

The Nash Certainty Equivalence Principle and McKean-Vlasov Systems: An Invariance Principle and Entry Adaptation

Minyi Huang Peter E. Caines Roland P. Malhamé

Abstract—We study large population stochastic dynamic games where the so-called Nash certainty equivalence based control laws are implemented by the individual players. We first show a martingale property for the limiting control problem of a single agent and then perform averaging across the population; this procedure leads to a constant value for the martingale which shows an invariance property of the population behavior induced by the Nash strategies. The situation of a new agent joining the population is also analyzed.

I. INTRODUCTION

Noncooperative dynamic game theory has attracted a long lasting research interest for decades. In such games, the states of the players (also called agents) are governed by certain dynamics and each agent chooses its strategy in a process of interaction with other players. The most basic formulation takes the form of two-person zero-sum dynamic games, and the classical solution notion is saddle strategies [15]. By extending to a general N -person situation, one can assign each player with its own cost function and adopt Nash equilibrium strategies as a basic solution scheme.

For noncooperative dynamic games, when each agent has perfect state information for all agents involved, a well known approach is to study feedback Nash strategies and employ dynamic programming to examine necessary conditions for the associated strategies and individual costs [1]. In general, the complexity of this approach is high in the case of many players, and the Nash equilibrium strategies, if existing, involves high implementational complexity since each player needs the state information of all other players.

On the other hand, in many social, economic and engineering scenarios [8], [9], [10], [17], [16], it is typical to have a large number of agents performing decision-making, and a characteristic feature of these systems is that each agent faces the average influence of the overall population while receiving a negligible impact from any other specific agent; also, relevant vaccination game models arise in public health research [2], [5]. Motivated by these phenomena, in our earlier work we formulated a class of stochastic dynamic games with many players and weak coupling. For obtaining low complexity solutions, the so-called Nash certainty equivalence (NCE) methodology has been developed in a series of works [10], [12], [13], [14], [11]. The key idea of this methodology is to specify a certain consistency relationship

between the individual strategies and the mass effect (i.e., the overall effect of the population on a given agent) within the population limit, and each decision-maker can ignore the fine details of the behavior of any individual player by only focusing on the overall impact of the population. This procedure leads to decentralized strategies for the individual players in a large but finite population.

In the NCE methodology, each agent essentially solves a local optimal control problem. And on the other hand, in the stochastic optimal control literature, it is well known that under very mild conditions, the sum of the past cost calculated up to the current time along the optimal state-control and the future optimal cost is given as a martingale [4], [6]. In contrast to the usual HJB equation characterization of the value function, this martingale representation reveals sample path properties for the optimally controlled process.

By extending the martingale results in optimal control to the population limit of the dynamic game and averaging across the population, one obtains a deterministic martingale and hence a constant value over time. This gives the so-called invariance property involving the instantaneous cost, the value function and the empirical distribution function of the states of all agents. In other words, we obtain an invariance property associated with the controlled population behavior when the NCE strategies are implemented.

In a further step, we address the issue of time-varying population on an infinite time horizon. We will focus on the LQG case and analyze the entry adaptation of a new player joining a group.

II. THE STOCHASTIC DYNAMIC GAME MODEL

In a population of N agents, consider the dynamics for an individual agent

$$dz_i(t) = (1/N) \sum_{j=1}^N f_{a_i}(z_i(t), u_i(t), z_j(t)) dt + \sigma dw_i(t), \quad (1)$$

$$1 \leq i \leq N, \quad t \geq 0,$$

where $\{w_i, 1 \leq i \leq N\}$ denotes N independent standard scalar Wiener processes and $a_i \in \mathbb{R}$ is a dynamic parameter indicating the type of agent i . The state variable z_i and control u_i are each a scalar, and the initial states $\{z_i(0), 1 \leq i \leq N\}$ are mutually independent and also independent of $\{w_i, 1 \leq i \leq N\}$. In addition, $E|z_i(0)|^2 < \infty$. The diffusion coefficient $\sigma > 0$ is a constant. The associated cost function is given as

$$J_i = E \int_0^T (1/N) \sum_{j=1}^N g(z_i(t), u_i(t), z_j(t)) dt. \quad (2)$$

M. Huang is with School of Math. & Stat., Carleton Univ., Ottawa, ON K1S 5B6, Canada. mhuang@math.carleton.ca.

P.E. Caines is with Dept. Electr. & Comput. Eng., McGill Univ., Montreal, QC H3A 2A7, Canada. peterc@cim.mcgill.ca.

R.P. Malhamé is with Dept. Electr. Eng., École Poly. de Montréal, Montreal, QC H3C 3A7, Canada. roland.malhamé@polymtl.ca.

We note that our analysis may be easily generalized to deal with variants of the set of costs (2). For the system configuration $z = (z_1, \dots, z_N)$, define the empirical distribution $\varepsilon_z = (1/N) \sum_{i=1}^N \delta_{z_i}$ where δ_{\cdot} is the Dirac measure. We introduce a set of costs of the form

$$\tilde{J}_i = E \int_0^T \tilde{g}(z_i(t), u_i(t), \varepsilon_z(t)) dt, \quad (3)$$

where $\varepsilon_z(t) = (1/N) \sum_{i=1}^N \delta_{z_i(t)}$. Here $\tilde{g}(\cdot, \cdot, \cdot)$ may be represented as a function from $\mathbb{R} \times U \times \mathbb{R}^N$ such that under the permutation of all other entries in (z_1, \dots, z_N) except z_i , \tilde{g} remains the same value when (z_i, u_i) is given. It can be checked that (3) includes the cost (2) as a special case.

For simplicity of exposition, in this paper, the general formulation of the game will be based upon the cost (2).

Let $u_i(\cdot)$ denote a control input on $[0, T]$, and $u_i(t)$ its value at time t from a closed set $U \subset \mathbb{R}$. Each $u_i(\cdot)$ is adapted to the σ -algebra $\sigma(z_i(0), w_i(s), s \leq t, 1 \leq i \leq N)$.

The sequence $\{a_i, i \geq 1\}$ takes values from a finite set $\mathcal{A} \triangleq \{\theta_1, \dots, \theta_K\}$, and denote the empirical distribution function

$$F_N(\{\theta_k\}) = (1/N) \sum_{i=1}^N 1_{(a_i=\theta_k)}.$$

We assume the existence of a limit empirical distribution function F on \mathcal{A} for the sequence $\{F_N, N \geq 1\}$ and denote $\pi = (\pi_1, \dots, \pi_K)$ where $\pi_k = F(\{\theta_k\}) = \lim_{N \rightarrow \infty} F_N(\{\theta_k\})$.

A. Interacting Particle Systems

In an interacting particle (IP) system, the state evolution of an individual particle is affected by an empirical average of coupling terms involving all other particles. Mathematically, this leads to a set of weakly coupled diffusions, each describing the motion of a single particle. We introduce the following dynamics [7] in the form of N coupled stochastic differential equations (SDE):

$$dx_i(t) = (1/N) \sum_{k=1}^N b(x_i(t), x_k(t)) dt + \sigma dw_i(t), \quad (4)$$

$$1 \leq i \leq N, \quad t \geq 0,$$

where $b(\cdot, \cdot)$ is a function from \mathbb{R}^2 to \mathbb{R} , N is the number of particles and all x_i 's have i.i.d. initial conditions at $t = 0$. Here we assume x_i is a scalar although the modelling is also applicable to the case of vector particle states. The noises $\{w_i, 1 \leq i \leq N\}$ are N independent Wiener processes independent of the initial conditions $x_i(0)$, $1 \leq i \leq N$.

For this class of particle models, heuristically one may approximate the aggregate coupling term in terms of an expectation over a typical individual's probability distribution function which evolves with time. This is based upon the following intuition: as the number of particles grows to infinity, the collective impact of all particles on a given particle is averaged into a deterministic effect. More specifically, as N tends to infinity, the individual dynamics may be written in the limiting form:

$$dx(t) = b[x(t), \mu_t] dt + \sigma dw(t), \quad t \geq 0, \quad (5)$$

which is the celebrated McKean-Vlasov (M-V) equation [18]. Here $b[x, \nu_t] = \int b(x, y) \nu_t(dy)$ for a probability distribution ν_t on \mathbb{R} . The noise $w(t)$ may be determined in different ways. For instance, if we intend to approximate $x_1(t)$ in (4) by $x(t)$, we may set $w(t) = w_1(t)$ and $x(0) = x_1(0)$ in (5).

Definition 1: A pair $(x(t), \mu_t)$, $t \geq 0$, is called a consistent pair if $x(t)$ is a solution to the SDE (5) and μ_t is its distribution at time t , i.e., $P(x(t) \leq \alpha) = \int_{(-\infty, \alpha]} \mu_t(dy)$ for all $\alpha \in \mathbb{R}$ and $t \geq 0$. \square

By introducing the density function $p_t(x)$ for μ_t (or $x(t)$) in a consistent pair, one may recast (5) in the form of a Fokker-Planck equation whose coefficients depend upon $p_t(x)$ itself.

B. Related Notation for the Large Population Game

Let $\mu_t^o \triangleq (\mu_t^1, \dots, \mu_t^K)$ for $0 \leq t \leq T$, where each μ_t^k is a probability distribution on \mathbb{R} . Define $f_a[x, u, \mu_t^o] = \sum_{i=1}^K \pi_i \int_{\mathbb{R}} f_a(x, u, y) \mu_t^i(dy)$, and $g[x, u, \mu_t^o] = \sum_{i=1}^K \pi_i \int_{\mathbb{R}} g(x, u, y) \mu_t^i(dy)$. Here $f_a[x, u, \mu_t^o]$ and $g[x, u, \mu_t^o]$ may be viewed as functions from $\mathbb{R} \times U \times [0, T]$ to \mathbb{R} .

III. THE NASH CERTAINTY EQUIVALENCE BASED STRATEGIES

The basic idea for circumventing the dimensionality difficulty to the game problem (1)-(2) is as follows. Similar to the approximation of weakly coupled diffusions by the McKean-Vlasov equation, we use K probability measures to approximate the effect of K classes of agents. Next, the dynamics and cost of agent i can be approximated using its own state, control and the above K probability measures. Note that these measures are only assumed at this stage and must still be determined. We complete the procedure by requiring that the individual local optimal reactions will collectively produce the same set of probability measures in the closed-loop system; this is essentially to extend the consistency relationship in the original McKean-Vlasov equation to the optimal control situation. We call this scheme of control synthesis as the NCE methodology.

Following [14], we introduce the NCE based equation system

$$dx_i(t) = f_{a_i}[x_i(t), \hat{u}_i(t, x_i(t)), \mu_t^o] dt + \sigma dw_i(t), \quad (6)$$

$$-\frac{\partial v_{a_i}(t, x_i)}{\partial t} = \min_{u_i \in U} \left\{ f_{a_i}[x_i, u_i, \mu_t^o] \frac{\partial v_{a_i}(t, x_i)}{\partial x_i} + g[x_i, u_i, \mu_t^o] \right\}$$

$$+ \frac{\sigma^2}{2} \frac{\partial^2 v_{a_i}(t, x_i)}{\partial x_i^2}, \quad (7)$$

$$\hat{u}_i(t, x_i) \triangleq \varphi_{a_i}(t, x_i)$$

$$= \arg \min_{u_i \in U} \left\{ f_{a_i}[x_i, u_i, \mu_t^o] \frac{\partial v_{a_i}(t, x_i)}{\partial x_i} + g[x_i, u_i, \mu_t^o] \right\}, \quad (8)$$

where $0 \leq t \leq T$ and μ_t^k denotes the distribution of $x_i(t)$ in (6) when $a_i = \theta_k$. In (7), $v_{a_i}(t, x_i)$ is the value function for the associated optimal control problem in the population limit.

An alternative method to characterize the closed-loop

behavior is to express (6) via a Fokker-Planck equation:

$$\frac{\partial p_{a_i}(t, x_i)}{\partial t} = - \frac{\partial \{f_{a_i}[x_i, \hat{u}_i(t, x_i), \mu_i^o] p_{a_i}(t, x_i)\}}{\partial x_i} + \frac{\sigma^2}{2} \frac{\partial^2 p_{a_i}(t, x_i)}{\partial x_i^2}, \quad (9)$$

where $p_{a_i}(t, x_i)$, $0 \leq t \leq T$, denotes the density function for $x_i(t)$. Note that in (9), the density $p_{a_i}(t, x_i)$ should generate the distribution μ_i^k (as a component in μ_i^o) if $a_i = \theta_k$. The description by (9) avoids sample path dependent information.

To distinguish from the original game model, a different variable x_i is used in (6), but the same driving noise w_i and initial condition $z_i(0)$ are used in both (1) and (6). We note that (6) is interpreted in the generalized McKean-Vlasov equation sense for multi-class particles, and the control law \hat{u}_i is a minimizer of the right hand side of (7). A solution to the above McKean-Vlasov-HJB system consists of a triple $(x_i(t), v_{a_i}(t, x_i), \hat{u}_i(t, x_i))$, $t \geq 0$, representing the closed-loop solution, the value function and the feedback control law.

A detailed analysis is developed in [14] about the existence of a solution to the above equation system. This amounts to a fixed point argument with certain nonlinear operators. In this paper, we will focus on the large population behavior subject to the optimal control laws of the individual agents which are assumed to exist.

IV. A MARTINGALE PROPERTY OF THE POPULATION

A. The Optimal Control Problem

Before considering the game problem, as a preliminary step, we first give a brief review of results on martingale representation in optimal stochastic control. Consider the control problem for a single agent

$$dz(t) = f(z(t), u(t))dt + \sigma dw(t), \quad 0 \leq t \leq T, \quad (10)$$

where $z(0)$ is independent of the standard Wiener process $w(t)$ and satisfies $E|z(0)|^2 < \infty$. The cost function is $J_0 = E \int_0^T g(z(t), u(t))dt$. For simplicity, we restrict the analysis to a scalar state $x(t)$ and control $u(t)$ taking values from a closed subset U of \mathbb{R} . A control $u(\cdot)$ is called admissible if $u(t) \in U$ and is adapted to the σ -algebra $\sigma(z(0), w(s), s \leq t)$. Assume f is Lipschitz with respect to (z, u) and $g \geq 0$ satisfies a polynomial growth condition with respect to (z, u) .

Let $v(t, y) = \inf_{u(\cdot)} E \left[\int_t^T g(z(s), u(s))ds | z(t) = y \right]$, where $t \in [0, T]$, be the value function. In addition, we assume that there exists an optimal control law $u^*(t, z) \in C([0, T] \times \mathbb{R})$ satisfying Lipschitz continuity in z , such that

$$v(t, y) = E \left[\int_t^T g(z(s), u^*(s, z(s)))ds | z(t) = y \right].$$

After assuming the existence of the optimal control law, Lemma 2 below essentially follows from the optimality principle. Its proof is similar to the method in [4].

Lemma 2: The process $\xi_t \triangleq \int_0^t g(z(s), u^*(s, z(s)))ds + v(t, z(t))$ is a martingale, where $t \in [0, T]$ and $z(t)$ is the closed-loop solution when the control law u^* is applied. \square

Note that, based upon our assumption on f and u^* , the closed-loop system for (10) has a unique strong solution.

B. The Martingale Property and Its Limiting Form

We now extend the martingale property to the controlled McKean-Vlasov equation. Before so doing, we need to make the existence and growth rate assumptions.

(A1) There exists a solution $(x_i(t), v_{a_i}(t, x_i), \hat{u}_i(t, x_i))$ to the McKean-Vlasov-HJB system (6)-(8). \square

(A2) The closed-loop drift coefficient $f_{a_i}(x_i, \hat{u}_i(t, x_i))$ is in $C([0, T] \times \mathbb{R})$ and Lipschitz continuous in x_i . \square

(A3) Under the control \hat{u}_i , $g[x_i, \hat{u}_i(t, x_i), \mu_i^o]$ is in $C([0, T] \times \mathbb{R})$ and has a polynomial growth rate with respect to x_i . \square

Theorem 3: Suppose (A1)-(A3) hold. Then the process $\int_0^t g[x_i(s), \hat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t, x_i(t))$ is a martingale.

Proof: For $t \in [0, T]$, let $\zeta_t = \int_0^t g[x_i(s), \hat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t, x_i(t))$, and \mathcal{F}_t be the σ -algebra generated by $(x_i(s), s \leq t)$. We have

$$\begin{aligned} E[\zeta_t | \mathcal{F}_{t_1}] &= E \left[\int_0^{t_2} g[x_i(s), \hat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t_2, x_i(t_2)) | \mathcal{F}_{t_1} \right] \\ &= E \left[\zeta_{t_1} + \int_{t_1}^{t_2} g[x_i(s), \hat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t_2, x_i(t_2)) \right. \\ &\quad \left. - v_{a_i}(t_1, x_i(t_1)) | \mathcal{F}_{t_1} \right]. \end{aligned}$$

On the other hand, given μ_s^o , \hat{u}_i may be interpreted as a usual optimal control. Hence, by the optimality principle, we have $v_{a_i}(t_1, x_i(t_1))$

$$= E \left[\int_{t_1}^{t_2} g[x_i(s), \hat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t_2, x_i(t_2)) | \mathcal{F}_{t_1} \right].$$

Hence $E[\zeta_t | \mathcal{F}_{t_1}] = \zeta_{t_1}$ and this completes the proof. \square

Theorem 4: In addition to (A1)-(A3), we assume i.i.d. initial conditions $\{x_i(0), 1 \leq i \leq N\}$ with $E|x_i(0)|^k < \infty$ for any finite $k > 0$. Then for all $t \in [0, T]$, the process

$$\Lambda_t^N \triangleq \frac{1}{N} \sum_{i=1}^N \int_0^t g[x_i(s), \hat{u}_i(s, x_i(s)), \mu_s^o]ds + \frac{1}{N} \sum_{i=1}^N v_{a_i}(t, x_i(t))$$

converges in L_2 to a constant value c , as $N \rightarrow \infty$. \square

Proof: For μ_i^o satisfying (6)-(8), let μ_i^j , $1 \leq i \leq N$, be its components. Theorem 3 implies

$$\begin{aligned} c &= \int_0^t \int_{\mathbb{R}^2} \sum_{i,j=1}^K \pi_i \pi_j g(x, \varphi_{\theta_i}(s, x), y) \mu_s^j(dy) \mu_s^i(dx) ds \\ &\quad + \int_{\mathbb{R}} \sum_{i=1}^K \pi_i v_{\theta_i}(t, x) \mu_t^i(dx), \quad 0 \leq t \leq T, \quad (11) \end{aligned}$$

where c is determined by the initial condition of the population. Then the theorem follows by standard estimates and the fact that $\{x_i(\cdot), i \geq 1\}$ are independent processes. \square

Define $\bar{g}_t = \int_{\mathbb{R}^2} \sum_{i,j} \pi_i \pi_j g(x, \varphi_{\theta_i}(s, x), y) \mu_s^j(dy) \mu_s^i(dx)$ and $\bar{v}_t = \int_{\mathbb{R}} \sum_{i=1}^K \pi_i v_{\theta_i}(t, x) \mu_t^i(dx)$. Then it follows that

$$\frac{d\bar{v}_t}{dt} = -\bar{g}_t, \quad 0 \leq t \leq T.$$

We may interpret \bar{g}_t as the instantaneous average cost over the population limit at time t and \bar{v}_t is the average future cost to go across the population. Thus (11) shows how the forward propagation of the state distribution should preserve this equality relation in terms of the instantaneous cost and the future cost to go based upon the population averaging.

C. The LQG System

Now we specialize the analysis to an LQG game. First, let the individual dynamics be given as

$$dz_i(t) = a_i z_i(t) dt + b u_i(t) dt + (\alpha/N) \sum_{j=1}^N z_j(t) dt + \sigma dw_i(t),$$

$$1 \leq i \leq N, \quad (12)$$

and the cost for the i th agent is given as

$$J_i = E \int_0^T \left\{ [z_i(t) - (\gamma/N) \sum_{j=1}^N (z_j(t) + \eta)]^2 + r u_i^2(t) \right\} dt. \quad (13)$$

For simplicity, we assume the independent initial states $z_i(0)$ have zero mean. For the dynamic parameters we make the more general assumption that all $a_i \in \mathcal{A}^0$ where \mathcal{A}^0 is a compact subset of \mathbb{R} . The dynamic parameters have a limit empirical distribution, still denoted by $F(a)$.

For (12)-(13), we introduce the auxiliary control problem:

$$\begin{cases} dz_i(t) = a_i z_i(t) dt + b u_i(t) dt + \alpha \bar{z}(t) dt + \sigma dw_i(t) \\ J_i = E \int_0^T \{ [z_i(t) - z^*(t)]^2 + r u_i^2(t) \} dt \end{cases} \quad (14)$$

where $\bar{z}, z^* \in C[0, T]$. The tracking problem (14) may be easily solved [3]. Write the Riccati differential equation

$$\frac{d\Pi_i(t)}{dt} + 2a_i \Pi_i(t) - \frac{b^2}{r} \Pi_i^2(t) + 1 = 0 \quad (15)$$

with terminal condition $\Pi_i(T) = 0$. Also write the equation

$$\frac{ds_i(t)}{dt} + (a_i - \frac{b^2}{r} \Pi_i(t)) s_i(t) + \alpha \Pi_i(t) \bar{z}(t) - z^*(t) = 0, \quad (16)$$

where the terminal condition is $s_i(T) = 0$.

Finally, we introduce the equation

$$\frac{dq_i(t)}{dt} - \frac{b^2}{r} s_i^2(t) + |z^*(t)|^2 + 2\alpha \bar{z}(t) s_i(t) + \sigma^2 \Pi_i(t) = 0, \quad (17)$$

where $q_i(T) = 0$.

The optimal control law is given as

$$u_i(t) = -\frac{b}{r} (\Pi_i(t) z_i(t) + s_i(t)).$$

It can be verified that the resulting individual optimal cost is

$$v_i(t, x) = \Pi_i(t) x^2 + 2s_i(t) x + q_i(t)$$

when the initial time-state pair is (t, x) with $t \in [0, T]$.

Using the solution to the auxiliary tracking problem (14), the NCE scheme may be expressed in a more explicit form:

$$\frac{ds_a(t)}{dt} + (a - \frac{b^2}{r} \Pi_a(t)) s_a(t) + \alpha \Pi_a(t) \bar{z}(t) - z^*(t) = 0, \quad (18)$$

$$\frac{d\bar{z}_a(t)}{dt} = (a - \frac{b^2}{r} \Pi_a(t)) \bar{z}_a(t) - \frac{b^2}{r} s_a(t) + \alpha \bar{z}(t), \quad (19)$$

$$\bar{z}(t) = \int_{a \in \mathcal{A}^0} z_a(t) dF(a), \quad (20)$$

$$z^*(t) = \gamma(\bar{z}(t) + \eta). \quad (21)$$

Compared with the original NCE equation system (6)-(8), here we only need the dynamics of the mean process of

the closed-loop for the individual agents, instead of the controlled diffusion process or its Fokker-Planck equation.

We give a sufficient condition to ensure the existence and uniqueness of a solution to the above equation system. For $a \in \mathcal{A}^0$, let $\Phi_a(t, s) = \exp \left\{ -\int_s^t (a - \frac{b^2}{r} \Pi_a(\tau)) d\tau \right\}$.

Theorem 5: There exists a unique solution to (18)-(21) if

$$\sup_{t \in [0, T]} \frac{b^2}{r} \int_{a \in \mathcal{A}^0} \int_0^T \Phi_a(\kappa, t) \times \left\{ \int_{\kappa}^T \Phi_a(\kappa, \tau) [|\alpha| \Pi_a(\tau) + |\gamma|] d\tau + |\alpha| \right\} d\kappa dF(a) < 1,$$

where F is the limit empirical distribution and $a \in \mathcal{A}^0$. \square

Now we end this section by computation with the concrete parameter set $\mathcal{A}^0 = \{\theta_1, \dots, \theta_K\}$ where $F(\{\theta_k\}) = \pi_k$, $1 \leq k \leq K$. In equations (15)-(17), when $a_i = \theta_k$, we denote the resulting solutions, respectively, by $\Pi_{\theta_k}(t)$, $s_{\theta_k}(t)$ and $q_{\theta_k}(t)$. Then the invariance property translates into the form

$$\begin{aligned} c = \sum_{k=1}^K \pi_k \left\{ \int_0^t \int_{\mathbb{R}} \left[(x - z^*(\tau))^2 + \frac{b^2}{r} (\Pi_{\theta_k}(\tau) x + s_{\theta_k}(\tau))^2 \right] \right. \\ \left. \times dF_{\theta_k}^{\tau}(x) d\tau \right. \\ \left. + \int_{\mathbb{R}} [x^2 \Pi_{\theta_k}(t) + 2x s_{\theta_k}(t)] dF_{\theta_k}^t(x) + q_{\theta_k}(t) \right\}, \quad (22) \end{aligned}$$

where $F_{\theta_k}^{\tau}(x)$ denotes the state distribution at time τ for an agent with dynamic parameter θ_k .

For the state distribution above, we assume the existence of a density $p'_{\theta_k}(x)$ with suitable regularity; then by taking differentiation with respect to t on both sides of (22), we get

$$\begin{aligned} 0 = \sum_{k=1}^K \pi_k \left\{ \int_{\mathbb{R}} \left[(x - z^*(t))^2 + \frac{b^2}{r} (\Pi_{\theta_k}(t) x + s_{\theta_k}(t))^2 \right] p'_{\theta_k}(x) dx \right. \\ \left. + \int_{\mathbb{R}} \frac{\partial \{ [x^2 \Pi_{\theta_k}(t) + 2x s_{\theta_k}(t)] p'_{\theta_k}(x) \}}{\partial t} dx + \frac{dq_{\theta_k}(t)}{dt} \right\}. \quad (23) \end{aligned}$$

Due to the quadratic cost structure, the NCE scheme (18)-(21) gives little information directly about the variance, or spread, of the population behavior, but (23) provides such information by placing specific constraints on the propagation of the state distribution.

V. VARIABLE POPULATION WITH RANDOM ENTRY TIME

For the random entry problem we consider an LQG model with uniform agents. Since the individual cost is defined on an infinite horizon, a discounted cost integrand is used.

We suppose that a system of N agents starts the game from the initial time $t = 0$. Now assume the $(N+1)$ th agent joins at time τ_e (away from the time origin of the system), but it does not know the value of τ_e although it may know its statistical properties. At the system level, τ_e is treated as a random variable and its value may be observed. However, this value is generally unknown to the agent in question. Such a situation easily arises in many practical scenarios. To illustrate this point, we consider a queuing model where the $(N+1)$ th customer arrives with an exponentially distributed time gap after the previous customer; but the new customer does not know the value of the time gap unless the existing customers can communicate that information to it.

Hence in this setting, the main problem is for the new agent to adapt with respect to its random entry time. We may term this as a clock free adaptation problem in that the new comer has no information about the time origin at which other players started their game.

A. Switched Dynamics due to Agent Entry

We assume that the entry time τ_e is bounded by a constant and is a Markov time, i.e., for any $t \geq 0$, $\{\tau_e \leq t\} \in \mathcal{F}_t \triangleq \mathcal{F}(z_i(0), w_i(\tau), 1 \leq i \leq N, \tau \leq t)$. For $1 \leq i \leq N$, let the individual dynamics be

$$dz_i = \begin{cases} az_i dt + bu_i dt + \alpha z^{(N)} dt + \sigma dw_i, & \text{for } t < \tau_e, \\ az_i dt + bu_i dt + \alpha z^{(N+1)} dt + \sigma dw_i, & \text{for } t \geq \tau_e, \end{cases}$$

where $z^{(N)} = 1/N \sum_{i=1}^N z_i$ and $z^{(N+1)} = 1/(N+1) \sum_{i=1}^{N+1} z_i$. Thus, starting from τ_e , the previously existing agents will receive a small impact from the new agent with its dynamics:

$$dz_{N+1} = az_{N+1} dt + bu_{N+1} dt + \alpha z^{(N+1)} dt + \sigma dw_{N+1}, \quad t \geq \tau_e,$$

where w_{N+1} is a Brownian motion (with the initial time $t = \tau_e$) which is independent of (w_1, \dots, w_N) .

For the first N agents, we accordingly use $z^{(N+1)}$ to modify the cost integrand starting at τ_e to get

$$J_i \triangleq E \int_0^\infty e^{-\rho t} \left\{ [z_i - \Phi(z^{(N)}) 1_{(t < \tau_e)} - \Phi(z^{(N+1)}) 1_{(t \geq \tau_e)}]^2 + ru_i^2 \right\} (t) dt, \quad (24)$$

where $r > 0$, $\Phi(y) = \gamma(y + \eta)$. The cost for the new agent is

$$J_{N+1} \triangleq E \int_{\tau_e}^\infty e^{-\rho(t-\tau_e)} \left\{ [z_{N+1} - \Phi(z^{(N+1)})]^2 + ru_{N+1}^2 \right\} (t) dt. \quad (25)$$

For $1 \leq k \leq N+1$, we may alternatively write $J_k(u_1, \dots, u_{N+1})$ to indicate the associated controls.

B. The LQG-NCE Equation System

We denote

$$\Pi_a = \left(\frac{b^2}{r}\right)^{-1} \left[a - \frac{\rho}{2} + \sqrt{\left(a - \frac{\rho}{2}\right)^2 + \frac{b^2}{r}} \right], \quad (26)$$

$$\beta_1(a) = -\frac{\rho}{2} + \sqrt{\left(a - \frac{\rho}{2}\right)^2 + \frac{b^2}{r}}, \quad (26)$$

$$\beta_2(a) = \frac{\rho}{2} + \sqrt{\left(a - \frac{\rho}{2}\right)^2 + \frac{b^2}{r}}. \quad (27)$$

Here Π_a is the solution to the algebraic Riccati equation: $\rho\Pi = 2a\Pi - \frac{b^2}{r}\Pi^2 + 1$.

To simplify the aggregation procedure we assume zero initial mean for all agents, i.e., $Ez_i(0) = 0$, $i \geq 1$. The NCE consistency requirement leads to the equation system:

$$\rho s_a = \frac{ds_a}{dt} + as_a - \frac{b^2}{r} \Pi_a s_a + \alpha \Pi_a \bar{z} - z^*, \quad (28)$$

$$\frac{d\bar{z}_a}{dt} = \left(a - \frac{b^2}{r} \Pi_a\right) \bar{z}_a - \frac{b^2}{r} s_a + \alpha \bar{z}, \quad (29)$$

$$\bar{z} = \bar{z}_a, \quad (30)$$

$$z^* = \Phi(\bar{z}). \quad (31)$$

See [12] for details on construction of this equation system in an LQG context. Due to the uniform population, the mass effect \bar{z} is given by any representative agent (i.e., by \bar{z}_a).

In (28) it is unnecessary to specify the initial condition $s_a(0)$ for the bounded solution s_a derived from optimal tracking [12].

(H1) $\beta_1(a) > 0$ and $\frac{|\alpha|}{\beta_1(a)} + \frac{b^2(\gamma + |\alpha|\Pi_a)}{r\beta_1(a)\beta_2(a)} < 1$, where $\beta_1(a), \beta_2(a)$ are defined by (26)-(27). \square

(H2) All agents have mutually independent initial conditions of zero mean, i.e. $Ez_i(0) = 0$, $i \geq 1$. In addition, $\sup_{i \geq 1} Ez_i^2(0) < \infty$. \square

Theorem 6: [12] Under (H1)-(H2), the equation system (28)-(31) admits a unique bounded solution. \square

In fact, $\bar{z}(t)$ and s_a may be given in an explicit form (see [12]). Let u_i^0 denote the optimal tracking based control law,

$$u_i^0 = -\frac{b}{r} (\Pi_a z_i + s_a), \quad (32)$$

where s_a is derived from (28)-(31).

For the control calculation of the new comer, the equation system (28)-(31) may be used for computing the mass effect $\bar{z}(t)$ off-line. Once the new agent joins the game, it needs to re-construct a truncated version (started from that moment) of $\bar{z}(t)$ based upon its online measurement of the population mean $z^{(N+1)}$ which is now supplied as the additional information in order to deal with the unknown entry time. Next, if $z^{(N+1)}(t)$ is approximated by $\bar{z}(t)$, $t \geq \tau_e$, it can construct its own control as a standard tracking problem by determining the function s_a appearing in (32).

C. Control Strategy for the New Agent

At the system level, (25) gives a well defined quantity once the control inputs are properly selected and the condition (x_1, \dots, x_{N+1}) at τ_e is given. In our problem setting, we first take the control laws u_i^0 given by (32), for $1 \leq i \leq N$, and the remaining issue is to find a control law u_{N+1} which may be implemented by the information of the $(N+1)$ th agent.

Heuristically, if the first N agents take the controls u_i^0 , $1 \leq i \leq N$, we may effectively approximate the term $\Phi(z^{(N+1)})$ by a deterministic function z^* and subsequently find a suboptimal strategy u_{N+1} based upon tracking z^* . Intuitively, the influence of the new agent is negligible, and we may use (28)-(31) to determine z^* for the set of uniform agents.

Denote $\lambda_1 = (\rho + \alpha - \sqrt{(\rho + \alpha)^2 + 4\Theta})/2 < 0$ where $\Theta = \beta_2(\beta_1 - \alpha) + \frac{b^2}{r}(\alpha\Pi_a - \gamma) > 0$. Since the initial value $\bar{z}(0) = 0$ by (H2), it is easy to verify the relation

$$\bar{z}(t) = \bar{z}(\tau_e) e^{\lambda_1(t-\tau_e)} + (1 - e^{\lambda_1(t-\tau_e)}) \bar{z}(\infty)$$

for all $t \geq \tau_e$. However, the new agent has no means to know the exact value of $\bar{z}(\tau_e)$ as the entry time is unknown. To circumvent this difficulty, we approximate the value of $\bar{z}(\tau_e)$ by the population mean sampled at τ_e as follows:

$$\bar{z}(\tau_e) \approx z^{(N+1)}(\tau_e) = \frac{1}{N+1} \sum_{i=1}^{N+1} z_i(\tau_e), \quad (33)$$

which is the measurement at the entry time. This quantity is available to the new agent without knowing the entry time

τ_e (relative to the system time origin) since it is marked by the *event* that the new player commences its play.

Thus, we take the approximation

$$\bar{z}(t) \approx z^{(N+1)}(\tau_e)e^{\lambda_1(t-\tau_e)} + (1 - e^{\lambda_1(t-\tau_e)})\bar{z}(\infty). \quad (34)$$

Now by use of (33) and a time shifted version of (28), we construct the corresponding approximation for s_a with a time shift, as a function on $[\tau_e, \infty)$ in the form:

$$\begin{aligned} \hat{s}(t) = & [z^{(N+1)}(\tau_e) - \bar{z}(\infty)](\alpha\Pi_a - \gamma) \frac{e^{\lambda_1(t-\tau_e)}}{\beta_2 + |\lambda_1|} \\ & + \frac{\bar{z}(\infty)(\alpha\Pi_a - \gamma)}{\beta_2} - \frac{\gamma\eta}{\beta_2}. \end{aligned}$$

The new player's control law is given as

$$u_{N+1}^*(t) = -\frac{b}{r}(\Pi_a z_{N+1} + \hat{s})(t) \quad (35)$$

for $t \geq \tau_e$. The control (35) may be implemented since it depends upon $z^{(N+1)}(\tau_e)$ and the post-entry time $t - \tau_e$ both assumed to be known by the new agent.

For the equilibrium analysis, we use the control sets \mathcal{U}_i , $1 \leq i \leq N$, consisting of Lipschitz continuous feedback controls using $(t, z_1(t), \dots, z_N(t))$ and $z_{N+1}(t)$ (when $t \geq \tau_e$), for the first N agents. Each control u_{N+1} in \mathcal{U}_{N+1} satisfies that it is continuous in $(t - \tau_e, z^{(N+1)}(\tau_e), z_1(t), \dots, z_{N+1}(t))$ and is Lipschitz in $(z_1(t), \dots, z_{N+1}(t))$. The definition for ε -Nash equilibria in [10], [13] may be adapted to the variable population model in an obvious manner. For simplicity in Theorem 7 below we assume zero initial condition for z_{N+1} at τ_e , although this may be relaxed.

Theorem 7: Under (H1)-(H2), the set of controls $(u_1^0, \dots, u_N^0, u_{N+1}^*)$ is an ε -Nash equilibrium where $\varepsilon \rightarrow 0$, as $N \rightarrow \infty$.

Proof: (Sketch) — Let $C > 0$ (or C_1) denote a generic constant independent of N . Given u_i^0 , $1 \leq i \leq N$, there exists a fixed constant C and u_{N+1} such that

$$J_{N+1}(u_1^0, \dots, u_N^0, u_{N+1}) \leq C. \quad (36)$$

For instance, we may take $u_{N+1} = \bar{u}_{N+1}(t) = -\frac{b}{r}[\Pi_a z_{N+1}(t) + s_a(t - \tau_e)]$ for $t \geq \tau_e$, and verify (36) by basic estimates for the stable closed-loop system. Below we restrict u_{N+1} to satisfy (36), and we can further get

$$E \int_{\tau_e}^{\infty} e^{-\rho(t-\tau_e)} (z_{N+1}^2 + u_{N+1}^2)(t) dt \leq C_1. \quad (37)$$

Next we can show that

$$E |\bar{z}(\tau_e) - z^{(N+1)}(\tau_e)|^2 = O\left(\frac{1}{N}\right), \quad (38)$$

$$E \int_{\tau_e}^{\infty} e^{-\rho(t-\tau_e)} |\bar{z}(t) - z^{(N+1)}(t)|^2 dt = O\left(\frac{1}{N}\right). \quad (39)$$

We get (38) by the closed-loop on $[0, \tau_e]$, and (39) by the closed-loop of the $N+1$ agents and the boundedness of s_a .

Now we introduce the auxiliary control model

$$dz'_{N+1} = \alpha z'_{N+1} dt + bu'_{N+1} dt + \alpha \bar{z} dt + \sigma dw_{N+1}, \quad t \geq \tau_e,$$

with the initial condition $z'_{N+1}(\tau_e) = 0$ and cost

$$J'_{N+1} \triangleq E \int_{\tau_e}^{\infty} e^{-\rho(t-\tau_e)} \{ [z'_{N+1} - \Phi(\bar{z})]^2 + ru_{N+1}^2 \}(t) dt.$$

J'_{N+1} is minimized to get $\inf J'_{N+1}$ by a standard optimal tracking control law. We can check that $J_{N+1}(u_1^0, \dots, u_N^0, u_{N+1}^*) = \inf J'_{N+1} + o(1)$ and, furthermore, by (37) we can show $J_{N+1}(u_1^0, \dots, u_N^0, u_{N+1}) \geq \inf J'_{N+1} + o(1)$ with any other control u_{N+1} satisfying (36). The ε -Nash equilibrium property of u_{N+1}^* then follows.

Similarly, we can show an ε -Nash equilibrium property for u_i^0 , $1 \leq i \leq N$. The estimates may be obtained by forming the closed-loop dynamics of N (resp., $N+1$) dimension before (resp., after) the entry time τ_e . \square

REFERENCES

- [1] T. Başar and G. J. Olsder. *Dynamic Noncooperative Game Theory*, 2nd ed. Academic Press, London, UK, 1995.
- [2] C. T. Bauch and D. J. D. Earn. Vaccination and the theory of games. *Proc. Nat. Acad. Sci., U.S.A.*, vol. 101, pp. 13391-13394, Sept. 2004.
- [3] A. Stoussan. *Stochastic Control of Partially Observable Systems*, Cambridge Univ. Press, 1992.
- [4] R. Boel and P. Varaiya. Optimal control of jump processes. *SIAM J. Control Optim.*, vol. 15, no. 1, pp. 92-119, 1977.
- [5] R. Breban, R. Vardavas, and S. Blower. Inductive reasoning games in influenza vaccination models: mean field analysis. arXiv:q-bio.PE/0608016.
- [6] M. H. A. Davis and P. Varaiya. Dynamic programming conditions for partially observable stochastic systems. *SIAM J. Control Optim.*, vol. 11, pp. 226-261, 1973.
- [7] D. A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, vol. 20, pp. 247-308, 1987.
- [8] G. M. Erickson. Differential game models of advertising competition. *Europ. J. Oper. Res.*, vol. 83, pp. 431-438, 1995.
- [9] E. J. Green. Continuum and finite-player noncooperative models of competition. *Econometrica*, vol. 52, no. 4, pp. 975-993, 1984.
- [10] M. Huang, P. E. Caines, and R. P. Malhamé. Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. *Proc. the 42nd IEEE Conf. Decision Contr.*, Maui, Hawaii, pp. 98-103, December 2003.
- [11] M. Huang, P. E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ε -Nash equilibria. *IEEE Transactions on Automatic Control*, 2007 (to appear).
- [12] M. Huang, R. P. Malhamé, and P. E. Caines. Nash equilibria for large-population linear stochastic systems of weakly coupled agents, in *Analysis, Contr. Optim. Complex Dyn. Syst.*, E. K. Boukas and R. J. Malhamé eds., pp. 215-252, Springer, New York, 2005.
- [13] M. Huang, R. P. Malhamé, and P. E. Caines. Nash certainty equivalence in large population stochastic dynamic games: connections with the physics of interacting particle systems. *Proc. 45th IEEE CDC Conf.*, San Diego, CA, pp. 4921-4926, Dec. 2006
- [14] M. Huang, R. P. Malhamé, and P. E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, vol. 6, no. 3, pp. 221-252, 2006.
- [15] R. P. Isaacs. *Differential Games*. John Wiley, 1965.
- [16] M. Ali Khan and Y. Sun. Non-cooperative games with many players. In *Handbook of Game Theory with Economic Applications*, vol. 3, R. J. Aumann and S. Hart eds., North-Holland, 2002.
- [17] V. E. Lambson. Self-enforcing collusion in large dynamic markets. *J. Econ. Theory*, vol. 34, pp. 282-291, 1984.
- [18] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'Été de Probabilités de Saint-Flour XX -1989*, Lect. Notes Math. 1464, pp. 165-252, Springer-Verlag, Berlin, 1991.