Oblivious Equilibrium for General Stochastic Games with Many Players

Vineet Abhishek, Sachin Adlakha, Ramesh Johari and Gabriel Weintraub

***Abstract*—This paper studies a solution concept for large stochastic games. A standard solution concept for a stochastic game is *Markov perfect equilibrium* (MPE). In MPE, each player's optimal action depends on his own state and the state of the other players. By contrast, *oblivious equilibrium (OE)* is an approximation introduced in [5] where each player makes decisions based on his own state and the "average" state of the other players. For this reason OE is computationally more tractable than MPE. It was shown in [5] that as the number of players becomes large, OE closely approximates MPE; however, this result was established under a set of assumptions specific to industry dynamic models. In this paper we develop a parsimonious set of assumptions under which the result of [5] can be generalized to a broader class of stochastic games with a large number of players.**

I. Introduction

Markov perfect equilibrium is a commonly used equilibrium concept for stochastic games [1]. It has been widely used to analyze interactions in dynamic systems with multiple players with competing objectives. In MPE, strategies of players depend only on the current state of all players, and not on the past history of the game. In general, finding an MPE is analytically intractable; MPE is typically obtained numerically using dynamic programming (DP) algorithms [2]. As a result, the complexity associated with MPE computation increases rapidly with the number of players, the size of the state space, and the size of the action sets [3]. This limits its application to problems with small dimensions.

The economics literature has used MPE extensively in the study of models of industry dynamics with heterogeneous firms, notably as proposed in the seminal work of [4]. However, MPE computation for the proposed model is nontrivial. Recently, a scheme for approximating MPE in such models was proposed in [5], via a novel solution concept called *oblivious equilibrium*. In oblivious equilibrium, a firm optimizes given only the long-run *average* industry statistics, rather than the entire instantaneous vector of its competitors' state. Clearly, OE computation is significantly simpler than MPE computation, since each firm only needs to solve a one-dimensional dynamic program. When there are a large number of firms, individual firms have a small impact on the aggregate industry statistics, provided that no firm is "large" relative to the entire market (refered to as a "light-tail" condition by [5]). It is reasonable to expect that under such a condition, if firms make decisions based only on the long-run industry average, they should achieve near-optimal performance. Indeed, it is established in [5] that under a reasonable set of technical conditions (including the "light-tail" condition), OE is a good approximation to MPE for industry dynamic models with many firms; formally, this is called the *asymptotic Markov equilibrium* (AME) property.

This paper presents a generalization of the approximation result of [5]. As presented in [5], the main approximation result is tailored to the class of firm competition models presented there. However, in principle OE can be defined for any class of stochastic games where the number of players grows large in an appropriate sense. Our main contribution is to isolate a parsimonious set of assumptions for a general class of stochastic games with many players, under which the main result of [5] continues to hold: namely, that OE is a good approximation to MPE. Because our assumptions generalize those in [5], the technical arguments are similar to those in [5]; in some cases the arguments are in fact simplified due to the more general game class considered.

The rest of the paper is organized as follows. In Section II, we outline our model of a stochastic game, notation, and definitions. In Section III, we introduce the AME property and the formal light-tail condition. In Section IV, we prove the main theorem of this paper using a series of lemmas. We conclude in Section V.

II. Model, Definitions and Notations

We consider an m-player stochastic game evolving over discrete time periods with an infinite horizon. The discrete time periods are indexed with non-negative integers $t \in \mathbb{N}$. The state of the player i at time t is denoted by $x_{i,t} \in \mathcal{X}$, where $\mathcal{X}$ is a totally ordered (possibly infinite-dimensional) discrete metric space.

Let Θ be the finite set of types, and corresponding to a type $\theta \in \Theta$, let π^θ be the non-negative type-dependent single period payoff function. We assume that state evolution for a player i with type θ_i depends only on its own current state and the action it takes. This can be represented by a type dependent conditional probability mass function (pmf)

$$x_{i,t+1} \sim h^{\theta_i}\left(x|x_{i,t}, a_{i,t}\right) \quad \text{for some} \quad \theta_i \in \Theta, \tag{1}$$

where $a_{i,t}$ is the action taken by the player i at time t. We denote the set of actions by $\mathcal{A}$. The single period payoff to player i with type θ_i is given as $\pi^{\theta_i}\left(x_{i,t}, a_{i,t}, \boldsymbol{x}_{-i,t}\right)$. Here $\boldsymbol{x}_{-i,t}$ is the state of all players except player i at time t. Note that the payoff to player i does not depend on the actions

V. Abhishek and S. Adlakha are with the Department of Electrical Engineering, Stanford University. {avineet, adlakha}@stanford.edu
R. Johari is with the Department of Management Science and Engineering, Stanford University. ramesh.johari@stanford.edu
G. Weintraub is with the Columbia Business School, Columbia University. gweintraub@columbia.edu

taken by other players. Furthermore, we assume that the payoff function is independent of the identity of other players. That is, it only depends on the current state $x_{i,t}$ of the player i, the total number m of the players at any time, and the fraction $f_{-i,t}^{(m)}(y)$, which is the fraction of the players excluding player i, that have their state as y. In other words, we can write the payoff function as $\pi^{\theta_i}\left(x_{i,t}, a_{i,t}, f_{-i,t}^{(m)}, m\right)$, where, θ_i is the type of the player i, and $f_{-i,t}^{(m)}$ can be expressed as

$$f_{-i,t}^{(m)}(y) \triangleq \frac{1}{m-1}\sum_{j\neq i}\mathbf{1}_{\{x_{j,t}=y\}}. \tag{2}$$

Each player i chooses an action $a_{i,t} = \mu^{m,\theta_i}(x_{i,t}, f_{-i,t}^{(m)})$ to maximize his expected present value. Note that the policy μ^{m,θ_i} depends on the type θ_i of the player and m because of the underlying dependence of the payoff function and the state evolution on θ_i and m. Let $\boldsymbol{\mu}^m$ be the vector of policies of all players, and $\boldsymbol{\mu}_{-i}^m$ be the vector of policies of all players except player i. We define $V^{\theta_i}(x, f, m|\mu^{m,\theta_i}, \boldsymbol{\mu}_{-i}^m)$ to be the expected net present value for player i with current state x, if the current aggregate state of players other than i is f, given that i follows the policy μ^{m,θ_i} and the policy vector of players other than i is given by $\boldsymbol{\mu}_{-i}^m$. In particular, we have

$$V^{\theta_i}(x, f, m|\mu^{m,\theta_i}, \boldsymbol{\mu}_{-i}^m) \triangleq \mathbb{E}\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\pi^{\theta_i}(x_{i,\tau}, a_{i,\tau}, f_{-i,\tau}^{(m)}, m)\,\middle|\, x_{i,t}=x, f_{-i,t}^{(m)}=f; \mu^{m,\theta_i}, \boldsymbol{\mu}_{-i}^m\right], \tag{3}$$

where $0<\beta<1$ is the discount factor. Note that the random variables $(x_{i,t}, f_{i,t}^{(m)})$ depend on the policy vector $\boldsymbol{\mu}^m$ and the state evolution function h.

We focus on *symmetric Markov perfect equilibrium*, where all firms with the same type θ use the same policy $\mu^{m,\theta}$. Let $\mathcal{M}^\theta$ be the set of all policies available to a player of type θ. Note that this set also depends on total number of players m.

Definition 1 (Markov Perfect Equilibrium): The vector of policies $\boldsymbol{\mu}^m$ is a *Markov perfect equilibrium* if for all j, we have

$$\sup_{\mu'\in\mathcal{M}^{\theta_j}} V^{\theta_j}\left(x, f, m|\mu', \boldsymbol{\mu}_{-j}^m\right) = V^{\theta_j}\left(x, f, m|\mu^{m,\theta_j}, \boldsymbol{\mu}_{-j}^m\right)\ \forall x, f.$$

The analysis of [5] approximates the MPE using a form of the law of large numbers: as the number of players becomes large, the changes in the players' states average out such that the state vector is well approximated by its long run average. In this case, each player can find his optimal policy based solely on his own state and the long-run average aggregate state of the other players. We therefore restrict attention to policies that are only a function of the player's own state, and an underlying constant aggregate distribution of the competitors. Such strategies are referred to as *oblivious strategies* since they do not take into account the complete state of the competitors at any time. Let us denote $\tilde{\mu}^{m,\theta_i}$ as an oblivious policy of player i with type θ_i; we let $\tilde{\mathcal{M}}^\theta$ denote the set of all oblivious policies available to a player of type θ. This set also depends on the number of players m. Note that if all players use oblivious strategies, their states evolve as independent Markov chains. We make the following assumption regarding the Markov chain of each player playing an oblivious policy.

Assumption 1: The Markov chain associated with the state evolution of each player i (with type θ_i) playing an oblivious policy $\tilde{\mu}^{m,\theta_i}$ is positive recurrent, and reaches a stationary distribution $q^{\tilde{\mu}^{m,\theta_i}}$.

Let $\tilde{\boldsymbol{\mu}}^m$ be the vector of oblivious policies for all players, $\tilde{\mu}^{m,\theta_i}$ be the oblivious policy for player i, and $\tilde{\boldsymbol{\mu}}_{-i}^m$ be the vector of oblivious policies of all player except player i. For simplification of analysis, we assume that the initial state of a player i is sampled from the stationary distribution $q^{\tilde{\mu}^{m,\theta_i}}$ of its state Markov chain; without this assumption, the OE approximation holds only after sufficient mixing of the individual players' state evolution Markov chains. Given $\tilde{\boldsymbol{\mu}}^m$, for a particular player i, the long-run average aggregate state of its competitors is denote by $\tilde{f}_{-i}^{(m)}$, and is defined as

$$\tilde{f}_{-i}^{(m)}(y) \triangleq \mathbb{E}\left(f_{-i,t}^{(m)}(y)\right) = \frac{1}{m-1}\sum_{j\neq i} q^{\tilde{\mu}^{m,\theta_j}}(y). \tag{4}$$

Note that, $\tilde{f}_{-i}^{(m)}$ is completely determined by the state evolution function function h and oblivious policy $\tilde{\boldsymbol{\mu}}_{-i}^m$.

As with the case of symmetric MPE defined above, we assume that players with the same type use the same oblivious policy. Let $\tilde{\mu}^{m,\theta}$ denote the oblivious policy employed by all players of type θ; note that then $\tilde{f}_{-i}^{(m)}$ is identical for all such players i, so we abbreviate this as $\tilde{f}^{(m),\theta}$. We define the *oblivious value function* $\tilde{V}^{\theta_i}(x, m|\tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m)$ to be the expected net present value for player i with type θ_i and current state x, if player i follows the oblivious policy $\tilde{\mu}^{m,\theta_i}$, and players other than i follow the oblivious policy vector $\tilde{\boldsymbol{\mu}}_{-i}^m$. Specifically, we have

$$\tilde{V}^{\theta_i}(x, m|\tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m) \triangleq \mathbb{E}\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\pi^{\theta_i}\left(x_{i,\tau}, a_{i,\tau}, \tilde{f}_{-i}^{(m)}, m\right)\,\middle|\, x_{i,t}=x; \tilde{\mu}^{m,\theta_i}\right]. \tag{5}$$

Note that the expectation does not depend explicitly on the policies used by players other than i; this dependence only enters through the long-run average aggregate state $\tilde{f}_{-i}^{(m)}$. In particular, the state evolution is *only* due to the policy of player i. Using the oblivious value function, we define oblivious equilibrium as follows.

Definition 2 (Oblivious Equilibrium): The vector of policies $\tilde{\boldsymbol{\mu}}^m$ represents an *oblivious equilibrium* if for all j, we have

$$\sup_{\mu'\in\tilde{\mathcal{M}}^{\theta_j}} \tilde{V}^{\theta_j}\left(x, m|\ \mu', \tilde{\boldsymbol{\mu}}_{-j}^m\right) = \tilde{V}^{\theta_j}\left(x, m|\ \tilde{\mu}^{m,\theta_j}, \tilde{\boldsymbol{\mu}}_{-j}^m\right),\ \forall x.$$

Compared to [5], our model has a more general action-dependent payoff function. In [5], action is in form of choice of investment and appears as an additive form in the payoff function for all players. We also allow the possibility of heterogeneity in state evolution and payoff functions. Finally, in [5], many assumptions on the payoff functions are required primarily to establish the stationarity of the underlying Markov chain. We abstract this by assuming existence of stationary distribution; verification of this assumption must be done on an application-by-application basis.

In this paper, we do not show the existence of Markov perfect equilibrium or of oblivious equilibrium. We assume that both the equilibrium points exist for the stochastic game under consideration [1].

Assumption 2: Markov perfect equilibrium and oblivious equilibrium exist for the stochastic game under consideration.

III. Asymptotic Markov Equilibrium and the Light-Tail Condition

The main result of [5] is that under mild conditions, MPE can be approximated by OE in models of industry dynamics with a large number of firms. In this section, we generalize the key assumptions used in that paper, so that we can develop a similar result for general stochastic games.

We begin by describing the asymptotic Markov equilibrium (AME) property; intuitively, this property says that an oblivious policy is approximately optimal even when compared against Markov policies. Formally, the AME ensures that as number of players in the game becomes large, the approximation error between the expected net present value obtained by deviating from the oblivious policy $\tilde{\mu}^{m,\theta}$ and instead following the optimal (non-oblivious) policy goes to zero for each state x of the player.

Definition 3 (Asymptotic Markov Equilibrium (AME)): A sequence of oblivious policies $\tilde{\boldsymbol{\mu}}^m$ possesses the asymptotic Markov equilibrium (AME) property if for all x and i, we have

$$\lim_{m\to\infty} \mathbb{E}\left[\sup_{\mu'\in\mathcal{M}^{\theta_i}} V^{\theta_i}\left(x,f,m \mid \mu', \tilde{\boldsymbol{\mu}}^m_{-i}\right) - V^{\theta_i}\left(x,f,m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}^m_{-i}\right)\right] = 0.$$

Notice that the expectation here is over the aggregate state of players other than i, denoted by f. MPE requires the error to be zero for all (x,f), rather than in expectation; of course, in general, it will not be possible to find a single oblivious policy that satisfies the AME property for any f. In particular, in OE, actions taken by a player will perform poorly if the other players' state is far from the long-run average aggregate state. Thus, AME implies that the OE policy performs nearly as well as the non-oblivious best policy for those aggregate states of other players that occur with high probability.

[1] In general MPE is not unique. As stated in [5], there are likely to be fewer OE than MPE, though no general result is known.

In order to establish AME, we make the following assumptions on the payoff functions π^θ. For notational convenience, we drop the subscripts i,t whenever it does not lead to any ambiguity. Also we abbreviate $\pi^\theta(x,a,f^{(m)},m)$ to be the payoff function for all players j with type $\theta_j=\theta$.

Assumption 3: We assume that the payoff function is uniformly bounded. That is

$$\sup_{x,a,f^{(m)},m} \pi^\theta(x,a,f^{(m)},m) < \infty \quad \forall\theta.$$

Assumption 4: We assume that the payoff π^θ is Gateaux differentiable with respect to $f^{(m)}(y)$. That is for all θ, if

$$\sum_y \Delta f^{(m)}(y)\left|\frac{\partial \pi^\theta(x,a,f^{(m)},m)}{\partial f^{(m)}(y)}\right| < \infty,$$

then

$$\left.\frac{\partial \pi^\theta\left(x,a,f^{(m)}+\gamma\Delta f^{(m)},m\right)}{\partial\gamma}\right|_{\gamma=0} = \sum_y \Delta f^{(m)}(y)\frac{\partial\pi^\theta\left(x,a,f^{(m)},m\right)}{\partial f^{(m)}(y)}.$$

We now proceed to define the *light-tail condition* formally. We define $g^\theta(y)$ as

$$g^\theta(y) \triangleq \sup_{x,a,f^{(m)},m}\left|\frac{\partial\pi^\theta(x,a,f^{(m)},m)}{\partial f^{(m)}(y)}\right| \qquad (6)$$

and make following assumption on $g^\theta(y)$.

Assumption 5 (Light-Tail): We assume that $g^\theta(y)$ is finite for all θ and y. Also, given $\epsilon>0$, $\forall\ \theta$, there exists a state value z^θ, such that

$$\mathbb{E}\left[g^\theta(\tilde{U}^{(m)})\mathbf{1}_{\tilde{U}^{(m)}>z^\theta}|\tilde{U}^{(m)}\sim\tilde{f}^{(m),\theta}\right] \le \epsilon, \quad \forall m. \qquad (7)$$

Here $\tilde{U}^{(m)}$ is a random variable distributed according to $\tilde{f}^{(m),\theta}$. The function $g^\theta(y)$ can be interpreted as the maximum rate of change of the single period payoff of any player, with respect to a small change in the fraction of competitors at any state value y. The first part of the assumption implies that this is finite. The second part of the assumption requires that the probability of competitors at larger y (tail probability) should go to zero quickly uniformly over m. The quantity $\mathbb{E}\left[g^\theta(\tilde{U}^{(m)})\mathbf{1}_{\tilde{U}^{(m)}>z^\theta}|\tilde{U}^{(m)}\sim\tilde{f}^{(m),\theta}\right]$ captures the effect of competitors at a higher state on the single period payoff of a player.

To summarize, our development to this point has led to five assumptions on our model:

1) Positive recurrence of the state evolution under oblivious policies (Assumption 1);
2) Existence of MPE and OE (Assumption 2);
3) Uniform boundedness of the payoff function of each player (Assumption 3);
4) Gateaux differentiability of the payoff function of each player (Assumption 4); and
5) The light-tail condition on the payoff function of each player (Assumption 5).

IV. Asymptotic Results for Oblivious Equilibrium

In this section, we prove the AME property using a series of lemmas; our technical development is similar to that in [5], but is streamlined by the use of a parsimonious set of assumptions. *Assumptions 1-5 are kept throughout the remainder of the paper.*

Lemma 1: For all x and θ, we have

$$\sup_{m,\mu^{m,\theta}\in\mathcal{M}^\theta} \mathbb{E}\Big[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\sup_f \pi^\theta\left(x_{i,\tau},a_{i,\tau},f,m\right) \\ |x_{i,t}=x\Big]<\infty.$$

Proof: Follows trivially from assumptions 3. ■

The next lemma shows that whenever two aggregate states f and $f^{'}$ are *close* to each other, the single period payoffs of any player under these two aggregate states are also close to each other. For an appropriate metric for the distance between two distributions, we define the $1-g$ norm as follows:

$$\|f\|_{1,g^\theta} \triangleq \sum_y |f(y)|g^\theta(y).$$

The $1-g$ norm puts higher weights on those states of competitors where a slight change in the fractional distribution of competitors causes a large change in the payoff function. Hence, it measures the distance between two distributions in terms of their effect on the payoff function.

Lemma 2: For all θ and for all f, f' such that $\|f^{(m)}\|_{1,g^\theta}<\infty$ and $\|f'^{(m)}\|_{1,g^\theta}<\infty$, the following holds

$$\left|\pi^\theta(x,a,f^{(m)},m)-\pi^\theta(x,a,f^{'(m)},m)\right| \\ \leq \left\|f^{(m)}-f^{'(m)}\right\|_{1,g^\theta}.$$

Proof: By the assumptions on the payoff function, for any x, a, $f, f^{'}$ and m, we have,

$$\begin{aligned}&\left|\pi^\theta(x,a,f^{(m)},m)-\pi^\theta(x,a,f^{'(m)},m)\right| \\ &=\left|\int_{\alpha=0}^{1}\frac{\partial\pi^\theta\left(x,a,f^{(m)}+\alpha(f^{(m)}-f^{'(m)}),m\right)}{\partial\alpha}d\alpha\right|, \\ &=\Bigg|\int_0^1\sum_y\left(f^{(m)}(y)-f^{'(m)}(y)\right)\cdot \\ &\qquad\left(\frac{\partial\pi^\theta\left(x,a,f^{(m)}+\alpha(f^{(m)}-f^{'(m)}),m\right)}{\partial\left(f^{(m)}+\alpha(f^{(m)}-f^{'(m)})\right)(y)}\right)d\alpha\Bigg|, \\ &\leq\int_0^1\sum_y\left|f^{(m)}(y)-f^{'(m)}(y)\right|g^\theta(y)d\alpha, \\ &=\left\|f^{(m)}-f^{'(m)}\right\|_{1,g^\theta}.\end{aligned}$$

■

The next lemma shows that, as $m\to\infty$, the distribution of the aggregate state $f^{(m)}$ approaches its mean $\tilde{f}^{(m)}$ in the $1,g^\theta$ norm define above. Here, both $f^{(m)}$ and $\tilde{f}^{(m)}$ are defined over same oblivious policy vector, $\tilde{\boldsymbol{\mu}}^m$.

Lemma 3: For all i with $\theta_i=\theta$, $\left\|f^{(m)}_{-i,t}-\tilde{f}^{(m),\theta}\right\|_{1,g^\theta}\to 0$ in probability as $m\to\infty$.

Proof: We can write

$$\left\|f^{(m)}_{-i,t}-\tilde{f}^{(m),\theta}\right\|_{1,g^\theta}=\sum_y g^\theta(y)\left|f^{(m)}_{-i,t}(y)-\tilde{f}^{(m),\theta}(y)\right|.$$

Now, let $\epsilon>0$ be given and let z^θ be such that the light-tail condition in (7) is satisfied for the given ϵ. Then,

$$\begin{aligned}\left\|f^{(m)}_{-i,t}-\tilde{f}^{(m),\theta}\right\|_{1,g^\theta}&\leq\underbrace{z^\theta\max_{y\leq z^\theta}g^\theta(y)\left|f^{(m)}_{-i,t}(y)-\tilde{f}^{(m),\theta}(y)\right|}_{\equiv A_z^{(m)}} \\ &+\underbrace{\sum_{y>z^\theta}g^\theta(y)f^{(m)}_{-i,t}(y)}_{\equiv B_z^{(m)}}+\underbrace{\sum_{y>z^\theta}g^\theta(y)\tilde{f}^{(m),\theta}(y)}_{\equiv C_z^{(m)}}.\end{aligned}$$

By the light-tail assumption, for any $\epsilon>0$ and sufficiently large z^θ, we have $C_z^{(m)}\leq\epsilon$. Hence, $\mathcal{P}\left[C_z^{(m)}>\epsilon\right]\to 0$ as $m\to\infty$.

Also, $\mathbb{E}\left[B_z^{(m)}\right]=C_z^{(m)}$ and hence by Markov inequality we have, for any $\delta>0$ and $\epsilon>0$, and sufficiently large z^θ, $\mathcal{P}\left[B_z^{(m)}>\delta\right]<\frac{\epsilon}{\delta}$. Hence, $\limsup_{m\to\infty}\mathcal{P}\left[B_z^{(m)}>\delta\right]=0$.

Now,

$$\begin{aligned}&\mathbb{E}\Big(f^{(m)}_{-i,t}(y)-\tilde{f}^{(m)}_{-i}(y)\Big)^2 \\ &=\frac{1}{(m-1)^2}\mathbb{E}\left(\sum_{j\neq i}\mathbf{1}_{\{x_{j,t}=y\}}-\mathbb{E}\left(\sum_{j\neq i}\mathbf{1}_{\{x_{j,t}=y\}}\right)\right)^2, \\ &=\frac{1}{(m-1)^2}\sum_{j\neq i}\text{Var}\left(\mathbf{1}_{\{x_{j,t}=y\}}\right), \\ &\leq\frac{1}{4(m-1)}\to 0 \text{ as m}\to\infty.\end{aligned}$$

since the random variable $\mathbf{1}_{\{x_{j,t}=y\}}$ is a Bernoulli random variable with $\mathbb{E}\left[\mathbf{1}_{\{x_{j,t}=y\}}\right]=q^{\theta_j}(y)$ and $\text{Var}\left(\mathbf{1}_{\{x_{j,t}=y\}}\right)=q^{\theta_j}(y)(1-q^{\theta_j}(y))\leq\frac{1}{4}$. Hence, $A_z^{(m)}\to 0$ in probability as $m\to\infty$. ■

The next lemma relates the present expected payoff when a player uses a policy $\mu^{m,\theta}$.

Lemma 4: For all x,μ^{m,θ_i} and θ_i,

$$\lim_{m\to\infty}\mathbb{E}\Bigg[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\Big|\pi^{\theta_i}\left(x_{i,\tau},a_{i,\tau},f^{(m)}_{-i,\tau},m\right)- \\ \pi^{\theta_i}\left(x_{i,\tau},a_{i,\tau},\tilde{f}^{(m)}_{-i},m\right)\Big|\ \Big|\ x_{i,t}=x;\mu^{m,\theta_i},\tilde{\boldsymbol{\mu}}^m_{-i}\Bigg]=0.$$

Proof: Let us define

$$\Delta^{m,\theta}_{i,t}\triangleq\left|\pi^{\theta_i}\left(x_{i,t},a_{i,t},f^{(m)}_{-i,t},m\right)-\pi^{\theta_i}\left(x_{i,t},a_{i,t},\tilde{f}^{(m)}_{-i},m\right)\right|.$$

For any $\delta > 0$, let us define $Z^{m,\theta}$ to be the event that $\left\|f_{-i,\tau}^{(m)}(y) - \tilde{f}_{-i}^{(m)}\right\|_{1,g^\theta} \geq \delta$. Then, we can write

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\Delta_{i,t}^{m,\theta}\right] \leq \left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\mathbb{E}\left(\Delta_{i,t}^{m,\theta}\right)\right], \\
&= \sum_{\tau=t}^{\infty}\beta^{\tau-t}\mathbb{E}\left(\Delta_{i,t}^{m,\theta}\mathbf{1}_{\neg Z^{m,\theta}}\right) + \sum_{\tau=t}^{\infty}\beta^{\tau-t}\mathbb{E}\left(\Delta_{i,t}^{m,\theta}\mathbf{1}_{Z^{m,\theta}}\right), \\
&\leq \frac{\delta}{1-\beta} + \sum_{\tau=t}^{\infty}\beta^{\tau-t}\mathbb{E}\left(\Delta_{i,t}^{m,\theta}\mathbf{1}_{Z^{m,\theta}}\right),
\end{aligned}$$

where the last inequality follows from Lemma 2. Now, $\Delta_{i,t}^{m,\theta} \leq 2\sup_f \pi^\theta(x, a, f^{(m)}, m)$. This implies that the second term in the above equation can be written as

$$\begin{aligned}
&\sum_{\tau=t}^{\infty}\beta^{\tau-t}\mathbb{E}\left(\Delta_{i,t}^{m,\theta}\mathbf{1}_{Z^{m,\theta}}\right) \\
&\leq 2\sum_{\tau=t}^{\infty}\beta^{\tau-t}\mathbb{E}\left(\sup_f \pi^\theta(x_{i,\tau}, a_{i,\tau}, f, m)\mathbf{1}_{Z^{m,\theta}}\right), \\
&\leq 2\sup_{\mu^{m,\theta}\in\mathcal{M}^\theta}\mathbb{E}\left(\sum_{\tau=t}^{\infty}\beta^{\tau-t}\sup_f \pi^\theta(x_{i,\tau}, a_{i,\tau}, f, m)\mathbf{1}_{Z^{m,\theta}}\right), \\
&= 2\mathcal{P}\left(Z^{m,\theta}\right)\sup_{\mu^{m,\theta}\in\tilde{\mathcal{M}}^\theta}\mathbb{E}\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\sup_f \pi^\theta(x_{i,\tau}, a_{i,\tau}, f, m)\right],
\end{aligned}$$

where the last equality follows because $\sup_{\mu^{m,\theta}}$ is attained by an oblivious policy and $x_{i,\tau}$ and $f_{-i,\tau}^{(m)}$ are independent. By Lemma 3, $\mathcal{P}\left(Z^{m,\theta}\right) \to 0$ as $m \to \infty$. This along with Lemma 1 gives the desired result. ■

Theorem 5 (Main Theorem): A sequence of oblivious equilibrium policies $\tilde{\boldsymbol{\mu}}^m$ satisfies the AME property. That is, for all i, x, we have

$$\begin{aligned}
\lim_{m\to\infty}\mathbb{E}\Big[\sup_{\mu'\in\mathcal{M}^{\theta_i}} &V^{\theta_i}\left(x, f, m \mid \mu', \tilde{\boldsymbol{\mu}}_{-i}^m\right) - \\
&V^{\theta_i}\left(x, f, m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right)\Big] = 0.
\end{aligned}$$

Proof: Let $\hat{\mu}^{m,\theta_i}$ be the (non-oblivious) optimal best response to $\tilde{\boldsymbol{\mu}}_{-i}^m$. That is for $\mu' \in \mathcal{M}^{\theta_i}$,

$$\sup_{\mu'} V^{\theta_i}\left(x, f, m \mid \mu', \tilde{\boldsymbol{\mu}}_{-i}^m\right) = V^{\theta_i}\left(x, f, m \mid \hat{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right).$$

Let us also define

$$\begin{aligned}
\Delta V^{\theta_i} \triangleq V^{\theta_i}&\left(x, f, m \mid \hat{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) - \\
&V^{\theta_i}\left(x, f, m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right).
\end{aligned}$$

Then we need to show that for all x, θ_i, $\lim_{m\to\infty}\mathbb{E}\left[\Delta V^{\theta_i}\right] = 0$. We can write ΔV^{θ_i} as

$$\begin{aligned}
\Delta V^{\theta_i} &= V^{\theta_i}\left(x, f, m|\hat{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) - \tilde{V}^{\theta_i}\left(x, m|\tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) \\
&+\tilde{V}^{\theta_i}\left(x, m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) - V^{\theta_i}\left(x, f, m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right), \\
&\leq V^{\theta_i}\left(x, f, m|\hat{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) - \tilde{V}^{\theta_i}\left(x, m|\hat{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) \\
&+\tilde{V}^{\theta_i}\left(x, m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right) - V^{\theta_i}\left(x, f, m \mid \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\right), \\
&\equiv T_1 + T_2.
\end{aligned}$$

The inequality follows since $\tilde{\mu}^{m,\theta_i}$ maximizes the oblivious value function. To prove the AME property, we need to show that $\mathbb{E}[T_1]$ and $\mathbb{E}[T_2]$ converge to zero as m becomes large. Using triangle inequality, we can write

$$\begin{aligned}
\mathbb{E}[T_1] \quad \leq \quad &\mathbb{E}\Big[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\Big|\pi^{\theta_i}\left(x_{i,\tau}, a_{i,\tau}, f_{-i,\tau}^{(m)}, m\right) - \\
&\pi^{\theta_i}\left(x_{i,\tau}, a_{i,\tau}, \tilde{f}_{-i}^{(m)}, m\right)\Big| \mid x_{i,t} = x; \hat{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\Big], \\
\mathbb{E}[T_2] \quad \leq \quad &\mathbb{E}\Big[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\Big|\pi^{\theta_i}\left(x_{i,\tau}, a_{i,\tau}, \tilde{f}_{-i}^{m}, m\right) - \\
&\pi^{\theta_i}\left(x_{i,\tau}, a_{i,\tau}, f_{-i,\tau}^{(m)}, m\right)\Big| \mid x_{i,t} = x; \tilde{\mu}^{m,\theta_i}, \tilde{\boldsymbol{\mu}}_{-i}^m\Big].
\end{aligned}$$

Here, expectation is also over the aggregate initial state f of competitors. Lemma 4 then implies the result. ■

Thus, for any type $\theta \in \Theta$, the AME property holds. Since $|\Theta| < \infty$, for a given x as $m \to \infty$, the AME property holds uniformly for all θ and hence for all players.

V. Conclusion

As an extension to the work done in [5], we have shown that the OE solution concept can be applied to a general class of stochastic games. Under certain mild technical conditions, the AME property holds and OE can be used for MPE computation. This allows analysis of problems with high dimension (large number of players) where MPE computation is intractable.

For the special case of a discrete-time infinite-horizon stochastic game with finite state space, the light tail condition automatically follows, and hence only Assumptions 1-3 are sufficient to imply AME.

VI. Acknowledgments

The authors gratefully acknowledge helpful conversations with Lanier Benkard and Benjamin Van Roy. This work was supported by the Stanford Clean Slate Internet Program, by the Defense Advanced Research Projects Agency under the Information Theory for Mobile Ad Hoc Networks (ITMANET) program, and by the National Science Foundation under grant number 0620811.

References

[1] L. S. Shapley, "Stochastic games," *Proceeding of the National Academy of Sciences*, 39, pp. 1095-1100, 1953.

[2] A. Pakes and P. McGuire, "Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model," *RAND Journal of Economics* 25(4), pp. 555 589, 1994.

[3] A. Pakes and P. McGuire, "Stochastic algorithms, symmetric Markov perfect equilibrium, and the curse of dimensionality," *Econometrica* 69(5), pp. 1261 - 1281, 2001.

[4] R. Ericson and A. Pakes, "Markov-perfect industry dynamics: A framework for empirical work," *Review of Economic Studies* 62(1), pp. 53 - 82, 1995.

[5] G. Y. Weintraub, L. C. Benkard, and B. Van Roy, "Markov perfect industry dynamics with many firms," *submitted for publication*.