AN INVARIANCE PRINCIPLE IN LARGE POPULATION STOCHASTIC DYNAMIC GAMES**

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.

Key words Large population, martingale representation, Nash equilibrium, optimal control, stochastic dynamic games.

1 Introduction

Noncooperative dynamic game theory has attracted a long lasting research interest for decades. In such games, the states of the players (also to be 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 games where the dynamics are given by differential or difference equations, and the classical solution notion is saddle strategies[1]. 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.

In the setting of 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[2]. In continuous time stochastic models, this procedure derives a set of coupled Hamilton-Jacobi-Bellman (HJB) equations. In general, the complexity of this approach is high, especially in the case of many players. Firstly, it is difficult to show existence and uniqueness

Minyi HUANG
Department of Information Engineering, Research School of Information Sciences and Engineering, The Australian National University, Canberra, ACT 0200, Australia. Email: minyi.huang@rsise.anu.edu.au.
Peter E. CAINES
Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 2A7, Canada; also affiliated with GERAD. Email: peterc@cim.mcgill.ca.
Roland P. MALHAMÉ
Department of Electrical Engineering, École Polytechnique de Montréal, Montreal, QC H3C 3A7, Canada; also affiliated with GERAD. Email: roland.malhame@polymtl.ca.

of solutions to the coupled HJB equations, and secondly, the set of 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[3−7], 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. 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[5,8−11]. 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. By this approach, each decision-maker can ignore the fine details of the behavior of all individual players by only focusing on the overall impact of the population. This methodology avoids the generation of a large coupled HJB equation system with as many equations as the population size. In the end, this procedure leads to decentralized strategies for the individual players in the system of N players.

In this paper, we study the large population behavior by identifying an invariance property in terms of the instantaneous cost, the value function and the empirical distribution function of the states of all agents. In our asymptotic analysis, we let $N \to \infty$. This essentially amounts to considering a family of game problems with an increasing number of agents.

In the NCE methodology, each agent can be viewed as essentially solving 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[12,13]. Compared to the usual HJB equation characterization of the value function, this martingale representation reveals sample path properties for the optimally controlled process.

Now, by extending the martingale results in optimal control to the population limit of the dynamic game and averaging across the population to eliminate randomness, one obtains a deterministic martingale and hence a constant value over time. This gives the so-called invariance property. Subsequently, by interpreting the measure process, appearing in the closed-loop McKean-Vlasov equation, as the limiting population state empirical distribution, we obtain an invariance property associated with the controlled population behavior when the NCE strategies are implemented by the agents.

The organization of the remaining part of the paper is as follows. In Section 2, we formulate the large population dynamic game problem in the general nonlinear context. The NCE based decentralized control synthesis is described in Section 3. In Section 4, we characterize the martingale property in the population limit and establish the invariance property. The special case of LQG systems is then analyzed in Section 5 for illustration. Section 6 concludes the paper.

2 The Stochastic Dynamic Game Model

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

$$dz_i(t) = \left(\frac{1}{N}\right) \sum_{j=1}^{N} f_{a_i}(z_i(t), u_i(t), z_j(t))dt + \sigma dw_i(t), \qquad 1 \le i \le N, \quad t \ge 0, \qquad (1)$$

where $\{w_i, 1 \le i \le N\}$ 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

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

These stochastic dynamic game models are well motivated by many practical decision scenarios arising from social, economic and engineering models where each decision maker faces the aggregate effect of all other players.

We note that the analysis in the paper may be easily generalized to deal with variants of the set of costs in (2). For the system configuration $z = (z_1, z_2, \cdots, z_N)$, define the empirical distribution $\varepsilon_z = (\frac{1}{N}) \sum_{i=1}^N \delta_{z_i}$ where $\delta_{(\cdot)}$ is the Dirac measure. We introduce a set of costs in the following form

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

where $\varepsilon_z(t) = (\frac{1}{N}) \sum_{i=1}^N \delta_{z_i(t)}$. Here $\widetilde{g}(\cdot, \cdot, \cdot)$ 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 $(z_1, z_2, \cdots, z_n)$ except z_i, g remains the same value when (z_i, u_i) is given. It can be checked that (3) includes the cost form in (2) as a special case.

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

In the analysis, we use $u_i(\cdot)$ to denote the control input on $[0, T]$, and $u_i(t)$ to denote its value at time t. For each $t \in [0, T]$, $u_i(t)$ takes values from U, which is a closed subset of $\mathbb{R}$. Each $u_i(\cdot)$ is adapted to the σ-algebra $\sigma(z_i(0), w_i(s), \ s \leq t, 1 \leq i \leq N)$.

We assume that the entries in the sequence $\{a_i, i \geq 1\}$ take values from a finite set $\mathcal{A} \triangleq \{\theta_1, \theta_2, \cdots, \theta_K\}$. For the set of dynamic parameters $\{a_i, 1 \leq i \leq N\}$, we denote the empirical distribution function

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

In this setup, we assume the existence of a limit empirical distribution function F on $\mathcal{A}$ for the sequence $\{F_N, N \geq 1\}$ and denote this distribution as $\pi = (\pi_1, \pi_2, \cdots, \pi_K)$, i.e., $\lim_{N \to \infty} F_N(\{\theta_k\}) = F(\{\theta_k\}) = \pi_k$ for each $\theta_k \in \mathcal{A}$.

2.1 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[14] in the form of N coupled stochastic differential equations (SDE):

$$dx_i(t) = \left(\frac{1}{N}\right) \sum_{k=1}^N b(x_i(t), x_k(t)) dt + \sigma dw_i(t), \quad 1 \leq i \leq N, \quad t \geq 0, \tag{4}$$

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 are assumed to 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 $x_i(0)$, $1 \leq i \leq N$. Let $\varepsilon_x = (\frac{1}{N}) \sum_{i=1}^{N} \delta_{x_i}$ denote the empirical measure of the particle configuration $x = (x_1, x_2, \cdots, x_N)$. Then the drift term in (4) may be expressed as a function of x_i and ε_x.

For this class of particle models, one can achieve a remarkable economy of representation by expressing 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. And furthermore, there is a decoupling effect such that a single particle's statistical properties can effectively approximate the empirical distribution produced by all particles; in other words, the large population may be viewed as independent copies of a given agent with certain statistical properties. More specifically, as N tends to infinity, the individual dynamics may be written in the limiting form:

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

which is the celebrated McKean-Vlasov (M-V) equation[15]. Here $b[x, \mu_t] = \int b(x,y)\mu_t(dy)$ for some probability distribution μ_t on $\mathbb{R}$. The noise $w(t)$ may be determined in different ways. For instance, if one intends to approximate $x_1(t)$ in (4) by $x(t)$, one may set $w(t) = w_1(t)$ as the driving Brownian motion in (5) and $x(0) = x_1(0)$. Note that by introducing the density function $p_t(x)$, associated with μ_t for $x(t)$, one may recast (5) in the form of a Fokker-Planck equation whose coefficients depend upon the density $p_t(x)$ itself.

Definition 1   A pair $(x(t), \mu_t)$, $t \geq 0$, is said to be 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 $t \geq 0$.

It is obvious that μ_0 in Definition 1 is determined as the distribution of $x(0)$.

2.2 Related Notation for the Large Population Game

Before further analysis for the game problem in a controlled McKean-Vlasov setting, we introduce some notation. For a set of K probability distributions $\mu_t^o \triangleq (\mu_t^1, \mu_t^2, \cdots, \mu_t^K)$, where each μ_t^k is defined for Borel subsets of $\mathbb{R}$ and $0 \leq t \leq T$, we define

$$f_a[x, u, \mu_t^1, \cdots, \mu_t^K] = \sum_{i=1}^{K} \pi_i \int_{\mathbb{R}} f_a(x, u, y)\mu_t^i(dy),$$

and similarly,

$$g[x, u, \mu_t^1, \cdots, \mu_t^K] = \sum_{i=1}^{K} \pi_i \int_{\mathbb{R}} g(x, u, y)\mu_t^i(dy).$$

Here $f_a[x, u, \mu_t^1, \cdots, \mu_t^K]$ and $g[x, u, \mu_t^1, \cdots, \mu_t^K]$ may be looked at as two functions from $\mathbb{R} \times U \times [0, T]$ to $\mathbb{R}$. We denote $f_a[x, u, \mu_t^o] = f_a[x, u, \mu_t^1, \cdots, \mu_t^K]$ and $g[x, u, \mu_t^o] = g[x, u, \mu_t^1, \cdots, \mu_t^K]$.

3 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 a set of K probability measures to approximate the effect of K classes of agents. Next, the dynamics and cost of agent i can be approximated in terms of 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 [10], we introduce the NCE based equation system

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

$$-\frac{\partial v_{a_i}(t, x_i)}{\partial t} = \min_{u_i \in U} \left\{ f_{a_i}[x_i, u_i, \mu_t^o]\frac{\partial v_{a_i}(t, x_i)}{\partial x_i} + g[x_i, u_i, \mu_t^o] \right\} + \frac{\sigma^2}{2}\frac{\partial^2 v_{a_i}(t, x_i)}{\partial x_i^2}, \tag{7}$$

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

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

where $0 \leq t \leq T$ and we denote the distribution of $x_i(t)$ in (6) by μ_t^k when $a_i = \theta_k$. In the above, $v_{a_i}(t, x_i)$ 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) in the form of a Fokker-Planck equation.

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

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

To distinguish from the original game model, here we use a different variable x_i 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 equation (6) is interpreted in the generalized McKean-Vlasov equation sense for multi-class particles, and the control law $\widehat{u}_i$ is a minimizer of the right hand side of (7). A solution to the above McKean-Vlasov-HJB system consists of a triple $(x_i(t), v_{a_i}(t, x_i), \widehat{u}_i(t, x_i))$, where $t \geq 0$, representing the closed-loop solution, the value function and the feedback control law.

A detailed analysis is developed in [10] about the existence of a solution to the above equation system. This amounts to a fixed point argument and identifying some gain conditions for 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.

4 The Martingale Property of the Population

4.1 The Optimal Control Problem

Before considering the game problem introduced above, 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

$$dz(t) = f(z(t), u(t))dt + \sigma dw(t), \tag{10}$$

where the initial condition $z(0)$ is independent of the standard Wiener process $w(t)$ and satisfies $E|z(0)|^2 < \infty$. The cost function is given as

$$J_0 = E \int_0^T g(z(t), u(t))dt.$$

Here for simplicity, we restrict the analysis to a scalar state and control input, respectively, $x(t)$ and $u(t)$. The control takes 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).

Denote the value function

$$v(t, y) = \inf_{u(\cdot)} E\left[\int_t^T g(z(s), u(s))ds | z(t) = y \right],$$

where $t \in [0, T]$. 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

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

The following lemma is easily proved by a method similar to that in [12]. After assuming the existence of the optimal control law, the lemma essentially follows from the optimality principle.

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.

Note that, based upon our assumption on the function f and the optimal control law u^*, the closed-loop system associated with (10) has a unique strong solution.

4.2 The Martingale Property and Its Limiting Form

Subsequently, we will 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 $(x_i(t), v_{a_i}(t, x_i), \widehat{u}_i(t, x_i))$ to the McKean-Vlasov-HJB system (6)–(8).

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

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

Theorem 3 *Suppose* (A1)–(A3) *hold. Then the process*

$$\int_0^t g[x_i(s), \widehat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t, x_i(t))$$

is a martingale.

Proof For $t \in [0, T]$, let

$$\zeta_t = \int_0^t g[x_i(s), \widehat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t, x_i(t)),$$

and let $\mathcal{F}_t$ denote the σ-algebra generated by $(x_i(s), s \leq t)$. We have

$$E[\zeta_{t_2}|\mathcal{F}_{t_1}] = E\bigg[\int_0^{t_2} g[x_i(s), \widehat{u}_i(s, x_i(s)), \mu_s^o]ds + v_{a_i}(t_2, x_i(t_2))|\mathcal{F}_{t_1}\bigg]$$

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

On the other hand, with respect to the measure process μ_s^o, $\widehat{u}_i$ may be interpreted as a usual optimal control. Hence, by the optimality principle, we have

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

Now it follows that $E[\zeta_{t_2}|\mathcal{F}_{t_1}] = \zeta_{t_1}$ and this completes the proof. ∎

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 the process*

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

converges in L_2 to a constant value c for all $t \in [0, T]$, as $N \to \infty$.

Proof Theorem 3 implies

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

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. ∎

Define

$$\bar{g}_t = \int_{\mathbb{R}^2} \sum_{i,j} \pi_i \pi_j g(x, \varphi_{\theta_i}(s, x), y) \mu_s^j(dy) \mu_s^i(dx), \quad \bar{v}_t = \int_{\mathbb{R}} \sum_{i=1}^N \pi_i v_{\theta_i}(t, x) \mu_t^i(dx),$$

where $\bar{v}_t$ and $\bar{g}_t$ are both deterministic functions on $[0, \infty)$. Then it follows that

$$\frac{d\bar{v}_t}{dt} = -\bar{g}_t.$$

In the above, $\bar{g}_t$ may be interpreted 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. Notice that by the closed-loop McKean-Vlasov equation, we may write a special Fokker-Planck equation provided that there is a well defined density. This is purely based upon the dynamics of the given agent. In contrast, (11) shows how the forward propagation of the distribution should preserve the equality relation in terms of the instantaneous cost and the future cost to go based upon the population averaging.

5 The LQG System

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

$$dz_i(t) = a_i z_i(t)dt + bu_i(t)dt + \left(\frac{\alpha}{N}\right) \sum_{j=1}^{N} z_j(t)dt + \sigma dw_i(t), \tag{12}$$

and the cost for the ith agent is given as

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

For simplicity, we assume the independent initial states $z_i(0)$ have zero mean. In the LQG model, 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 the function $F(a)$.

Corresponding to the LQG game model (12)–(13), we first construct the auxiliary control problem:

$$\begin{cases} dz_i(t) = a_i z_i(t)dt + bu_i(t)dt + \alpha \bar{z}(t)dt + \sigma dw_i(t), \\ J_i = E \int_0^T \left\{ [z_i(t) - z^*(t)]^2 + ru_i^2(t) \right\} dt, \end{cases} \tag{14}$$

where $\bar{z}, z^* \in C[0, T]$.

The above LQG tracking problem may be solved using a similar method as in [16]. Write the Riccati differential equation

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

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

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

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

Finally, we introduce the equation

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

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

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

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

By use of the solution to the above auxiliary LQG tracking problem, the NCE scheme may be expressed in a more explicit form in terms of the following equation system

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

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

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

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

Compared to the original NCE equation system (6)–(8), here we only need the dynamics of the mean process of the closed-loop for the individual agent in question, instead of the controlled diffusion process or its Fokker-Planck equation. For convenience of presentation, we shall also refer (18)–(21) as the NCE equation system without causing a conceptual problem.

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

$$\Phi_a(t, s) = \exp \left\{ - \int_s^t \left(a - \frac{b^2}{r} \Pi_a(\tau) \right) d\tau \right\}.$$

Theorem 5 *The sufficient condition for the existence of a unique solution to the equation system* (18)–(21) *is*

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

where $F(a)$ is the empirical distribution of the dynamic parameter $a \in \mathcal{A}^0$.

The proof is given in the Appendix.

Now we end this section by computation with the concrete parameter set $\mathcal{A}^0 = \{\theta_1, \cdots, \theta_K\}$ where $F(\{\theta_k\}) = \pi_k$, $1 \le k \le K$. 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

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

$$\left. + \int_{\mathbb{R}} [x^2 \Pi_{\theta_k}(t) + 2x s_{\theta_k}(t)] dF_{\theta_k}^t(x) + q_{\theta_k}(t) \right\}, \tag{22}$$

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 function $p_{\theta_k}^t(x)$ with suitable regularity; then by taking differentiation with respect to t on both sides of (22), we get

$$0 = \sum_{k=1}^K \pi_k \left\{ \int_{\mathbb{R}} \left[(x - z^*(t))^2 + \frac{b^2}{r} (\Pi_{\theta_k}(t) x + s_{\theta_k}(\tau))^2 \right] p_{\theta_k}^t(x) dx \right.$$

$$\left. + \int_{\mathbb{R}} \frac{\partial [x^2 \Pi_{\theta_k}(t) + 2x s_{\theta_k}(t)] p_{\theta_k}^t(x)}{\partial t} dx + \frac{dq_{\theta_k}(t)}{dt} \right\}. \tag{23}$$

Due to the quadratic cost structure, the fundamental 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 when the control strategies are employed in the population limit.

6 Conclusion

In this paper, we consider large population stochastic dynamic games. By extending the classical martingale representation results in optimal control to the large population conditions, we have shown that one obtains an invariance property for the population behavior when the mass is individually subject to NCE feedback control laws.

Appendix: Proof of Theorem 5

We prove the existence and uniqueness of a solution by a fixed point argument. By (18) and (21), we write the equation

$$\frac{ds_a(t)}{dt} = -\left(a - \frac{b^2}{r}\Pi_a(t)\right)s_a(t) - \alpha\Pi_a(t)\bar{z}(t) + \gamma(\bar{z}(t) + \eta) = 0.$$

By the terminal condition $s_a(T) = 0$, the solution may be expressed as

$$s_a(t) = \int_t^T \Phi_a(t,\tau)[\alpha\Pi_a(\tau)\bar{z}(\tau) - \gamma(\bar{z}(\tau) + \eta)]d\tau \triangleq \Gamma_1(\bar{z})(t).$$

Next we have

$$\bar{z}_a(t) = \int_0^t \Phi_a(\kappa,t)\left\{-\frac{b^2}{r}\int_0^T \Phi_a(\kappa,\tau)[\alpha\Pi_a(\tau)\bar{z}(\tau) - \gamma(\bar{z}(\tau) + \eta)]d\tau + \alpha\bar{z}(\kappa)\right\}d\kappa.$$

We further define the operator

$$\Gamma(\bar{z})(t) \triangleq \int_{\mathcal{A}^0}\int_0^t \Phi_a(\kappa,t)\left\{-\frac{b^2}{r}\int_0^T \Phi_a(\kappa,\tau)[\alpha\Pi_a(\tau)\bar{z}(\tau) - \gamma(\bar{z}(\tau) + \eta)]d\tau + \alpha\bar{z}(\kappa)\right\}d\kappa dF(a).$$

It is easy to show that the above gives a well-defined operator Γ from $C[0,T]$ to $C[0,T]$.

Now, by the inequality condition, we see there exists a unique solution to the fixed point equation

$$\bar{z}(t) = \Gamma(\bar{z})(t). \tag{24}$$

Subsequently, we may determine

$$s_a(t) = \Gamma_1(\bar{z})(t)$$

and $\bar{z}_a(t)$ is determined in an obvious manner. It is easy to verify that $(s_a(t), \bar{z}_a(t), \bar{z}(t))$ thus determined satisfies the NCE equation system. On the other hand, in any solution $(s'_a(t), \bar{z}'_a(t), \bar{z}'(t))$ to the equation system, the component $\bar{z}'(t)$ necessarily satisfies the fixed point equation (24), and uniqueness of the solution follows easily. ∎

References

[1] R. P. Isaacs, *Differential Games*, John Wiley, 1965.

[2] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory* (2nd Ed.), Academic Press, London, UK, 1995.

[3] G. M. Erickson, Differential game models of advertsing competition, *Europ. J. Oper. Res.*, 1995, **83**: 431–438.

[4] E. J. Green, Continuum and finite-player noncooperative models of competition, *Econometrica*, 1984, **52**(4): 975–993.

[5] 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, in *Proc. the* 42*nd IEEE Conf. Decision Contr.*, Maui, Hawaii, December 2003, 98–103.

[6] V. E. Lambson, Self-enforcing collusion in large dynamic markets, *J. Econ. Theory*, 1984, **34**: 282–291.

[7] M. Ali Khan and Y. Sun, Non-cooperative games with many players, in *Handbook of Game Theory with Economic Applications* (ed. by R. J. Aumann and S. Hart), North-Holland, 2002, **3**.

[8] M. Huang, R. P. Malhamé, and P. E. Caines, Nash equilibria for large-population linear stochastic systems of weakly coupled agents, in *Analysis, Control and Optimization of Complex Dynamic Systems* (ed. by E. K. Boukas and R. P. Malhamé), Springer, New York, 2005, 215–252.

[9] 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, in *Proc. the* 45*th IEEE Conference on Decision and Control*, San Diego, CA, December 2006, 4921–4926.

[10] 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 Control*, 2007, to appear.

[11] M. Huang, P. E. Caines, and R. P. Malhamé, Large-population cost-coupled LQG problems with non-uniform agents: Individual-mass behavior and decentralized ε-Nash equilibria, *IEEE Transactions on Automatic Control*, 2007, **52**, to appear.

[12] R. Boel and P. Varaiya, Optimal control of jump processes, *SIAM J. Control Optim.*, 1977, **15**(1): 92–119.

[13] M. H. A. Davis and P. Varaiya, Dynamic programming conditions for partially observable stochastic systems, *SIAM J. Control Optim.*, 1973, **11**: 226–261.

[14] D. A. Dawson and J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions, *Stochastics*, 1987, **20**: 247–308.

[15] A. S. Sznitman, Topics in propagation of chaos, in *Ecole d'Eté de Probabilitiés de Saint-Flour XIX* -1989, Lect. Notes Math. 1464, Springer-Verlag, Berlin, 1991, 165–252.

[16] A. Bensoussan, *Stochastic Control of Partially Observable Systems*, Cambridge Univ. Press, 1992.