43rd IEEE Conference on Decision and Control
December 14-17, 2004
Atlantis, Paradise Island, Bahamas

Large-population cost-coupled LQG problems: generalizations to non-uniform individuals

Minyi Huang[†], Peter E. Caines[‡], and Roland P. Malhamé[§]

Abstract— **We consider LQG games in large population conditions where the agents have non-uniform dynamics and are coupled by their individual costs. A state aggregation technique is developed to obtain a set of decentralized control laws which possesses an ε-Nash equilibrium property. An attraction property of the mass is also established. The methodology and the results contained in this paper illuminate individual and mass behaviour in such large complex systems.**

I. Introduction

In this paper, we investigate the optimization of large-scale linear quadratic Gaussian (LQG) control systems wherein many agents (also to be called players) have similar dynamics and will evolve independently when state regulation is not included. To facilitate our exposition the individual cost based optimization shall be called the *dynamic LQG game*, or simply *LQG game*. In this framework, each agent is coupled with the other agents only through its cost function. We view this to be the characteristic property of a class of situations which we term (*distributed*) *cost-coupled* control problems. The study of such large-scale cost-coupled systems is motivated by a variety of scenarios, for instance, dynamic economic models involving agents linked via a market, and power control in mobile wireless communications. In the latter case, different users have independent power control mechanisms and statistically independent communication channels, but they interact with each other via the signal-to-interference ratio (SIR) based performance measure [7], [9], [6].

For the LQG game we analyze the ε-Nash equilibrium properties for a decentralized control law by which each individual optimizes its cost function depending upon the state of the individual agent and the average state of all other agents taken together, hereon referred to as “the mass”. In preceding work [7] we considered the LQG game for a population of uniform agents and introduced a state aggregation procedure for the design of decentralized control. In the non-uniform case studied here a given agent only has exact information on its own dynamics. The information concerning other agents is available in a statistical sense as described by a randomized parametrization for agents' dynamics across the population. Due to the particular structure of the individual cost, the mass formed by all other agents impacts any given agent as a nearly deterministic quantity. In response to any known mass influence, a given individual will select its localized control strategy to minimize its own cost. In a practical situation the mass influence cannot be assumed known *a priori*. It turns out, however, that this does not present any difficulty for applying the individual-mass interplay methodology as described below.

In the noncooperative game setup studied here, an overall rationality assumption for the population, to be characterized further down, implies the potential of achieving a stable predictable mass behaviour in the following sense: if some deterministic mass behaviour were to be given, rationality would require that each agent synthesize its individual cost based optimal response as a *tracking* action. Thus the mass trajectory corresponding to rational behaviour would guide the agents to collectively generate the trajectory which, individually, they were assumed to be reacting to in the first place. Indeed, if a mass trajectory with the above fixed point property existed, if it were unique, and, furthermore, if each individual had enough information to compute it, then rational agents who were assuming all other agents to be rational would anticipate their collective state of agreement and select a control policy consistent with that state. Thus, in the context of this paper, we make the following rationality assumption: Each agent is rational in the sense that it both (i) optimizes its own cost function, and (ii) assumes that all other agents are being simultaneously rational when evaluating their competitive behaviour. This justifies and motivates the search for mass trajectories with the fixed point property.

The central results of this paper consist of the precise characterization of (1) the Nash equilibrium associated with the individual cost functions depending on both the individual and mass behaviour, (2) the consistency (fixed point property) of the mass trajectory under the Nash equilibrium inducing individual feedback controls, and (3) the global attraction property of the mass behaviour in the function space with respect to policy iterations associated with such individual optimizing behaviour. This equilibrium then has the rationality and optimality interpretations but we underline that these hypotheses are not employed in the mathematical derivation of the results.

The framework presented in this paper is particularly

Work partially supported by ARC and NSERC.

[†]Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia. Email: m.huang@ee.mu.oz.au

[‡]Department of Electrical and Computer Engineering, McGill University, Montreal H3A 2A7, Canada. Email: peterc@cim.mcgill.ca

[§]Department of Electrical Engineering, École Polytechnique de Montréal, Montreal H3C 3A7, Canada. Email: roland.malhame@polymtl.ca

0-7803-8682-5/04/$20.00 ©2004 IEEE

suitable for optimization of large-scale systems where individuals seek to optimize for their own reward and where it is effectively impossible to achieve global optimality through close coordination between all agents. In this context, the methodology of noncooperative games and state aggregation developed in this paper provides a feasible approach for building simple (decentralized) optimization rules which under appropriate conditions lead to stable population behaviour. Our methodology could potentially provide effective methods for analyzing complex systems arising in socio-economic and engineering areas [9].

It is worthwhile to note that the large population limit formulation presented in this paper is relevant to research in the economic community on (mainly static) models with a large number or a continuum of agents; see e.g. [5]. However, instead of directly assigning a priori measure in a continuum space for labelling an infinite number of agents [5], we induce a probability distribution on a parameter space via empirical statistics, and furthermore, based on the induced measure we develop *state aggregation* for the underlying *dynamic* models.

In the paper we omit almost all of the proofs which may be found in [8].

II. Dynamically Independent and Cost-Coupled Systems

Consider an n dimensional linear stochastic system where the evolution of each state component is described by

$$dz_i = (a_i z_i + b_i u_i)dt + \sigma_i dw_i,\ 1 \le i \le n,\ t \ge 0, \quad (1)$$

where $\{w_i, 1 \le i \le n\}$ denotes n independent standard scalar Wiener processes. The initial state $z_i(0)$ are mutually independent and are also independent of $\{w_i, 1 \le i \le n\}$. In addition, $E|z_i(0)|^2 < \infty$ and $b_i \ne 0$. Each state component shall be referred to as the state of the corresponding individual (also to be called an agent or a player).

Evidently, the state of a given individual is not subject to direct influence from any other individual apart from possible feedback effects. In this paper we investigate the behaviour of the agents when they only interact with each other through coupling terms in their costs; this is displayed in the following set of *individual* cost functions which shall be used henceforth in the analysis:

$$J_i(u_i, v_i) \triangleq E \int_0^\infty e^{-\rho t}[(z_i - v_i)^2 + r u_i^2]dt. \quad (2)$$

We term this type of model a *dynamically independent and cost-coupled system*. For simplicity of analysis we assume in this paper that $b_i = b > 0,\ 1 \le i \le n$. In particular we assume the cost-coupling to be of the following form for most of our analysis: $v_i = \gamma(\frac{1}{n}\sum_{k\ne i}^n z_k + \eta)$, and we study the large scale system behaviour in the dynamic noncooperative game framework where each individual is linked with others via v_i. Evidently the linking term v_i gives a measure of the average effect of the mass formed by all other agents. Here we assume $\rho, r, \gamma, \eta > 0$, and throughout the paper z_i is described by the dynamics (1).

A. A Production Output Planning Example

The production output adjustment problem is based upon the early work of Basar and Ho [3] where a quadratic nonzero-sum game was considered for a static duopoly model and where it was assumed that the price of the commodity decreases linearly as the overall production level of the two firms increases. Here we study a dynamic model consisting of many players.

Consider n firms $F_i,\ 1 \le i \le n$, supplying the same product to the market. First, let x_i be the production level of firm F_i and suppose x_i is subject to adjustment by the following model:

$$dx_i = u_i dt + \sigma_i dw_i, \qquad t \ge 0, \quad (3)$$

which is a special form of (1). Here u_i denotes the action of increasing or decreasing the production level x_i, and $\sigma_i dw_i$ denotes uncertainty in the change of x_i.

Second, by generalizing the affine linear price model of [3] to the case of many players, we assume the price of the product is given by

$$p = \overline{\eta} - \overline{\gamma}(\frac{1}{n}\sum_{i=1}^n x_i), \quad (4)$$

where $\overline{\eta}, \overline{\gamma} > 0$. In (4) the overall production level $Q \triangleq \sum_{i=1}^n x_i$ is scaled by $\frac{1}{n}$. A justification for doing so is that this may be used to model the situation when an increasing number of firms distributed over different areas join together to serve an increasing number of consumers. $\frac{1}{n}\sum_{i=1}^n x_i$ measures the average production level in an expanding market. Thus this is a useful paradigm in an increasingly globalized market. Following [3], $\overline{\gamma}$ may be interpreted as the "slope of the demand curve". In fact, (4) may be regarded as a simplified form of a more general price model introduced by Lambson in a large market [10].

We assume that firm F_i adjusts its production level x_i by referring to the current price. Indeed, an increasing price calls for more supplies of the product to consumers and a decreasing price for less. We seek a production level which is approximately in proportion to the price that the current market provides, i.e., $x_i \approx \beta p = \beta[\overline{\eta} - \overline{\gamma}(\frac{1}{n}\sum_{i=1}^n x_i)]$, where $\beta > 0$ is a constant; based upon this requirement we introduce a penalty term $\{x_i - \beta[\overline{\eta} - \overline{\gamma}(\frac{1}{n}\sum_{i=1}^n x_i)]\}^2 \triangleq (x_i - v_i^0)^2$. On the other hand, in the adjustment of x_i, the control u_i corresponds to actions of shutting down or restarting production lines, or even the construction of new ones; these may further lead to hiring or laying off workers [4]. Each of these actions will incur certain costs to the firm; for simplicity we denote the instantaneous cost of the adjustment by $r u_i^2$, where $r > 0$. We now write the infinite horizon discounted cost for firm F_i as follows:

$$J_i^x(u_i, v_i^0) = E \int_0^\infty e^{-\rho t}[(x_i - v_i^0)^2 + r u_i^2]dt, \quad (5)$$

where $\rho > 0$ and the superscript in J_i^x indicates the associated dynamics (3). Due to the penalty on the change rate u_i, this situation may be regarded as falling into the framework of smooth planning [4]. Here obviously $v_1^0 = \cdots = v_n^0$. Notice that $v_i^0 = \beta[\overline{\eta} - \overline{\gamma}(\frac{1}{n}\sum_{i=1}^n x_i)]$ in this example and $v_i = \gamma(\frac{1}{n}\sum_{k\neq i}^n z_k + \eta)$ in the generic case (2) share the common feature of averaging over a mass.

In this production planning example, each firm has its independent individual dynamics and all the firms interact with each other through the market in which they seek to optimize their individual costs. This gives rise to what is termed here a cost-coupled situation (see e.g. [10], [2]).

III. The Preliminary Linear Tracking Problem

To begin with, for large n, assume $z_{-i}^* \stackrel{\triangle}{=} \gamma(\frac{1}{n}\sum_{k\neq i}^n z_k + \eta)$ in Section II is approximated by a *deterministic* continuous function z^* defined on $[0,\infty)$. For a given z^*, we construct the *individual cost* for the i-th player as follows:

$$J_i(u_i, z^*) = E\int_0^\infty e^{-\rho t}\{[z_i - z^*]^2 + ru_i^2\}ds. \qquad (6)$$

Here we shall consider the general tracking problem with bounded z^*. For minimization of J_i, the admissible control set is taken as $\mathcal{U}_i \stackrel{\triangle}{=} \{u_i | u_i$ adapted to $\sigma(z_i(0), w_i(s), s \leq t)$, and $E\int_0^\infty e^{-\rho t}(z_i^2 + u_i^2)dt < \infty\}$. Define

$$C_b[0,\infty) \stackrel{\triangle}{=} \{x \in C[0,\infty),\ |x|_\infty < \infty\},$$

where $|x|_\infty = \sup_{t\geq 0}|x(t)|$, for $x \in C[0,\infty)$. Under the norm $|\cdot|_\infty$, $C_b[0,\infty)$ is a Banach space [11]. Let Π_i be the positive solution to the algebraic Riccati equation

$$\rho\Pi_i = 2a_i\Pi_i - \frac{b^2}{r}\Pi_i^2 + 1. \qquad (7)$$

It is easy to verify that $-a_i + \frac{b^2\Pi_i}{r} + \frac{\rho}{2} > 0$. Denote

$$\beta_1 = -a_i + \frac{b^2}{r}\Pi_i, \quad \beta_2 = -a_i + \frac{b^2}{r}\Pi_i + \rho. \qquad (8)$$

Clearly, $\beta_2 > \frac{\rho}{2}$.

Proposition 3.1: Assume (i) $E|z_i(0)|^2 < \infty$ and $z^* \in C_b[0,\infty)$; (ii) $\Pi_i > 0$ is the solution to (7) and $\beta_1 = -a_i + \frac{b^2}{r}\Pi_i > 0$; and (iii) $s_i \in C_b[0,\infty)$ is determined by

$$\rho s_i = \frac{ds_i}{dt} + a_i s_i - \frac{b^2}{r}\Pi_i s_i - z^*. \qquad (9)$$

Then the control law $\widehat{u}_i = -\frac{b}{r}(\Pi_i z_i + s_i)$ minimizes $J_i(u_i, z^*)$, for all $u_i \in \mathcal{U}_i$. □

Proposition 3.2: Suppose assumptions (i)-(iii) in Proposition 3.1 hold and $q \in C_b[0,\infty)$ satisfies

$$\rho q = \frac{dq}{dt} - \frac{b^2}{r}s_i^2 + (z^*)^2 + \sigma_i^2\Pi_i. \qquad (10)$$

Then the cost for the control law $\widehat{u}_i$ is given by $J_i(\widehat{u}_i, z^*) = \Pi_i E z_i^2(0) + 2s(0)Ez_i(0) + q(0)$. □

Remark: In Proposition 3.1, assumption (ii) means that the resulting closed-loop system has a stable pole. □

Remark: s_i in Proposition 3.1 may be uniquely determined by only utilizing its boundedness, and it is unnecessary to specify the initial condition for (9) separately. Similarly, after $s_i \in C_b[0,\infty)$ is obtained, q in Proposition 3.2 can be uniquely determined from its boundedness. □

Proposition 3.3: Under the assumptions of Proposition 3.1, there exists a unique initial condition $s_i(0) \in \mathbb{R}$ such that the associated solution s_i to (9) is in $C_b[0,\infty)$. And moreover, for the obtained $s_i \in C_b[0,\infty)$, there is a unique $q(0) \in \mathbb{R}$ for (10) such that the solution $q \in C_b[0,\infty)$.

Proof: Consider (9) with initial condition $s_i(0)$ which leads to $s_i(t) = s_i(0)e^{\beta_2 t} + e^{\beta_2 t}\int_0^t e^{-\beta_2\tau}z^*(\tau)d\tau$. Since $\beta_2 > 0$ always holds, the integral $\int_0^\infty e^{-\beta_2\tau}z^*(\tau)d\tau$ exists and is finite. We take $s_i(0) = -\int_0^\infty e^{-\beta_2\tau}z^*(\tau)d\tau$ which yields $s_i(t) = -e^{\beta_2 t}\int_t^\infty e^{-\beta_2\tau}z^*(\tau)d\tau \in C_b[0,\infty)$, and it is easily verified that any initial condition other than $s_i(0)$ yields an unbounded solution. Similarly, a unique $q(0)$ in (10) may be determined to give $q \in C_b[0,\infty)$. □

IV. Competitive Behaviour and Continuum Mass Behaviour

In the cost coupled situation with individual costs each agent is assumed to be rational in the sense that it both optimizes its own cost and its strategy is based upon the assumption that the other agents are rational. In other words each agent believes (i.e., has as a hypothesis in the derivation of its strategy) the other agents are optimizers.

Then under the rationality assumption it is possible to approximate the linking term v_i by a purely deterministic process z^*, and if a deterministic tracking control law is employed by the i-th agent, its optimality loss is negligible in large population conditions. Hence, over the large population, all agents would tend to adopt such a tracking based control strategy if an approximating z^* were to be given.

However, we stress that the rationality notion is only used to construct the aggregation procedure, and the main theorems in the paper will be based solely upon their mathematical assumptions.

A. State Aggregation

Assume $z^* \in C_b[0,\infty)$ is given for approximation of the mass effect, and $s_i \in C_b[0,\infty)$ is a solution to (9). For the i-th agent, after applying the optimal tracking based control law $\widehat{u}_i$ in Proposition 3.1, the closed-loop equation is

$$dz_i = (a_i - \frac{b^2}{r}\Pi_i)z_i dt - \frac{b^2}{r}s_i dt + \sigma_i dw_i. \qquad (11)$$

Denoting $\overline{z}_i(t) = Ez_i(t)$ and taking expectation on both sides of (11) yields

$$\frac{d\overline{z}_i}{dt} = (a_i - \frac{b^2}{r}\Pi_i)\overline{z}_i - \frac{b^2}{r}s_i, \qquad (12)$$

where the initial condition is $\overline{z}_i|_{t=0} = Ez_i(0)$.

We further define the population average of means (simply called population mean) as $\overline{z}^{(n)} \stackrel{\triangle}{=} \frac{1}{n}\sum_{i=1}^n \overline{z}_i$. Note that in the case all agents have i.i.d. dynamics the evolution of $\overline{z}^{(n)}$ is simply expressed using the dynamics of any $\overline{z}_i$ combined with the initial condition $\overline{z}^{(n)}|_{t=0}$ [7].

So far, the individual reaction is determined in a straightforward manner if a mass effect z^* is given *a priori*.

Here one naturally comes up with the important question: how is z^* chosen to approximate the overall influence of all other players on the given player, and in what way does it capture the dynamic behaviour of the collection of many individuals? Since we wish to have $z^*(t) \approx z^*_{-i} = \gamma(\frac{1}{n}\sum_{k\neq i}^{n} z_k + \eta)$, for large n it is reasonable to express z^* in terms of the population mean $\overline{z}^{(n)}$ as

$$z^*(t) = \gamma(\overline{z}^{(n)}(t) + \eta). \tag{13}$$

After introducing such an equality relation, a dynamic interaction is built up between the individual and the mass: by averaging over the individual mean trajectories, the mass effect z^* is constructed, in response to which the individuals, in turn, optimize their own objectives.

Our analysis below will be based upon the observation that the large population limit may be employed to determine the effect of the mass of the population on any given individual, and that the population limit is characterized by an empirical distribution for parametrization of individual dynamics, which is assumed to exist. Specifically, our interest is in the case when a_i, $i \geq 1$, is "adequately randomized" in the sense that the population exhibits certain statistical properties. In this context, the association of the value a_i, $i \geq 1$ and the specific index i plays no essential role, and the more important fact is the frequency of occurrence of a_i on different segments in the range space of the sequence a_i, $i \geq 1$.

Within this setup, we assume that the sequence a_i, $i \geq 1$, has an empirical distribution function $F(a)$ for which a more detailed specification will be stated in Section V. For the Riccati equation (7), when the coefficient a is used in place of a_i, denote the corresponding solution by Π_a. Accordingly, we express $\beta_1(a)$ and $\beta_2(a)$ when a and Π_a are substituted into (8). Straightforward calculation gives

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

$$\beta_1(a) = -\tfrac{\rho}{2} + \sqrt{(a-\tfrac{\rho}{2})^2 + \tfrac{b^2}{r}}, \tag{14}$$

$$\beta_2(a) = \tfrac{\rho}{2} + \sqrt{(a-\tfrac{\rho}{2})^2 + \tfrac{b^2}{r}}. \tag{15}$$

Example 4.1: $a = 1$, $b = 1$, $\sigma = 0.3$, $\rho = 0.5$, $\gamma = 0.6$, $r = 0.1$, $\eta = 0.25$. We get $\Pi = 0.4$, $\beta_1 = 3$, $\beta_2 = 3.5$. □

To simplify the aggregation procedure we assume zero mean for initial conditions of all agents, i.e., $Ez_i(0) = 0$, $i \geq 1$. The above analysis suggests we introduce the equation system:

$$\rho s_a = \frac{ds_a}{dt} + as_a - \frac{b^2}{r}\Pi_a s_a - z^*, \tag{16}$$

$$\frac{d\overline{z}_a}{dt} = (a - \frac{b^2}{r}\Pi_a)\overline{z}_a - \frac{b^2}{r}s_a, \tag{17}$$

$$\overline{z} = \int_{\mathcal{A}} \overline{z}_a dF(a), \tag{18}$$

$$z^* = \gamma(\overline{z} + \eta). \tag{19}$$

In the above, each individual equation is indexed by the parameter a. For the same reasons as noted in Proposition 3.3, here it is unnecessary to specify the initial condition for s_a. (17) with $\overline{z}_a|_{t=0} = 0$ is based upon (12). Hence $\overline{z}_a$ is regarded as the expectation given the parameter a in the individual dynamics. Also, in contrast to the arithmetic average for $\overline{z}^{(n)}$ in (13), (18) is derived using an empirical distribution function $F(a)$ is for the sequence of parameters $a_i \in \mathcal{A}$, $i \geq 1$, with the range space $\mathcal{A}$. (19) is the large population limit form for the equality relation (13).

Remark: In the more general case with non-zero $Ez_i(0)$, we may introduce a joint empirical distribution $F_{a,z}$ for the two dimensional sequence $\{(a_i, Ez_i(0)), i \geq 1\}$. Then the function in (17) is to be labelled by both the dynamic parameter a and an associated initial condition, and furthermore, the integration in (18) is computed with respect to $F_{a,z}$. In this paper we only consider the case of zero mean $Ez_i(0)$ in order to avoid notational complication. □

(**H1**) $\beta_1(a) > 0$ for $a \in \mathcal{A}$ and $\int_{\mathcal{A}} \frac{M}{\beta_1(a)\beta_2(a)} dF(a) < 1$, where $M = \frac{b^2\gamma}{r}$ and $\beta_1(a), \beta_2(a)$ are defined by (14)-(15). Here $\mathcal{A}$ is an interval containing all a_i, $i \geq 1$ and $F(a)$ is the empirical distribution function for $\{a_i, i \geq 1\}$, which is assumed to exist. □

(**H2**) All agents have zero mean initial condition, i.e. $Ez_i(0) = 0$, $i \geq 1$. □

Proposition 4.2: If $b^2 > \frac{r\rho^2}{4}$, then $\beta_1(a) > 0$ for all $a \in (-\infty, \infty)$. □

Remark: Under (**H1**), we have $-\beta_2(a) < -\beta_1(a) < 0$ where $-\beta_1(a)$ is the stable pole of the closed-loop system for the agent with parameter a. $|\beta_1(a)|$ measures the stability margin. The ratio $\frac{M}{\beta_1(a)\beta_2(a)} = \frac{b^2\gamma}{r\beta_1(a)[\beta_1(a)+\rho)]}$ depends on the stability margin and γ. □

The following procedure is used to illustrate the interaction between the individual and the mass. First, given $z^* \in C_b[0,\infty)$, Proposition 3.3 implies that equation (16) leads to the bounded solution

$$s_a(t) = -e^{\beta_2(a)t}\int_t^\infty e^{-\beta_2(a)\tau} z^*(\tau)d\tau \stackrel{\triangle}{=} \mathcal{T}_1 z^*. \tag{20}$$

By use of (20) and (17), we compute the individual mean trajectory $\overline{z}_a$, which combined with (18)-(19) leads to

$$z^*(t) = \frac{\gamma b^2}{r}\int_{\mathcal{A}}\int_0^t\int_s^\infty e^{-\beta_1(a)(t-s)} e^{\beta_2(a)s} e^{-\beta_2(a)\tau} \times z^*(\tau)d\tau ds dF(a) + \gamma\eta \stackrel{\triangle}{=} (\mathcal{T}z^*)(t). \tag{21}$$

Lemma 4.3: Under (**H1**), we have $\mathcal{T}x \in C_b[0,\infty)$, for any $x \in C_b[0,\infty)$. □

Theorem 4.4: Under (**H1**), the map $\mathcal{T} : C_b[0,\infty) \to C_b[0,\infty)$ has a unique fixed point which is uniformly Lipschitz continuous on $[0,\infty)$.

Proof: We only show there exists a unique fixed point. By Lemma 4.3, $\mathcal{T}$ is a map from the Banach space $C_b[0,\infty)$ to itself. For any $x, y \in C_b[0,\infty)$ we have

$$|(\mathcal{T}x - \mathcal{T}y)(t)| \leq |x-y|_\infty \int_{\mathcal{A}} \frac{M}{\beta_1(a)\beta_2(a)} dF(a). \tag{22}$$

Then from (**H1**) it follows that $\mathcal{T}$ is a contraction and therefore has a unique fixed point $z^* \in C_b[0,\infty)$. □

Theorem 4.5: Under (**H1**)-(**H2**), the equation system (16)-(19) admits a unique bounded solution. □

B. The Virtual Agent, Policy Iteration and Attraction to Mass Behaviour

We proceed to investigate certain asymptotic properties on the interaction between the individual and the mass, and the formulation shall be interpreted in the large population limit (i.e., an infinite population) context. Assume each agent is assigned a cost according to (6), i.e.,

$$J_i(u_i, z^*) = E\int_0^\infty e^{-\rho t}\{[z_i - z^*]^2 + ru_i^2\}ds, \ i \geq 1. \quad (23)$$

We now introduce a so-called *virtual agent* to represent the mass effect and use $z^* \in C_b[0,\infty)$ to describe the behaviour of the virtual agent. Here the virtual agent acts as a passive player in the sense that z^* appears as an exogenous function of time to be tracked by the agents.

Then after each selection of the set of individual control laws, a new z^* will be induced as specified below; subsequently, the individual shall consider its optimal policy (over the whole time horizon) to respond to this updated z^*. Thus, the interplay between a given individual and the virtual agent representing the mass may be described as a sequence of virtual plays which may be employed by the individual as a calculation device to learn the mass behaviour. In the following policy iteration analysis in function spaces, we take the virtual agent as a *passive leader* and the individual agents as *active followers*.

Now, we describe the iterative update of an agent's policy from its *policy space*. For a fixed iteration number $k \geq 0$, suppose that there is *a priori* $z^{*(k)} \in C_b[0,\infty)$. Then by Proposition 3.1 the optimal control for the i-th agent using the cost (23) with respect to $z^{*(k)}$ is given as $u_i^{(k+1)} = -\frac{b}{r}(\Pi_i z_i + s_i^{(k+1)})$ where $s_i^{(k+1)} \in C_b[0,\infty)$ is given by

$$\rho s_i^{(k+1)} = \frac{ds_i^{(k+1)}}{dt} + a_i s_i^{(k+1)} - \frac{b^2}{r}\Pi_i s_i^{(k+1)} - z^{*(k)}.$$

By Proposition 3.3, the unique solution $s_i^{(k+1)} \in C_b[0,\infty)$ may be represented by $s_i^{(k+1)} = -e^{\beta_2 t}\int_t^\infty e^{-\beta_2\tau}z^{*(k)}(\tau)d\tau$.

Subsequently, the control laws $\{u_i^{(k+1)}, i \geq 1\}$ produce a mass trajectory $\overline{z}^{(k+1)} = \int_{\mathcal{A}} \overline{z}_a^{(k+1)}dF(a)$, where

$$\frac{d\overline{z}_a^{(k+1)}}{dt} = -\beta_1(a)\overline{z}_a^{(k+1)} - \frac{b^2}{r}s_a^{(k+1)}, \quad (24)$$

with initial condition $\overline{z}_a^{(k+1)}|_{t=0} = 0$ by (**H2**). Then the virtual agent's state (as a function) z^* corresponding to $u_i^{(k+1)}$ is determined as $z^{*(k+1)} = \gamma(\overline{z}^{(k+1)} + \eta)$. From the above and using the operator introduced in (21), we get the recursion for $z^{*(k)}$ as $z^{*(k+1)} = \mathcal{T}z^{*(k)}$, where $z^{*(k+1)}|_{t=0} = \gamma(\overline{z}^{(k+1)}|_{t=0} + \eta) = \gamma\eta$ for all k.

By the iterative adjustments of the individual strategies in response to the virtual agent, we induce the mass behaviour by a sequence of functions $z^{*(k)} = \mathcal{T}z^{*(k-1)} = \mathcal{T}^k z^{*(0)}$. Now we establish that as the population grows, a statistical mass equilibrium exists and it is globally attracting.

Proposition 4.6: Under (**H1**)-(**H2**), $\lim_{k\to\infty} z^{*(k)} = z^*$ for any $z^{*(0)} \in C_b[0,\infty)$, where $z^* \in C_b[0,\infty)$ is given by (16)-(19). □

C. Explicit solution with uniform agents

In the special case of uniform agents, $F(a)$ in (**H1**) degenerates to point mass. Omitting the subscript a for the functions involved, then (16)-(19) specializes to

$$\rho s = \frac{ds}{dt} + as - \frac{b^2}{r}\Pi_a s - z^*, \quad (25)$$

$$\frac{d\overline{z}}{dt} = (a - \frac{b^2}{r}\Pi_a)\overline{z} - \frac{b^2}{r}s, \quad (26)$$

$$z^* = \gamma(\overline{z} + \eta). \quad (27)$$

Here we shall take a general initial condition $\overline{z}(0)$ which is not necessarily zero. Setting the derivatives in (25)-(27) to zero, we write a set of linear algebraic equations which under (**H1**) has a unique solution $(s_\infty, \overline{z}_\infty, z_\infty^*)$ [7].

Proposition 4.7: Under (**H1**), the unique bounded solution $(\overline{z}, s)$ in (25)-(26), is given by $\overline{z}(t) = \overline{z}_\infty + (\overline{z}(0) - \overline{z}_\infty)e^{\lambda_1 t}$, $s(t) = s_\infty + \frac{\gamma}{\beta_2 - \lambda_1}(\overline{z}_\infty - \overline{z}(0))e^{\lambda_1 t}$, where $\lambda_1 = \frac{\rho - \sqrt{\rho^2 + 4(\beta_1\beta_2 - M)}}{2} < 0$, $\beta_1 = -a + \frac{b^2}{r}\Pi$, $\beta_2 = -a + \frac{b^2}{r}\Pi + \rho$ and $\beta_1\beta_2 - M > 0$. □

V. THE DECENTRALIZED ε-NASH EQUILIBRIUM

Let $J_i(u_i, v_i(u_1, \cdots, u_{i-1}, \cdots, u_{i+1}, \cdots, u_n))$ denote the individual cost with respect to the linking term $v_i = \gamma(\frac{1}{n}\sum_{k\neq i}^n z_k(u_k) + \eta)$ for the i-th player when the k-th player takes control u_k, $1 \leq k \leq n$.

$$J_i(u_i, v_i(u_1^0, \cdots, u_{i-1}^0, u_{i+1}^0 \cdots, u_n^0))$$
$$\triangleq E\int_0^\infty e^{-\rho t}\{[z_i(u_i) - \gamma(\frac{1}{n}\sum_{k\neq i}^n z_k(u_k^0) + \eta)]^2 + ru_i^2\}dt,$$

where $z_k(u_k^0) = z_k(u_k^0(z^*, z_k))$. Here we use u_i^0 to denote the optimal tracking based control law,

$$u_i^0 = -\frac{b}{r}(\Pi_i z_i + s_i), \quad (28)$$

where s_i and the associated z^* are derived from (16)-(19). It should be noted that in the following asymptotic analysis the control law u_i^0 for the i-th agent among a population of n agents is constructed using the limit empirical distribution $F(a)$ involved in (18).

Within the context of a population of n agents, for any $1 \leq k \leq n$, the k-th agent's admissible control set $\mathcal{U}_k$ consists of all u_k adapted to the σ-algebra $\sigma(z_i(s), s \leq t, 1 \leq i \leq n)$. In this setup we give the definition.

Definition 5.1: A set of controls $u_k \in \mathcal{U}_k, 1 \leq k \leq n$, for n players is called an *ε-Nash equilibrium* with respect to the costs $J_k, 1 \le k \le n$, if there exists $\varepsilon \ge 0$ such that for any fixed $1 \le i \le n$, we have

$$\begin{aligned} &J_i(u_i, v_i(u_1, \cdots, u_{i-1}, u_{i+1} \cdots, u_n)) \\ \le &J_i(u_i', v_i(u_1, \cdots, u_{i-1}, u_{i+1}, \cdots, u_n)) + \varepsilon, \end{aligned} \quad (29)$$

when any alternative $u_i' \in \mathcal{U}_k$ is applied by the i-th player. □ If $\varepsilon = 0$ in (29), then Definition 5.1 specializes to the usual Nash equilibrium [1].

Remark: The admissible control set $\mathcal{U}_k$ is not decentralized since the k-th agent has perfect information on other agents' states. In effect, such admissible control sets lead to a stronger qualification of the ε-Nash equilibrium property for the decentralized control analyzed in this section. □

For the sequence $\{a_i, i \ge 1\}$ we define the empirical distribution associated with the first n agents

$$F_n(x) = \frac{\sum_{i=1}^n 1_{(a_i \le x)}}{n}, \qquad x \in \mathbb{R}. \quad (30)$$

(**H3**) There exits a distribution function F on $\mathbb{R}$ such that $F_n \to F$ weakly as $n \to \infty$, i. e., $\lim_{n\to\infty} F_n(x) = F(x)$ whenever F is continuous at $x \in \mathbb{R}$. □

(**H3'**) There exists a distribution function F such that $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) \to F(x)| = 0$. □

Remark: Obviously (**H3'**) implies (**H3**). If the sequence $a_1^\infty \stackrel{\triangle}{=} \{a_i, i \ge 1\}$ is sufficiently "randomized" such that a_1^∞ is generated by independent observations on the same underlying distribution function F, then with probability one (**H3'**) holds by Glivenko – Cantelli theorem [8]. □

Given the distribution function F and $z^* \in C_b[0, \infty)$, from (16)-(17) it is seen that both s_a and $\overline{z}_a$ may be explicitly expressed as a function of $a \in \mathcal{A}$.

(**H4**) There exists a closed interval $\mathcal{A}$ such that (i) $\{a_i, i \ge 1\} \subset \mathcal{A}$ with $\int_\mathcal{A} dF(a) = 1$, and (ii) $\sup_{a\in\mathcal{A}} |\overline{z}_a|_\infty < \infty$, and (iii) $\lim_{a'\to a} \sup_t |\overline{z}_a(t) - \overline{z}_{a'}(t)| = 0$ with a vanishing rate depending only on $|a-a'|$, for $a, a' \in \mathcal{A}$. In addition, $\sup_{i\ge1}[\sigma_i^2 + Ez_i^2(0)] < \infty$. □

Proposition 5.2: Assume (**H1**) holds and in addition, there exists $\widehat{\varepsilon} > 0$ such that $\beta_1(a) \ge \widehat{\varepsilon}$ for all $a \in \mathcal{A}$. Then $\overline{z}_a(t)$ satisfies conditions (ii) and (iii) in (**H4**). □

Now we define

$$\epsilon_n(t) = |\int_\mathcal{A} \overline{z}_a(t) dF_n(a) - \int_\mathcal{A} \overline{z}_a(t) dF(a)|, \ t \ge 0. \quad (31)$$

Lemma 5.3: Under (**H3**)-(**H4**), we have $\lim_{n\to\infty} \overline{\epsilon}_n \stackrel{\triangle}{=} \lim_{n\to\infty} \sup_{t\ge0} \epsilon_n(t) = 0$, where $\epsilon_n(t)$ is given by (31). □

Lemma 5.4: Under (**H1**)-(**H4**), for z^* determined by (16)-(19), we have

$$E\int_0^\infty e^{-\rho t}[z^* - \gamma(\frac{1}{n}\sum_{k\ne i}^n z_k(u_k^0) + \eta)]^2 dt = O(\overline{\epsilon}_n^2 + \frac{\gamma^2}{n}),$$

$$|J_i(u_i^0, \gamma(\frac{1}{n}\sum_{k\ne i}^n z_k(u_k^0) + \eta)) - J_i(u_i^0, z^*)| = O(\overline{\epsilon}_n + \frac{\gamma}{\sqrt{n}}),$$

where $\overline{\epsilon}_n$ is given in Lemma 5.3 and the state $z_k(u_k^0)$, $k \ne i$, is generated by the control law u_k^0 given by (28). □

Theorem 5.5: Under (**H1**)-(**H4**), the set of controls $u_i^0, 1 \le i \le n$, for the n players is an ε-Nash equilibrium with respect to the costs $J_i(u_i, \gamma(\frac{1}{n}\sum_{k\ne i}^n z_i(u_k) + \eta))$, i.e.,

$$\begin{aligned} & J_i(u_i^0, \gamma(\tfrac{1}{n}\textstyle\sum_{k\ne i}^n z_i(u_k^0) + \eta)) - \varepsilon \\ \le \ & \inf_{u_i} J_i(u_i, \gamma(\tfrac{1}{n}\textstyle\sum_{k\ne i}^n z_i(u_k^0) + \eta)) \\ \le \ & J_i(u_i^0, \gamma(\tfrac{1}{n}\textstyle\sum_{k\ne i}^n z_i(u_k^0) + \eta)) \end{aligned}$$

where $0 < \varepsilon = O(\overline{\epsilon}_n + \frac{\gamma}{\sqrt{n}})$ with $\overline{\epsilon}_n$ given in Lemma 5.3, u_k^0 is the control law (28) for the k-th player, and u_i is any alternative control which depends on $(t, z_1, \cdots, z_n)$. □

VI. Concluding Remarks

Within the state aggregation framework, in the case an individual has inaccurate statistics on the dynamics over the competing population, it will naturally optimize with respect to an incorrectly calculated mass trajectory. Then the interesting issue concerns (i) the offset between its predicted mass effect and the observed one, and (ii) the offset between its actually attained cost and the expected cost as calculated by the deviant individual. The magnitude of the offset between the predicated mass effect and the actual one is of importance in that it may lead the individual to detect its statistical inaccuracy. On the other hand, a continuous dependence of these offsets on the statistical inaccuracy is useful to show robustness of the decentralized control design based on aggregation. The analysis is given in [8] for an isolated agents with *inaccurate* estimate of the density $p(a)$ associated with the distribution $F(a)$.

Further generalization of the methodology to nonlinear stochastic models and the investigation of statistical mechanics methods for such cost-coupled systems are of interest.

References

[1] J.-P. Aubin. *Optima and Equilibria: An Introduction to Nonlinear Analysis*, Springer, 2nd Ed., 1998.

[2] E.J. Balder. Remarks on Nash equilibria for games with additively coupled payoffs, *Economic Theory*, vol.9, pp.161-167, 1997.

[3] T. Basar and Y.-C. Ho. Informational properties of the Nash solutions of two stochastic nonzero-sum games. *J. Econ. Theory*, vol.7, pp.370-387, 1974.

[4] A. Bensoussan, M. Crouhy, and J.M. Proth. *Mathematical Theory of Production Planning*, North Holland, Amsterdam, 1983.

[5] E.J. Green. Continuum and finite-player noncooperative models of competition, *Econometrica*, vol.52, no.4, pp.975-993, 1984.

[6] M. Huang, P.E. Caines, C.D. Charalambous, and R.P. Malhamé. Stochastic power control for wireless systems: classical and viscosity solutions. *Proc. IEEE CDC*, Orlando, FL., pp.1037-1042, Dec. 2001.

[7] 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. 42nd IEEE Conf. Decision and Control*, Hawaii, pp.98-103, Dec 2003.

[8] M. Huang, P.E. Caines, and R.P. Malhamé. Large-population cost-coupled LQG problems: individual-mass behaviour and decentralized ε-Nash equilibria. Submitted to *IEEE Trans. Automat. Contr.*, Sept. 2004.

[9] M. Huang, R.P. Malhamé, and P.E. Caines. On a class of large-scale cost-coupled Markov games with applications to decentralized power control. *Proc. the 43rd IEEE CDC*, Bahamas, Dec. 2004.

[10] V. E. Lambson. Self-forcing collusion in large dynamic markets. *J. Econ. Theory*, vol. 34, pp. 282-291, 1984.

[11] K. Yosida. *Functional Analysis*, 6th Ed., Springer-Verlag, 1980.