it is also worthwhile to examine MD2 in a discrete-time setting whereby the mirror map is replaced with a proximal operator.

Symbol	Definition
$x/x(t)/X_k$	Strategy/at time t/at time k
$z/z(t)/Z_k$	Dual aggregate/at time t/at time k
$\xi/\xi(t)/\Xi_k$	Primal aggregate/at time t/at time k
$\mathcal{U}^p/U^p/U$	Payoff/partial-gradient/pseudo-gradient
$\mathbf{J}_U/\mathfrak{J}_U$	Jacobian of U/symmetric game Jacobian of U
$\vartheta^p/\psi_\epsilon^p/\psi_\epsilon^{p\star}$	Regularizer/$\psi_\epsilon^p = \epsilon\vartheta^p$/convex conjugate of ψ_ϵ^p
$C_\epsilon^p/C_\epsilon/\mathbf{J}_{C_\epsilon}$	Player p's/overall mirror map/Jacobian of C_ϵ
$\gamma,\epsilon,\alpha,\beta$	Parameters associated with MD/MD2/DA2
I/O	Identity matrix/zero matrix

TABLE 1: A list of main notations used in this paper.

X. APPENDIX

The following two lemmas are standard in this area of literature, see [18], [22], [28]. In particular, Lemma 2 follows directly from the Legendre theorem [52].

Lemma 2. *Let $\psi_\epsilon^p := \epsilon\vartheta^p$, $\epsilon > 0$, where ϑ^p satisfies Assumption 3(ii), and let $\psi_\epsilon^{p\star}$ be the convex conjugate of ψ_ϵ^p. Then,*

(i) *$\psi_\epsilon^{p\star} : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{\infty\}$ is closed, proper, strictly convex, steep, and finite-valued over $\mathbb{R}^{n_p}$.*

(ii) *$\nabla\psi_\epsilon^p : \text{int}(\text{dom}(\psi_\epsilon^p)) \to \text{int}(\text{dom}(\psi_\epsilon^{p\star}))$ is a homeomorphism with inverse mapping $(\nabla\psi_\epsilon^p)^{-1} = \nabla\psi_\epsilon^{p\star} = C_\epsilon^p$.*

(iii) *C_ϵ^p is strictly monotone on $\text{int}(\text{dom}(\psi_\epsilon^{p\star}))$.*

Lemma 3. *Let $\psi_\epsilon^p := \epsilon\vartheta^p$, $\epsilon > 0$, where ϑ^p satisfies Assumption 3(i), and let $\psi_\epsilon^{p\star}$ be the convex conjugate of ψ_ϵ^p. Then,*

(i) *$\psi_\epsilon^{p\star} : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{\infty\}$ is closed, proper, convex and finite-valued over $\mathbb{R}^{n_p}$, i.e., $\text{dom}(\psi_\epsilon^{p\star}) = \mathbb{R}^{n_p}$.*

(ii) *$\psi_\epsilon^{p\star}$ is continuously differentiable and $\nabla\psi_\epsilon^{p\star} = C_\epsilon^p$.*

(iii) *C_ϵ^p is surjective from $\mathbb{R}^{n_p}$ onto $\text{rint}(\Omega^p)$ whenever ψ_ϵ^p is steep, and onto Ω^p whenever ψ_ϵ^p is non-steep.*

Proof. (**Proof of Proposition 1**) Let $x, x' \in \Omega$ be two arbitrarily chosen strategies. Consider $\overline{x} = \theta x + (1-\theta)x'$, $\theta \in [0, 1]$. Then the proof for all these claims boil down to the equality condition found in [59, Prop. 1.44], in which it be can shown,

$$(x - x')^\top (U(x) - U(x')) = \int_0^1 (x' - x)^\top \mathfrak{J}_U(\overline{x})(x' - x)\, \mathrm{d}\theta \quad (48)$$

Rearranging (48), we have,

$$\begin{aligned}(x - x')^\top U(x) &= \int_0^1 (x' - x)^\top \mathfrak{J}_U(\overline{x})(x' - x)\, \mathrm{d}\theta \\ &\quad + (x - x')^\top U(x').\end{aligned} \quad (49)$$

Let $x' = x^\star$, then $(x - x^\star)^\top U(x^\star) \leq 0$ and we have,

$$(x - x^\star)^\top U(x) \leq \int_0^1 (x^\star - x)^\top \mathfrak{J}_U(\overline{x})(x^\star - x)\, \mathrm{d}\theta. \quad (50)$$

By the definiteness assumptions of $\mathfrak{J}_U(x)$ on $T_\Omega(x)$ for all $x \in \Omega$ and $\overline{x} = \theta x + (1 - \theta)x^\star$ is contained in Ω for any $\theta \in [0, 1]$, (50) implies statements $(i), (ii), (iii)$. Next, suppose the definiteness condition on $\mathfrak{J}_U(x^\star)$ holds for all $y \in T_\Omega(x^\star)$. Since U is assumed to be continuously differentiable, therefore $y^\top \mathfrak{J}_U(x)y$ is continuous for all $x \in \Omega$. Using the standard $\epsilon - \delta$ definition of continuity, this means for every $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that for all $x \in \Omega$, $\|x - x^\star\| < \delta(\epsilon) \implies |y^\top \mathfrak{J}_U(x)y - y^\top \mathfrak{J}_U(x^\star)y| < \epsilon$. The latter statement, $|y^\top \mathfrak{J}_U(x)y - y^\top \mathfrak{J}_U(x^\star)y| < \epsilon$, implies $y^\top \mathfrak{J}_U(x)y < \epsilon + y^\top \mathfrak{J}_U(x^\star)y$. Since this holds for every $\epsilon > 0$, therefore this is equivalent to $y^\top \mathfrak{J}_U(x)y \leq y^\top \mathfrak{J}_U(x^\star)y$, and hence $\mathfrak{J}_U(x)$ shares the same definiteness property as $\mathfrak{J}_U(x^\star)$ for all x in some ball $\mathcal{D} = \{x \in \Omega | \|x - x^\star\| < \delta(\epsilon)\}$. Let $\overline{x} = \theta x + (1 - \theta)x^\star$,

$\theta \in [0, 1]$, $x \in \mathcal{D}$. This implies (50) holds locally on the set of all $\overline{x}$, and (i'), (ii'), (iii') follows. The isolation condition of (i), (i') are proven in [1, Prop. 2.7]. ∎

Proof. (**Proof of Theorem 1**) (i) Consider the Lyapunov function,

$$V(\xi, z) = \frac{\alpha}{2\beta}\|\xi(t) - x^\star\|_2^2 \qquad (51)$$
$$+ \gamma^{-1} \sum_{p \in \mathcal{N}} \psi_\epsilon^p(x^{p\star}) - z^{p\top} x^{p\star} + \psi_\epsilon^{p\star}(z^p).$$

Here, V is composed of a quadratic distance term, which measures the progress of $\xi(t)$ towards $\xi^\star = x^\star$, added onto a function which consists the collection of all the terms associated with the Fenchel-Young inequality [48, p. 88], also known as the Fenchel coupling in [1]. Since ϑ^p satisfies Assumption 3(i), $V(\xi, z)$ is continuous, positive definite (due to the Legendre theorem), $V(\xi, z) = 0$ iff $\xi = x^\star$ and $z = C_\epsilon^{-1}(x^\star)$. Taking the time-derivative of $V(\xi, z)$ along the solutions of MD2, using $x = C_\epsilon(z)$, $\nabla \psi_\epsilon^{p\star} = C_\epsilon^p$ and $\dot{z} = \gamma(U(x) - \alpha\beta^{-1}\dot{\xi})$,

$$\dot{V}(\xi, z) = \frac{\alpha}{\beta}(\xi - x^\star)^\top \dot{\xi} + \gamma^{-1}(C_\epsilon(z) - x^\star)^\top \dot{z}$$
$$= \frac{\alpha}{\beta}(\xi - x^\star)^\top \dot{\xi} + (C_\epsilon(z) - x^\star)^\top (U(x) - \frac{\alpha}{\beta}\dot{\xi})$$
$$= \frac{\alpha}{\beta}(\xi - x^\star - x + x^\star)^\top \dot{\xi} + (C_\epsilon(z) - x^\star)^\top U(x)$$
$$= \frac{\alpha}{\beta}(\xi - x)^\top \dot{\xi} + (x - x^\star)^\top U(x).$$

Substituting in $\dot{\xi} = \beta(x - \xi)$, we have,

$$\dot{V}(\xi, z) = -\alpha\beta^{-2}\|\dot{\xi}\|_2^2 + (x - x^\star)^\top U(x) \qquad (52)$$
$$\leq -\alpha\beta^{-2}\|\dot{\xi}\|_2^2 \leq 0,$$

where we have used $x^\star$ is a mere VSS. Observe that $\dot{V}(\xi, z) = 0$ only if $\dot{\xi} = \mathbf{0} \iff x = \xi$. By the Legendre theorem [52], $\psi_\epsilon^{p\star}$ is coercive, which implies $V(\xi, z)$ is coercive (radially unbounded) on $\mathbb{R}^n \times \mathbb{R}^n$ and hence all of its sublevel sets are compact. Let $\mathcal{D}_c = \{(\xi, z) \in \mathbb{R}^n \times \mathbb{R}^n | V(\xi, z) \leq c\}$ be a compact sublevel set for some $c > 0$. Consider $\mathcal{E} = \{(\xi, z) \in \mathcal{D}_c | \dot{V}(\xi, z) = 0\} = \{(\xi, z) \in \mathcal{D}_c | \xi = x = C_\epsilon(z)\}$. Let $(\xi(t), z(t))$ be some solution starting in $\mathcal{E}$, then $\xi = x = C_\epsilon(z) \implies \dot{\xi} = \dot{x} = \mathbf{J}_{C_\epsilon}(z)\dot{z} \implies \mathbf{0} = \mathbf{J}_{C_\epsilon}(z)\dot{z}$. By Legendre theorem and the $\mathcal{C}^2$ assumption, $\mathbf{J}_{C_\epsilon}(z)$ exists and is invertible, therefore $\dot{z} = \mathbf{J}_{C_\epsilon}(z)^{-1}\mathbf{0} = \mathbf{0}$. This means the largest invariant set contained in $\mathcal{E}$ is $\mathcal{S} = \{(\xi, z) \in \mathcal{E} | \dot{z} = \mathbf{0}\}$. By LaSalle's invariance principle [42, Theorem 3.3], any solution starting from $\mathcal{D}_c$ converges to $\mathcal{S}$ as $t \to \infty$. On $\mathcal{S}$, $\dot{z} = \mathbf{0} \implies U(x) = U(C_\epsilon(z)) = \mathbf{0}$, which means $x = C_\epsilon(z)$ converges to some interior NE (which we may denote as $x^\star$, since $x^\star$ is arbitrary), which by assumption is a globally mere VSS. Global convergence follows from coercivity of $V(\xi, z)$.

(ii) Suppose instead Ω^p is compact for all players p. Let x^ω be an ω-limit point of $x(t) = C_\epsilon(z(t))$ of MD2. Since $x(t)$ is bounded for all $t \geq 0$, the existence of x^ω follows from Bolzano-Weierstrass theorem. Suppose that x^ω is not a globally mere VSS $x^\star$. Take $\mathcal{O}_\omega$ be an open ball around x^ω. By definition, there exists a sequence $\{x(t_k)\}_{k \in \mathbb{N}}$, $x(t_k) \in \mathcal{O}_\omega$, converging towards x^ω, where $\{t_k\}_{k \in \mathbb{N}}$ is an increasing sequence of times. Following the technique in [19], we wish to build an auxiliary sequence in $\mathcal{O}_\omega$ and show that V is unbounded below along this sequence, thereby obtaining a contradiction.

First, note that since ϑ^p is ρ-strongly convex, $\psi_\epsilon^p := \epsilon\vartheta^p$ is $\epsilon\rho$-strong convex, and hence by the conjugate correspondence theorem [48, Theorem 5.26], $\psi_\epsilon^{p\star}$ is $(\epsilon\rho)^{-1}$-smooth, i.e., $C_\epsilon^p = \nabla \psi_\epsilon^{p\star}$ is $(\epsilon\rho)^{-1}$-Lipschitz. Let $\tau > 0$ be a time increment, then, by $z(t_k + \tau) - z(t_k) = \int_{t_k}^{t_k+\tau} \dot{z}(s)ds = \gamma \int_{t_k}^{t_k+\tau} U(x(s)) - \alpha(x(s) - \xi(s))ds$,

we have,

$$\|x(t_k + \tau) - x(t_k)\| = \|C_\epsilon(z(t_k + \tau) - C_\epsilon(z(t_k))\| \qquad (53)$$
$$\leq (\epsilon\rho)^{-1}\|z(t_k + \tau) - z(t_k)\|_\star \qquad (54)$$
$$\leq (\epsilon\rho)^{-1}\gamma\| \int_{t_k}^{t_k+\tau} U(x(s)) - \alpha(x(s) - \xi(s))ds\|_\star \qquad (55)$$
$$\leq (\epsilon\rho)^{-1}\gamma \int_{t_k}^{t_k+\tau} \|U(x(s))\|_\star + \alpha\|x(s) - \xi(s)\|_\star ds \qquad (56)$$
$$\leq (\epsilon\rho)^{-1}\tau\gamma \max_{x \in \Omega}(\|U(x)\|_\star + \alpha\|x(s) - \xi(s)\|_\star) \leq b(\tau), \qquad (57)$$

where $b(\tau)$ is a τ-dependent upper-bound. Let $\mathcal{B}_\tau(x(t_k)) = \{x \in \Omega | \|x - x(t_k)\| \leq b(\tau), x(t_k) \in \mathcal{O}_\omega\}$ be a ball around $x(t_k)$. Since $\mathcal{O}_\omega$ is open, therefore there exists some $\delta > 0$, such that $\mathcal{B}_\tau(x(t_k)) \subseteq \mathcal{O}_\omega$ for all $\tau \in [0, \delta]$ and hence $x(t_k + \tau) \in \mathcal{O}_\omega$.

Integrating (52), we have,

$$V(\xi(t), z(t)) = V_0 - \int_0^t \alpha\|x - \xi\|_2^2 + (x - x^\star)^\top U(x)d\tau, \qquad (58)$$

where $V_0 := V(\xi(0), z(0))$ is some constant. Since $x^\star$ is assumed to be a globally mere VSS, therefore $(x - x^\star)^\top U(x) \leq 0, \forall x$. Tossing out $(x - x^\star)^\top U(x) \leq 0, \forall x$, (58), we obtain,

$$V(\xi(t), z(t)) \leq V_0 - \int_0^t \alpha\|x(\tau) - \xi(\tau)\|_2^2 d\tau.$$

Expressing this inequality in terms of the sequence $\{x(t_k)\}_{k \in \mathbb{N}}$ and $\{\xi(t_k)\}_{k \in \mathbb{N}}$, $-\|x(\tau) - \xi(\tau)\|_2 \leq 0$, and using the δ which we have previous found, we have,

$$V(\xi(t_k + \delta), z(t_k + \delta)) \leq V_0 - \sum_{\ell=1}^k \int_{t_\ell}^{t_\ell+\delta} \alpha\|x(\tau) - \xi(\tau)\|_2^2 d\tau$$

for some k. Since Ω is compact, x is continuous, ξ is a continuous function of x (15), and ξ is not eventually identically equal to x (otherwise, using the rest point condition for MD2 (14) with our assumption that all interior NE are globally mere, we arrive at $x^\omega = x^\star$ hence a contradiction), $\|x - \xi\|_2^2$ achieves a lower-bound on an interval $[t_\ell, t_\ell + \delta]$. Let $c = \max\{l > 0 | \|x(\tau) - \xi(\tau)\|_2^2 > l, \tau \in [t_\ell, t_\ell + \delta], \ell = 1, \ldots, k\}$. This means,

$$V(\xi(t_k + \delta), z(t_k + \delta)) \leq V_0 - \sum_{\ell=1}^k \int_{t_\ell}^{t_\ell+\delta} \alpha c d\tau \leq V_0 - \alpha c \delta k,$$

which shows that V is unbounded below as $k \to \infty$, a contradiction. ∎

Lemma 4. *Suppose $\vartheta^p : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{\infty\}$ satisfies Assumption 3(i) or (ii). Let $h = \sum_{p \in \mathcal{N}} \vartheta^p$, $x^{p\star} \in \text{dom}(\vartheta^p)$, $x^p \in \text{dom}(\partial\vartheta^p)$, $z^p \in \partial\psi_\epsilon^p(x^p)$, $\psi_\epsilon^p = \epsilon\vartheta^p, \forall p$, then,*

$$D_h(x^\star, x) = \epsilon^{-1} \sum_{p \in \mathcal{N}} \psi_\epsilon^p(x^{p\star}) - \psi_\epsilon^{p\star}(z^p) - z^{p\top} x^{p\star}, \qquad (59)$$

where $x^\star = (x^{p\star})_{p \in \mathcal{N}}, x = (x^p)_{p \in \mathcal{N}}$.

Proof. (**Proof of Lemma 4**) By definition,

$$D_h(x^\star, x) = \sum_{p \in \mathcal{N}} \vartheta^p(x^{p\star}) - \vartheta^p(x^p) - \epsilon^{-1} z^{p\top}(x^{p\star} - x^p) \qquad (60)$$
$$= \sum_{p \in \mathcal{N}} \vartheta^p(x^{p\star}) + \vartheta^{p\star}(\epsilon^{-1}z^p) - \epsilon^{-1}z^{p\top}x^{p\star} \qquad (61)$$
$$= \epsilon^{-1} \sum_{p \in \mathcal{N}} \epsilon\vartheta^p(x^{p\star}) + \epsilon\vartheta^{p\star}(\epsilon^{-1}z^p) - z^{p\top}x^{p\star} \qquad (62)$$
$$= \epsilon^{-1} \sum_{p \in \mathcal{N}} \psi_\epsilon^p(x^{p\star}) + \psi_\epsilon^{p\star}(z^p) - z^{p\top}x^{p\star}, \qquad (63)$$

where we used $\psi_\epsilon^p(x^{p\star}) = \epsilon\vartheta^p(x^{p\star}), z^p \in \partial\psi_\epsilon^p(x^p) = \epsilon\partial\vartheta^p(x^p), \psi_\epsilon^{p\star}(z^p) = \epsilon\vartheta^{p\star}(\epsilon^{-1}z^p), x^p \in \partial\psi_\epsilon^{p\star}(z^p)$. We note that $\psi_\epsilon^{p\star}(z^p) = \epsilon\vartheta^{p\star}(\epsilon^{-1}z^p)$ follows from the conjugate property $\epsilon\vartheta^p(x^p) \iff \epsilon\vartheta^{p\star}(\epsilon^{-1}z^p)$ [48, p. 93]. ∎

Proof. (**Proof of Theorem 2**) Consider the following Lyapunov function,

$$V(\xi, z, \gamma) = \frac{\alpha\gamma}{2\beta}\|\xi - x^\star\|_2^2 + \sum_{p \in \mathcal{N}} \psi_\epsilon^p(x^{p\star}) - z^{p\top}x^{p\star} + \psi_\epsilon^{p\star}(z^p),$$

Then taking the time-derivative of $V(\xi, z, \gamma)$ along MD2γ, and using $\nabla \psi_\epsilon^{p\star} = C_\epsilon^p$, $x^p = C_\epsilon^p(z^p)$, $\dot{z} = U(x) - \gamma\alpha\beta^{-1}\dot{\xi}$, we have,

$$\dot{V}(\xi, z, \gamma) = \frac{\alpha\dot{\gamma}}{2\beta}\|\xi - x^\star\|_2^2 + \frac{\alpha\gamma}{\beta}(\xi - x^\star)^\top \dot{\xi} + (C_\epsilon(z) - x^\star)^\top \dot{z}$$

$$= \frac{\alpha\dot{\gamma}}{2\beta}\|\xi - x^\star\|_2^2 + \frac{\alpha\gamma}{\beta}(\xi - x^\star - x + x^\star)^\top \dot{\xi} + (x - x^\star)^\top U(x)$$

$$= \frac{-\alpha\eta\gamma}{2\beta\epsilon}\|\xi - x^\star\|_2^2 - \frac{\alpha\gamma}{\beta^2}\|\dot{\xi}\|_2^2 + (x - x^\star)^\top U(x)$$

$$\leq \frac{-\alpha\eta\gamma}{2\beta\epsilon}\|\xi - x^\star\|_2^2 + (x - x^\star)^\top U(x)$$

$$\leq \frac{-\alpha\eta\gamma}{2\beta\epsilon}\|\xi - x^\star\|_2^2 - \eta(D_h(x^\star, x) + D_h(x, x^\star))$$

$$\leq \frac{-\alpha\eta\gamma}{2\beta\epsilon}\|\xi - x^\star\|_2^2 - \eta D_h(x^\star, x)$$

$$= -\eta\epsilon^{-1}\left(\frac{\alpha\gamma}{2\beta}\|\xi - x^\star\|_2^2 + \sum_{p\in\mathcal{N}}\psi_\epsilon^p(x^{p\star}) - \psi_\epsilon^{p\star}(z^p) - z^{p\top}x^{p\star}\right)$$

$$= -\eta\epsilon^{-1}V(\xi, z, \gamma), \tag{64}$$

where we have used $D_h(x^\star, x) = \epsilon^{-1}\sum_{p\in\mathcal{N}}\psi_\epsilon^p(x^{p\star}) - \psi_\epsilon^{p\star}(z^p) - z^{p\top}x^{p\star}, \forall x = C_\epsilon(z), z \in C_\epsilon^{-1}(x)$ (Lemma 4). (64) implies $V(\xi, z, \gamma) \leq e^{-\eta\epsilon^{-1}t}V(\xi_0, z_0, \gamma_0)$. Substituting in the exact expression for $V(\xi, z, \gamma)$ and $V(\xi_0, z_0, \gamma_0)$, then rearranging the resulting expression, we obtain (21). (22) follows from $D_h(x^\star, x) \geq 2^{-1}\rho\|x^\star - x\|_2^2$ whenever ϑ^p is ρ-strongly convex. ∎

Proof. (**Proof of Theorem 3**) Without loss of generality, let $\gamma = 1$. Let $y^p \in \Omega^p$ be an arbitrary strategy. Since $\mathcal{U}^p(y^p; x^{-p}(t))$ is concave in y^p, $\mathcal{U}^p(y^p; x^{-p}(t)) \leq \mathcal{U}^p(x^p(t); x^{-p}(t)) + \nabla_{x^p}\mathcal{U}^p(x^p(t); x^{-p}(t))^\top(y^p - x^p(t)) = \mathcal{U}^p(x^p(t); x^{-p}(t)) + U^p(x)^\top(y^p - x^p(t)), \forall t \geq 0$, therefore, the integral term of (23), can be written as (suppressing time-index for brevity),

$$\int_0^t \mathcal{U}^p(y^p; x^{-p}) - \mathcal{U}^p(x)d\tau \leq \int_0^t(y^p - x^p)^\top U^p(x)d\tau, \tag{65}$$

Next, define,

$$Q^p(t) := \psi_\epsilon^{p\star}(z^p(t)) - y^{p\top}z^p(t) + \frac{\alpha}{\beta}\left(\frac{1}{2}\|\xi^p(t)\|_2^2 - y^{p\top}\xi^p(t)\right). \tag{66}$$

Taking the time-derivative of Q^p along the solutions of MD2 and using $\nabla\psi_\epsilon^{p\star} = C_\epsilon^p$, $x^p = C_\epsilon^p(z^p)$, $\dot{z}^p = U^p(x) - \beta^{-1}\alpha\dot{\xi}^p$, yields,

$$\dot{Q}^p(t) = \nabla\psi_\epsilon^{p\star}(z^p)^\top \dot{z}^p - y^{p\top}\dot{z}^p + \frac{\alpha}{\beta}(\xi^{p\top}\dot{\xi}^p - y^{p\top}\dot{\xi}^p) \tag{67}$$

$$= C_\epsilon^p(z^p)^\top \dot{z}^p - y^{p\top}\dot{z}^p + \frac{\alpha}{\beta}(\xi^{p\top}\dot{\xi}^p - y^{p\top}\dot{\xi}^p) \tag{68}$$

$$= (x^p - y^p)^\top \dot{z}^p + \frac{\alpha}{\beta}(\xi^p - y^p)^\top \dot{\xi}^p \tag{69}$$

$$= (x^p - y^p)^\top\left(U^p(x) - \frac{\alpha}{\beta}\dot{\xi}^p\right) + \frac{\alpha}{\beta}(\xi^p - y^p)^\top \dot{\xi}^p \tag{70}$$

$$= (x^p - y^p)^\top U^p(x) + \frac{\alpha}{\beta}(-x^p + \xi^p)^\top \dot{\xi}^p \tag{71}$$

$$= (x^p - y^p)^\top U^p(x) - \frac{\alpha}{\beta^2}\|\dot{\xi}^p\|_2^2. \tag{72}$$

Carrying on from (65) and using (72), we have,

$$\int_0^t \mathcal{U}^p(y^p; x^{-p}) - \mathcal{U}^p(x)d\tau = \int_0^t -\dot{Q}^p(\tau) - \frac{\alpha}{\beta^2}\|\dot{\xi}^p\|_2^2 d\tau \tag{73}$$

$$\leq Q^p(0) - Q^p(t). \tag{74}$$

Applying Fenchel's inequality [52] to the function Q^p, we obtain, $Q^p(t) \geq -\psi_\epsilon^p(y^p) - \frac{\alpha}{2\beta}\|y^p\|_2^2$. Hence $Q^p(0) - Q^p(t) \leq Q^p(0) + \psi_\epsilon^p(y^p) + \frac{\alpha}{2\beta}\|y^p\|_2^2$, where $Q^p(0)$ is some finite constant (due to Lemma 3(i)). Therefore we obtain the following inequality,

$$\int_0^t \mathcal{U}^p(y^p; x^{-p}) - \mathcal{U}^p(x)d\tau \leq Q^p(0) + \psi_\epsilon^p(y^p) + \frac{\alpha}{2\beta}\|y^p\|_2^2. \tag{75}$$

Plugging into (75) the definition of regret, we obtain,

$$\mathcal{R}^p(t) = \max_{y^p\in\Omega^p} t^{-1}\int_0^t \mathcal{U}^p(y^p; x^{-p}(\tau)) - \mathcal{U}^p(x(\tau))d\tau$$

$$\leq \max_{y^p\in\Omega^p} t^{-1}\left(Q^p(0) + \psi_\epsilon^p(y^p) + \frac{\alpha}{2\beta}\|y^p\|_2^2\right). \tag{76}$$

Since Ω^p is assumed to be compact, and the numerator is continuous on all of Ω^p, therefore by Weierstrass theorem [48, Theorem 2.12] a maximum is achieved. Taking the limsup yields the desired result. When $\alpha = 0$, this proof recovers no-regret bound of MD [21]. ∎

Proof. (**Proof of Proposition 2**) By Lemma 3(ii), $x^p = C_\epsilon^p(z^p) = \nabla\psi_\epsilon^{p\star}(z^p)$, therefore $x^p \in \partial\psi_\epsilon^{p\star}(z^p)$. Using the fact that $\partial\psi_\epsilon^p$ is the inverse map of $\partial\psi_\epsilon^{p\star}$, for all $x^p \in \text{dom}(\partial\psi_\epsilon^{p\star}) = \text{rint}(\Omega^p)$,

$$z^p \in \partial\psi_\epsilon^p(x^p) = \{\nabla\psi_\epsilon^p(x^p)\} + N_{\Omega^p}(x^p), \tag{77}$$

From (77), we can write $z^p = \nabla\psi_\epsilon^p(x^p) + n^p(x^p)$ where $n^p(x^p) \in N_{\Omega^p}(x^p)$. Plugging this expression into (25), we have,

$$\begin{cases} \dot{\xi}^p \in \beta(x^p - \xi^p) \\ \dot{x}^p \in \gamma\nabla^2\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p) + n^p(x^p))(U^p(x) - \alpha(x^p - \xi^p)). \end{cases} \tag{78}$$

To reduce (78) from a set of differential inclusions to a set of ODEs, we will show, for all $x^p \in \text{rint}(\Omega^p)$,

$$\nabla^2\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p) + n^p(x^p)) = \nabla^2\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p)). \tag{79}$$

Indeed, $x^p \in \partial\psi_\epsilon^{p\star}(z^p) \iff z^p \in \partial\psi_\epsilon^p(x^p) \overset{(77)}{\iff} \nabla\psi_\epsilon^p(x^p) + n^p(x^p) \in \partial\psi_\epsilon^p(x^p)$. The last equation, $\nabla\psi_\epsilon^p(x^p) + n^p(x^p) \in \partial\psi_\epsilon^p(x^p)$ means, $\forall y^p \in \Omega^p \ \psi_\epsilon^p(y^p) \geq \psi_\epsilon^p(x^p) + (\nabla\psi_\epsilon^p(x^p) + n^p(x^p))^\top(y^p - x^p)$. Since $n^p(x^p) \in N_{\Omega^p}(x^p)$, therefore $n^p(x^p)^\top(y^p - x^p) \leq 0$, hence it is also true that $\psi_\epsilon^p(y^p) \geq \psi_\epsilon^p(x^p) + \nabla\psi_\epsilon^p(x^p)^\top(y^p - x^p)$ or equivalently, $\nabla\psi_\epsilon^p(x^p) \in \partial\psi_\epsilon^p(x^p)$, or $\partial\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p)) = \partial\psi_\epsilon^{p\star}(\partial\psi_\epsilon^p(x^p))$. Hence by re-substituting in (77) and using the single-valuedness of $\partial\psi_\epsilon^{p\star}$, we obtain, $\nabla\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p) + n^p(x^p)) = \nabla\psi_\epsilon^{p\star}(\nabla\psi_\epsilon^p(x^p))$. Taking the Jacobian of the preceding equality yields our desired result. ∎

The proof of Theorem 4 requires the following definitions from [54]–[56]. Recall that a *semiflow* Φ on a metric space $(\mathcal{M}, d)$ is a continuous map $\Phi : \mathbb{T} \times \mathcal{M} \to \mathcal{M}, (t, x) \mapsto \Phi(t, x) = \Phi_t(x)$, where $\mathbb{T} = \mathbb{R}_{\geq 0}$, such that, $\Phi_0 = Id$ and $\Phi_{t+s} = \Phi_t \circ \Phi_s$, for all $(t, s) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$. If $\mathbb{T} = \mathbb{R}$, then Φ defines a flow. We say that a continuous function $f : \mathbb{R}_{>0} \to \mathcal{M}$ is an *asymptotic pseudotrajectory* of Φ if $\lim_{t\to\infty}\sup_{0\leq h\leq T} d(f(t+h), \Phi_h(f(t))) = 0$ for any $T > 0$. A set $\mathcal{A} \subset \mathcal{M}$ is *positively invariant* if $\Phi_t(\mathcal{A}) \subset \mathcal{A}$ for all $t \geq 0$ and *invariant* if $\Phi_t(\mathcal{A}) = \mathcal{A}$ for all $t \in \mathbb{T}$.

Proof. (**Proof of Theorem 4**) For simplicity, we assume that all constants associated with MD2 are set to 1. We break up our proof into the following steps:

1) We first need to show that DA2 is the stochastic approximation of MD2.

   Rearranging the Z_k^p update for DA2, $\frac{Z_{k+1}^p - Z_k^p}{t_{k+1}} = \widehat{U}_{k+1}^p = U^p(X_k) + \zeta_{k+1}^p$ and taking the expectation with respect to the filtration $\mathcal{F}_k = \sigma(\Xi_0, Z_0, \zeta_0, \ldots, \zeta_k)$ yields,

$$\mathbb{E}\left(\frac{Z_{k+1}^p - Z_k^p}{t_{k+1}}\Big|\mathcal{F}_k\right) = \mathbb{E}(U^p(X_k) + \zeta_{k+1}^p - (X_k^p - \Xi_k^p)|\mathcal{F}_k), \tag{80}$$

   Since $\mathbb{E}(U^p(X_k)|\mathcal{F}_k) = U^p(X_k) - (X_k^p - \Xi_k^p)$ and $\mathbb{E}(\zeta_{k+1}^p|\mathcal{F}_k) = 0$ by Assumption 6, therefore,

$$\mathbb{E}\left(\frac{Z_{k+1}^p - Z_k^p}{t_{k+1}}\Big|\mathcal{F}_k\right) = U^p(X_k) - (X_k^p - \Xi_k^p), \tag{81}$$

And similarly for Ξ_{k+1}^p, we have,

$$\mathbb{E}(\frac{\Xi_{k+1}^p - \Xi_k^p}{\tau_{k+1}}|\mathcal{F}_k) = X_k^p - \Xi_k^p, \tag{82}$$

which shows that MD2 is the mean ODE of DA2.

2) Next, we need to build interpolated processes for Z_k and Ξ_k and show that the interpolated processes converge to a rest point of MD2. The full construction process is provided as follows: let $\{\ell_k\}_{k\in\mathbb{N}}$ be a sequence such that,

$$\ell_0 = 0 \quad \ell_k = \sum_{i=1}^k t_i, k \geq 1, \tag{83}$$

and define the interpolated process,

$$\underline{Z}^p(\ell_k + s) = Z_k^p + s\frac{Z_{k+1}^p - Z_k^p}{\ell_{k+1} - \ell_k}, \tag{84}$$

where $0 \leq s < t_{k+1}$. Similarly, let $\{\ell_k'\}_{k\in\mathbb{N}}$ be a sequence such that,

$$\ell_0' = 0 \quad \ell_k' = \sum_{i=1}^k \tau_i, k \geq 1, \tag{85}$$

and define the interpolated process associated with Ξ^p as,

$$\underline{\Xi}^p(\ell_k' + s') = \Xi_k^p + s'\frac{\Xi_{k+1}^p - \Xi_k^p}{\ell_{k+1}' - \ell_k'}, \tag{86}$$

where $0 \leq s' < \tau_{k+1}$.

3) Let $\underline{Z}^p : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_p}$ and $\underline{\Xi}^p : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_p}$ denote the continuous functions (as functions of s, s' respectively) associated with the above processes and let $\underline{Z} = (\underline{Z}^p)_{p\in\mathcal{N}}, \underline{\Xi} = (\underline{\Xi}^p)_{p\in\mathcal{N}}$ be the stacked-vector of all the individual interpolated processes.

4) Next, define the overall process,

$$\underline{W} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{2n} \quad \underline{W} = (\underline{\Xi}, \underline{Z}), \tag{87}$$

Define the interpolated process $\underline{W}$. By Proposition 4.1 and Proposition 4.2 of [54], under Assumption 5 - 8 $\underline{W}$ is an asymptotic pseudo-trajectory of the semiflow $\Phi : \mathbb{R}_{\geq 0} \times \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ induced by $\dot{\omega} = F(\omega)$ (39), that is,

$$\lim_{t\to\infty}\sup_{0\leq h\leq T}\|\underline{W}(t+h) - \Phi(h, \underline{W}(t))\|_2 = 0, \tag{88}$$

for any $T > 0$.

5) By Assumption 7, $\underline{W}$ has compact closure, i.e., is *pre-compact*. Since we have shown that $\underline{W}$ is a pre-compact APT of Φ induced by $\dot{\omega}$, by [54, Theorem 5.7(i)] the limit set

$$L(\underline{W}) = \bigcap_{t\geq 0} \mathrm{cl}(\underline{W}([t, \infty))), \tag{89}$$

is internally chain transitive, which by [54, Prop 5.3], $L(\underline{W})$ is a compact invariant set.

6) Recall that $V : \mathbb{R}^{2n} \to \mathbb{R}$ was used as the Lyapunov function associated with $\dot{\omega} = F(\omega)$ in Theorem 1. Consider the set of critical points of V, $\mathcal{E} = \{(\xi, z) \in \mathbb{R}^{2n} | \dot{V} = 0\} = \{(\xi, z) \in \mathbb{R}^n \times \mathbb{R}^n | \xi = x = C(z)\}$. By the arguments in the proof of Theorem 1, since $V(\xi', z') < V(\xi, z)$ for all $(\xi, z) \in \mathbb{R}^{2n}\backslash\mathcal{E}$, $(\xi', z') = \Phi(t, (\xi, z))$ and $V(\xi', z') \leq V(\xi, z)$ for all $(\xi, z) \in \mathcal{E}, (\xi', z') = \Phi(t, (\xi, z))$, therefore V is a Lyapunov function for $\mathcal{E}$ [55].

7) Since $V(\xi, z)$ is a Lyapunov function for $\mathcal{E} \subset \mathbb{R}^{2n}$, and $V(\mathcal{E})$ is a constant, which implies $\mathrm{int}(V(\mathcal{E})) = \emptyset$, therefore by [55, Prop. 3.27], $L(\underline{W})$ is contained in $\mathcal{E}$. By our assumption that every NE is a mere VSS, hence every point in $\mathcal{E}$ corresponds to an interior mere VSS and $L(\underline{W}) \subset \mathcal{E}$, therefore $L(\underline{W})$ contains a compact subset of the rest points of $\dot{\omega}$.

8) By the definition of a limit set, for any $\underline{W}(0)$, the interpolated process $\underline{W}(t)$ converges as $t \to \infty$.

9) From our construction of the interpolated process and the dimin-

ishing step-size assumption, i.e., $0 \leq s < \tau_{k+1}$, $0 \leq s' < t_{k+1}$ and $\lim_{k\to\infty} t_k = \lim_{k\to\infty} \tau_k = 0$, the convergence of the interpolated processes $\underline{\Xi}$ and $\underline{Z}$ implies the convergence of Ξ_k, Z_k respectively.

10) By continuity of C_ϵ, it follows that $X_k = C_\epsilon(Z_k)$ converges almost surely an interior mere VSS $x^\star$.

∎

REFERENCES

[1] P. Mertikopoulos and Z. Zhou, "Learning in games with continuous action sets and unknown payoff functions," *Mathematical Programming*, vol. 173, no. 1-2, pp. 465-507, 2019.

[2] T. Migot and M.-G. Cojocaru, "On Minty-variational inequalities and evolutionary stable states of generalized monotone games," *Operations Research Letters*, vol. 49, no. 1, pp. 96-100, 2021.

[3] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization (Discrete Mathematics)*. Wiley, 1983.

[4] M. Raginsky and J. Bouvrie, "Continuous-time stochastic mirror descent on a network: Variance reduction, consensus, convergence", in *Proc. IEEE CDC*, pp. 6793-6800, 2012.

[5] W. Krichene, A. Bayen, and P. Bartlett, "Accelerated Mirror Descent in Continuous and Discrete Time", *in NIPS 29*, 2015.

[6] A. Beck, M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Oper. Research Lett.*, vol. 31, no. 3, pp. 167-175, 2003.

[7] T. T. Doan, S. Bose, D. H. Nguyen and C. L. Beck, "Convergence of the Iterates in Mirror Descent Methods," in *IEEE CSL*, vol. 3, no. 1, pp. 114-119, Jan. 2019.

[8] Y. Yu and B. Açıkmeşe, "RLC Circuits-Based Distributed Mirror Descent Method," in *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 548-553, July 2020.

[9] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE TAC*, vol. 57, no. 3, pp. 592-606, Mar. 2012.

[10] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proc. IEEE CDC*, pp. 5453-5458, 2012.

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

[12] J. Hofbauer and W. H. Sandholm, "Stable games and their dynamics," *J. Econ. Theory*, vol. 144, no. 4, pp. 1665-1693, 2009.

[13] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. Cambridge, MA, USA: MIT Press, 2010.

[14] J. Harsanyi, "Oddness of the number of equilibrium points: A new proof," *International Journal of Game Theory*, vol. 2, no. 1, pp. 235-250, 1973.

[15] P. Coucheney, B. Gaujal, and P. Mertikopoulos,"Penalty-regulated dynamics and robust learning procedures in games," *Math. Oper. Res.*, vol. 40, no. 3, pp. 611-633, 2015.

[16] Z. Zhou, P. Mertikopoulos, N. Bambos, P. Glynn, and C. Tomlin, "Countering feedback delays in multi-agent learning," *in Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, pp. 5776-5783, 2017.

[17] Z. Zhou, P. Mertikopoulos, S. Athey, N. Bambos, P. W. Glynn, and Y. Ye, "Learning in games with lossy feedback," in Proc. Advances NIPS, pp. 5140-5150, 2018.

[18] P. Mertikopoulos, W. Sandholm, "Learning in Games via Reinforcement and Regularization", *Mathem. of Oper. Research*, vol. 41, no. 4, pp. 1297-1324, 2016.

[19] P. Mertikopoulos and M. Staudigl, "Convergence to Nash equilibrium in continuous games with noisy first-order feedback," in Proc. 56th IEEE Conf. Decis Control, pp. 5609-5614, 2017.

[20] L. Flokas, E.-V. Vlatakis-Gkaragkounis, T. Lianeas, P. Mertikopoulos, and G. Piliouras, "No-Regret Learning and Mixed Nash Equilibria: They Do Not Mix," in *34th NeurIPS*, 2020.

[21] P. Mertikopoulos, C. Papadimitriou, G. Piliouras, "Cycles in adversarial regularized learning", *ACM-SIAM*, pp. 2703-2717, 2018.

[22] A. Giannou, E.-V. Vlatakis-Gkaragkounis, and P. Mertikopoulos, "Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information," in *Proceedings of MLR*, vol 134, pp. 2147–2148, 2021.

[23] J. Cohen, A. Héliou, and P. Mertikopoulos, "Hedging under uncertainty: Regret minimization meets exponentially fast convergence," in *Algorithmic Game Theory, V. Bilò and M. Flammini*, Eds. Cham, Switzerland: Springer, pp. 252–263, 2017.

[24] A. Kannan and U. V. Shanbhag, "Distributed Computation of Equilibria in Monotone Nash Games via Iterative Regularization Techniques", *SIAM J. on Optimization*, vol. 22, no. 4, pp. 1177-1205, 2012.

[25] J. Koshal, A. Nedich, and U. V. Shanbhag, "Regularized iterative stochastic approximation methods for stochastic variational inequality problems," *IEEE TAC*, vol. 58, pp. 594–609, 2013.

[26] D. Gadjov and L. Pavel, "On the exact convergence to Nash equilibrium in monotone regimes under partial-information," *IEEE CDC*, 2020, pp. 2297-2302.

[27] B. Gao and L. Pavel, "On Passivity, Reinforcement Learning and Higher-Order Learning in Multi-Agent Finite Games", *IEEE TAC*, vol. 66, no. 1, pp. 121-136, 2021.

[28] B. Gao and L. Pavel, "Continuous-Time Discounted Mirror Descent Dynamics in Monotone Concave Games," in *IEEE TAC*, vol. 66, no. 11, pp. 5451-5458, 2021.

[29] B. Gao and L. Pavel, "Continuous-Time Convergence Rates in Potential and Monotone Games," *SIAM Journal on Control and Optimization*, vol. 60, no. 3, pp. 1712–1731, 2022.

[30] M. Sylvestre and L. Pavel, "Q-Learning with Side Information in Multi-Agent Finite Games," in *Proc. of the 58th IEEE Conf. on Decision and Control*, 2019.

[31] S. Amari, *Information Geometry and Its Applications*. Springer Japan, 2016.

[32] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Comp. Math. and Math. Phy.*, vol. 4, no. 5, pp. 1-17, 1964.

[33] R. Laraki and P. Mertikopoulos, "Higher order game dynamics", *Journal of Economic Theory*, vol. 148, no. 6, pp. 2666-2695, 2013.

[34] S. D. Flam and J. Morgan, "Newtonian mechanics and Nash play," *International Game Theory Review*, vol. 6, no. 2, pp. 181-194, 2004.

[35] G. Arslan and J. S. Shamma, "Anticipatory learning in general evolutionary games," *Proceedings of the 45th IEEE CDC*, pp. 6289-6294, 2002.

[36] A. Nagurney and D. Zhang, "Projected dynamical systems in the formulation, stability analysis, and computation of fixed-demand traffic network equilibria," *Transp. Sci.*, vol. 31, no. 2, pp. 147-158, 1997.

[37] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, no. 5427, pp. 15–18, 1973.

[38] J. M. Smith, *Evolution and the Theory of Games*. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[39] H. Attouch, Z. Chbani, J. Fadili and H. Riahi, "First-order optimization algorithms via inertial systems with Hessian driven damping," *Mathematical Programming*, pp. 1-43, 2020.

[40] I. M. J. Sutskever, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. ICML*, pp. 1139-1147, 2013.

[41] A. Krogh and J. A. Hertz, "A simple weight decay can improve generalization", in *Proc. NIPS*, 1992.

[42] W. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control*. Princeton: Princeton University Press, 2011.

[43] C. De Persis and S. Grammatico, "Continuous-time integral dynamics for a class of aggregative games with coupling constraints," *IEEE TAC*, vol. 65, pp. 2171-2176, 2020.

[44] S. Sorin and C. Wan, "Finite composite games: Equilibria and dynamics," *Journal of Dynamics & Games*, vol. 3, p. 101, 2016.

[45] T. Tatarenko and M. Kamgarpour, "Learning Nash Equilibria in Monotone Games," *In Proc. IEEE Conf. Decision and Control*, pp. 3104-3109, 2019.

[46] P. Mertikopoulos, H. Zenati, B. Lecouat, C.-S. Foo, V. Chandrasekhar, and G. Piliouras, "Mirror descent in saddle-point problems: Going the extra (gradient) mile," *ICLR 2019*, pp. 1-23 2019.

[47] E. Mazumdar, L. Ratliff and S. Sastry, "On Gradient-Based Learning in Continuous Games," *SIMODS*, vol. 2, no. 1, pp. 103-131, 2020.

[48] A. Beck, *First-Order Methods in Optimization*, 1st ed. SIAM, 2017.

[49] T. Başar, G. Olsder, *Dynamic noncooperative game theory*. SIAM, 1999.

[50] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*. Vol.I & II, Springer-Verlag, NY, 2003.

[51] J. Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games", *Econometrica*, vol. 33, no. 3, p. 520, 1965.

[52] J. M. Borwein, H. H. Bauschke, "Legendre functions and the method of random Bregman projections," *J. of Convex Analysis*, vol. 4, pp. 27-67, 1997.

[53] C. Daskalakis, A. Ilyas, V. Syrgkanis and H. Zeng, "Training GANs with Optimism", In Proc. *6th Int. Conf. on Learning Representation*, 2018.

[54] M. Benaim, "Dynamics of stochastic approximation algorithms," in *Séminaire de Probabilités XXXIII (Lecture Notes in Mathematics)*, vol. 1709. Berlin, Germany: Springer-Verlag, pp. 1-68, 1999.

[55] M. Benaim, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions," *SIAM J. Control Optimiz.*, vol. 44, no. 1, pp. 328–348, 2005.

[56] M. Benaim and M. Faure, "Stochastic approximation, cooperative dynamics and supermodular games," *Annals of App. Prob.*, vol. 22, no. 5, pp. 2133-2164, 2012.

[57] M. Benaim, J. Hofbauer, and E. Hopkins, "Learning in games with unstable equilibria," *Journal of Economic Theory*, vol. 144, pp. 1694-1709, 2009.

[58] T. Tatarenko, "Stochastic learning in multi-agent optimization: Communication and payoff-based approaches," *Automatica*, vol. 99, pp. 1–12, 2019.

[59] I. Menache and A. Ozdaglar, "Network Games: Theory, models, and dynamics," *Synthesis Lectures Commun. Netw.*, vol. 4, no. 1, pp. 1–159, Mar. 2011.

[60] S. Bervoets, M. Bravo, and M. Faure, "Learning with minimal information in continuous games," *Theoretical Economics*, vol. 15, no. 4, pp. 1471-1508, 2020.

[61] U. Ghat, E. Hazan, and Y. Singer, "Exponentiated gradient meets gradient descent", in *ALT*, vol. 117, pp. 386-407, 2020.

[62] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 3rd ed. Heidelberg: Springer, 2009.

[63] A. Fiacco, "Perturbed variations of penalty function methods. Example: Projective SUMT," *Annals of Operations Research*, vol. 27, no. 1, pp. 371-380, 1990.

Bolin Gao received the B.A.Sc. (with Hons.) and M.A.Sc. degrees in electrical engineering in 2015 and 2017, respectively, from the University of Toronto, Toronto, ON, Canada, where he is currently working toward the Ph.D. degree with the Systems Control Group. His research interests include design and control of game algorithms with applications to reinforcement learning in multiagent systems.

Lacra Pavel (M'92 - SM'04) received the Dipl. Engineer from Technical University of Iasi, Romania and the Ph.D. degree in Electrical Engineering from Queen's University, Canada. After a postdoctoral stage at the National Research Council and four years of industry experience, in 2002 she joined University of Toronto, where she is now a Professor in the Systems Control Group, Department of Electrical and Computer Engineering. Her research interests are in game theory and distributed optimization in networks, with emphasis on dynamics and control. She is the author of the book Game Theory for Control of Optical Networks (Birkhäuser-Springer Science, ISBN 978-0-8176-8321-4, 2012). She is a Senior Editor of IEEE Transactions on Control of Network Systems, a Senior Editor of the IEEE Open Journal of Control Systems and a Member of the Conference Editorial Board of the IEEE Control Systems Society; she acted as Publications Chair of the 45th IEEE Conference on Decision and Control.