Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

MoB11.5

Nash Strategies and Adaptation for Decentralized Games Involving Weakly-coupled Agents

Minyi Huang, Roland P. Malhamé and Peter E. Caines

Abstract— We consider dynamic games in large population conditions where the agents evolve according to non-uniform dynamics and are weakly coupled via their individual dynamics and costs. A state aggregation technique is employed to obtain a set of decentralized control laws for the individuals which ensures closed-loop stability and possesses an ε-Nash equilibrium property. We then propose a scheme for Nash strategy adaptation when the agents have unknown parameters. The issue of transient performance improvement is addressed by introducing dither signals.

I. INTRODUCTION

In this paper, we study the optimization of large-scale linear systems wherein many agents are each coupled with others via the individual dynamics and costs. The study of such large-scale weakly coupled systems is motivated by a variety of complex phenomena arising in engineering and socio-economic settings, for instance, dynamic economic models involving competing agents [3], [10], [7], and power control in wireless communications where different users compete for quality of services [6], [8]. The model studied is also related to research on swarming, flocking and formation control of autonomous mobile agents, where each agent has its individual dynamics in which an average effect by all others or the surrounding agents acts as a driving term; see, e.g., [4], [20], [12].

In the literature, within the optimal control context weakly interconnected systems were studied in [1]. For differential games with weakly coupled agents, iterative computational techniques for ε-Nash strategies can be found in [15] for linear models, and in [19] for nonlinear models. In contrast to existing work (especially for dynamic LQG games [14]), our concentration is on games with large populations. We analyze the ε-Nash equilibrium properties for a control law by which each individual optimizes using *local information* while interacting with the average effect of all agents taken together, hereon referred to as "the mass". In preceding work [6] we considered the LQG game for a population of uniform agents with only cost coupling and introduced a state aggregation procedure for the design of decentralized control with an ε-Nash equilibrium property, and this

methodology is generalized in [7] to a system of non-uniform agents described by a randomized parametrization for agents' dynamics across the population.

In this paper we consider the more general model where the aggregated population effect is incorporated into the individual dynamics. The resulting dynamic coupling leads to more complicated mutual impact between the agents than in [7], and for establishing our equilibrium results, the performance estimate heavily relies on the closed-loop behaviour of the overall system of all dynamically coupled agents. Based upon the empirical parameter distribution for non-uniform agents and the interaction between the individual and mass, we develop *state aggregation* for the underlying dynamic models, and our approach differs from the well-known aggregation techniques based upon time-scales which lead to a form of hierarchical optimization [18], [16].

Within the noncooperative game setup, an overall rationality assumption for the large population serves as a motivating point for our state aggregation, and implies for a given population effect, each agent tends to apply an optimal tracking action. The key step is then to construct a mutually consistent pair of mass effect and individual strategies such that the latter not only possess optimality with respect to the former for the individuals but also collectively produce it.

In a further step, we consider adaptive strategy selection for the agents. As a motivating example, in wireless power control applications, each agent needs to adjust its power for its own quality of service satisfaction, and such a process is accompanied by adaptively estimating its channel states with the aid of pilot signals [17], [8]. In contrast to classical adaptive control, this part aims to initiate cooperative adaptation/learning in large scale systems of random agents driving by individual incentives. The previously obtained Nash solution is intended as a target for adaptation using available measurements. The cost structure with discount naturally calls for an appropriate parameter adaptation during the transient phase in order to attain an ideal performance.

In this paper we omit all proofs. For the proof of the main theorems on asymptotic equilibrium analysis, see [9].

Minyi Huang is with the Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia. Email: m.huang@ee.mu.oz.au

Roland P. Malhamé is with the Department of Electrical Engineering, École Polytechnique de Montréal, Montreal H3C 3A7, Canada. Email: roland.malhame@polymtl.ca

Peter E. Caines is with the Department of Electrical and Computer Engineering, McGill University, Montreal H3A 2A7, Canada. Email: peterc@cim.mcgill.ca

II. THE WEAKLY COUPLED SYSTEMS

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

$$dz_i = (a_i z_i + b_i u_i)dt + \alpha z^{(n)}dt + \sigma_i dw_i, \ 1 \le i \le n, \ t \ge 0, \tag{1}$$

where $\{w_i, 1 \le i \le n\}$ denotes n independent standard scalar Wiener processes and $z^{(n)} = \frac{1}{n}\sum_{i=1}^{n} z_i$, $\alpha \in \mathbb{R}$.

Hence, $z^{(n)}$ may be looked at as a nominal driving term imposed by the population. The initial states $z_i(0)$ are mutually independent and 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 agent (also to be called player).

We investigate the behaviour of the agents when they interact with each other through specific coupling terms appearing in their individual costs:

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

For simplicity of analysis we assume in this paper that

$$b_i = b > 0, \qquad 1 \le i \le n.$$

In particular we assume the cost-coupling to be of the following form:

$$v_i = \Phi(z^{(n)}) = \Phi(\frac{1}{n}\sum_{k=1}^n z_k),$$

where Φ is a continuous function on $\mathbb{R}$. The linking term v_i gives a measure of the average effect of the mass formed by all agents in this type of group tracking problem. Here we assume $\rho, r > 0$ and unless otherwise stated, throughout the paper z_i is described by the dynamics (1).

III. THE PRELIMINARY LINEAR TRACKING PROBLEM

In the tracking analysis, we begin by replacing the average driving term $z^{(n)}$ in (1) by a deterministic function f. This suggests we introduce the auxiliary dynamics

$$d\hat{z}_i = a_i\hat{z}_i dt + bu_i dt + \alpha f dt + \sigma_i dw_i, \quad (3)$$

where f is bounded and continuous on $[0,\infty)$. For distinction, the state variable $\hat{z}_i$ is used in (3), and all other terms are specified in a similar manner as in (1).

For large n, we intend to approximate $v_i = \Phi(\frac{1}{n}\sum_{k=1}^n z_k)$ by a *deterministic* continuous function z^* defined on $[0,\infty)$. For a given z^* without being related to f above, we construct the *individual cost* associated with (3) as follows:

$$J_i^0(u_i, z^*) = E \int_0^\infty e^{-\rho t}\{[\hat{z}_i - z^*]^2 + ru_i^2\}dt. \quad (4)$$

We consider bounded z^*. For minimizing J_i^0, the admissible control set is $\mathcal{U}_i \triangleq \{u_i | u_i$ adapted to $\sigma(\hat{z}_i(0), w_i(s), s \le t)$, and $E \int_0^\infty e^{-\rho t}(\hat{z}_i^2 + u_i^2)dt < \infty\}$. Define $C_b[0,\infty) \triangleq \{x \in C[0,\infty), |x|_\infty < \infty\}$, where $|x|_\infty = \sup_{t\ge 0}|x(t)|$, for $x \in C[0,\infty)$. With the norm $|\cdot|_\infty$, $C_b[0,\infty)$ is a Banach space.

Let $\Pi_i > 0$ be the solution to the algebraic Riccati equation

$$\rho\Pi_i = 2a_i\Pi_i - \frac{b^2}{r}\Pi_i^2 + 1. \quad (5)$$

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

$$\beta_1 = -a_i + \frac{b^2}{r}\Pi_i, \quad \beta_2 = -a_i + \frac{b^2}{r}\Pi_i + \rho > \frac{\rho}{2}. \quad (6)$$

Proposition 1: Assume (i) $E|\hat{z}_i(0)|^2 < \infty$ and $f, z^* \in C_b[0,\infty)$; (ii) $\Pi_i > 0$ is the solution to (5) and $\beta_1 = -a_i + \frac{b^2}{r}\Pi_i > 0$; and (iii) $s_i \in C_b[0,\infty)$ and $q \in C_b[0,\infty)$ satisfy

$$\rho s_i = \frac{ds_i}{dt} + a_i s_i - \frac{b^2}{r}\Pi_i s_i + \alpha\Pi_i f - z^*, \quad (7)$$

$$\rho q = \frac{dq}{dt} - \frac{b^2}{r}s_i^2 + (z^*)^2 + 2\alpha f s_i + \sigma_i^2\Pi_i. \quad (8)$$

Then the control law

$$\widehat{u}_i = -\frac{b}{r}(\Pi_i\hat{z}_i + s_i) \quad (9)$$

minimizes $J_i^0(u_i, z^*)$ for all $u_i \in \mathcal{U}_i$, and the optimal cost is $J_i^0(\widehat{u}_i, z^*) = \Pi_i E\hat{z}_i^2(0) + 2s(0)E\hat{z}_i(0) + q(0)$. $\qquad\square$

Proposition 2: Under the assumptions of Proposition 1, there exists a unique initial condition $s_i(0) \in \mathbb{R}$ such that the associated solution s_i to (7) is bounded, i.e., $s_i \in C_b[0,\infty)$. And moreover, for the obtained $s_i \in C_b[0,\infty)$, there is a unique initial condition $q(0) \in \mathbb{R}$ for (8) such that the solution $q \in C_b[0,\infty)$. $\qquad\square$

In fact, the unique initial condition for s_i is given by $s_i(0) = -\int_0^\infty e^{-\beta_2\tau}[z^*(\tau) - \alpha\Pi_i f(\tau)]d\tau$ which yields

$$s_i(t) = e^{\beta_2 t}\int_t^\infty e^{-\beta_2\tau}[\alpha\Pi_i f(\tau) - z^*(\tau)]d\tau \in C_b[0,\infty).$$

and any other initial condition yields an unbounded solution.

IV. COMPETITIVE BEHAVIOUR AND CONTINUUM MASS BEHAVIOUR

In the weakly 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.

Combined with the rationality assumption, the specific structure of the driving term $z^{(n)}$ and linking term $v_i = \Phi(z^{(n)})$ will be employed in the approximation developed below. 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 via large population limit

Assume $f \in C_b[0,\infty)$ is given for approximation of $z^{(n)}$, and $s_i \in C_b[0,\infty)$ is a solution to (7) computed with $z^* = \Phi(f)$. For the ith agent, after applying the control law (9), the closed-loop equation is approximated by

$$dz_i = (a_i - \frac{b^2}{r}\Pi_i)z_i dt - \frac{b^2}{r}s_i dt + \alpha f dt + \sigma_i dw_i, \quad (10)$$

where f replaces $z^{(n)}$ in (1). We have

$$\frac{d\overline{z}_i}{dt} = (a_i - \frac{b^2}{r}\Pi_i)\overline{z}_i - \frac{b^2}{r}s_i + \alpha f, \quad (11)$$

where $\overline{z}_i(t) = Ez_i(t)$ and 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 $\overline{z}^{(n)} \triangleq \frac{1}{n}\sum_{i=1}^n \overline{z}_i$.

Since we wish to have $f \approx \frac{1}{n}\sum_{k=1}^{n} z_k$, for large n it is plausible to express

$$f = \overline{z}^{(n)}, \qquad z^*(t) = \Phi(\overline{z}^{(n)}(t)). \qquad (12)$$

In our analysis below, the large population limit is employed to determine the effect of the mass of the population on any given individual. 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. Within this setup, we assume that the sequence $\{a_i, i \geq 1\}$, has an empirical distribution function $F(a)$.

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

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

(H1) There exists a distribution function F on $\mathbb{R}$ such that $F_n \to F$ weakly as $n \to \infty$, i.e., $\lim_{n\to\infty} F_n(x) = F(x)$ if F is continuous at $x \in \mathbb{R}$. □

(H1') There exists a distribution function F on $\mathbb{R}$ such that $F_n \to F$ uniformly as $n \to \infty$, i.e., $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0$. □

Remark. Obviously **(H1')** implies **(H1)**. See [7] for an interpretation of the assumptions in terms of random sampling of $\{a_i, i \geq 1\}$ from an underlying distribution. □

For the Riccati equation (5), when the coefficient a is used in place of a_i, we 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 (6). We have

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

$$\beta_1(a) = -\frac{\rho}{2} + \sqrt{(a - \frac{\rho}{2})^2 + \frac{b^2}{r}}, \qquad (13)$$

$$\beta_2(a) = \frac{\rho}{2} + \sqrt{(a - \frac{\rho}{2})^2 + \frac{b^2}{r}}. \qquad (14)$$

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 consider the large population limit and introduce the equation system:

$$\rho s_a = \frac{ds_a}{dt} + as_a - \frac{b^2}{r}\Pi_a s_a + \alpha \Pi_a \overline{z} - z^*, \qquad (15)$$

$$\frac{d\overline{z}_a}{dt} = (a - \frac{b^2}{r}\Pi_a)\overline{z}_a - \frac{b^2}{r}s_a + \alpha \overline{z}, \qquad (16)$$

$$\overline{z} = \int_{\mathcal{A}} \overline{z}_a dF(a), \qquad (17)$$

$$z^* = \Phi(\overline{z}). \qquad (18)$$

In the above, each individual differential equation is indexed by the parameter a. For the same reasons as in Proposition 2, here it is unnecessary to specify the initial condition $s_a(0)$ for s_a derived from optimal tracking. Equation (16) with $\overline{z}_a|_{t=0} = 0$ is based upon (11). Also, in contrast to the arithmetic average for computing $\overline{z}^{(n)}$ appearing in (12), (17) is derived by use of the empirical distribution function $F(a)$ for the parameters $a_i \in \mathcal{A}$, $i \geq 1$, with the range space $\mathcal{A}$. With a little abuse of terminology, we shall refer to either $\overline{z}$, or in some cases $\Phi(\overline{z})$, as the mass trajectory.

(H2) The function Φ is Lipschitz continuous on $\mathbb{R}$ with a Lipschitz constant $\gamma > 0$, i.e., $|\Phi(y_1) - \Phi(y_2)| \leq \gamma|y_1 - y_2|$ for all $y_1, y_2 \in \mathbb{R}$. □

(H3) For all $a \in \mathcal{A}$, $\beta_1(a) > 0$, and $\int_{\mathcal{A}}[\frac{|\alpha|}{\beta_1(a)} + \frac{b^2(\gamma + |\alpha|\Pi_a)}{r\beta_1(a)\beta_2(a)}]dF(a) < 1$, where $\beta_1(a), \beta_2(a)$ are defined by (13)-(14), $\mathcal{A}$ is a measurable subset of $\mathbb{R}$ and contains $\{a_i, i \geq 1\}$, and $F(a)$ is the empirical distribution function for $\{a_i, i \geq 1\}$. The constant $\gamma > 0$ is specified in **(H2)** □

(H4) All agents have mutually independent initial conditions of zero mean, i.e. $Ez_i(0) = 0$, $i \geq 1$. In addition, $\sup_{i\geq 1}[\sigma_i^2 + Ez_i^2(0)] < \infty$. □

Remark. Under **(H3)**, 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. To avoid triviality for the linking term, we assume $\gamma > 0$ in **(H2)**. □

Given $\overline{z} \in C_b[0, \infty)$, Proposition 2 implies that (15) has the bounded solution $s_a(t) = e^{\beta_2(a)t}\int_t^{\infty} e^{-\beta_2(a)\tau}[\alpha\Pi_a\overline{z}(\tau) - \Phi(\overline{z}(\tau))]d\tau \overset{\triangle}{=} (\Gamma_1\overline{z})(t)$.

Then under **(H4)**, equations (16) and (17) give the fixed point equation below:

$$\overline{z}(t) = \int_{\mathcal{A}} \int_0^t e^{-\beta_1(a)(t-s)}\left[\alpha\overline{z}(s) + \frac{b^2}{r}e^{\beta_2(a)s}\right.$$
$$\times \left[\int_s^{\infty} e^{-\beta_2(a)\tau}[\Phi(\overline{z}(\tau)) - \alpha\Pi_a\overline{z}(\tau)]d\tau\right]\right] ds dF(a)$$
$$\overset{\triangle}{=} (\Gamma\overline{z})(t).$$

Theorem 3: Under **(H2)**-**(H3)**, the map Γ is from $C_b[0, \infty)$ to $C_b[0, \infty)$, and has a unique fixed point which is uniformly Lipschitz continuous on $[0, \infty)$. □

Theorem 4: Under **(H2)**-**(H4)**, the equation system (15)-(18) admits a unique bounded solution. □

B. The diagram for the individual and mass interaction

The state aggregation procedure is illustrated in Fig. 1 based on the individual and mass interaction within a large population limit, where S^{-1} denotes the integrator. The reader should be cautioned that the involved signals are not fed through each block in real time since for each $t > 0$, $s_a(t)$ is computed using the overall trajectory $\overline{z}$ on $[0, \infty)$. Such a diagram is more appropriately described by the so-called virtual play. The key assumption **(H3)** is used to ensure that the "loop gain" for the interconnected operators is less than one, as indicated in the proof of Theorem 3 [9].

In the virtual play, we introduce a *virtual agent* (representing the mass effect) and describe its behaviour by $\overline{z} \in C_b[0, \infty)$. The virtual agent acts as a passive player in the sense that $\overline{z}$ appears as an exogenous function of time and $\Phi(\overline{z})$ is to be tracked by the agents. Then after each selection of the set of individual control laws, a new $\overline{z}$ will be induced; subsequently, the individual shall consider its optimal policy (over the whole time horizon) to respond to this updated $\overline{z}$. Thus, the interplay between a given individual and the virtual agent may be described as a sequence of virtual plays which

may be employed by the individual as a calculation device to eventually learn the mass behaviour.

By the iterative adjustments of the individual strategies in response to the virtual agent, we induce the mass behaviour by a sequence of functions $\overline{z}^{(k)} = \Gamma\overline{z}^{(k-1)} = \Gamma^k\overline{z}^{(0)}$. Then we can show by Theorem 3 that under (H2)-(H4), $\lim_{k\to\infty} \overline{z}^{(k)} = \overline{z}$ for any $\overline{z}^{(0)} \in C_b[0,\infty)$, where $\overline{z}$ is determined by (15)-(18) [9].

It is of interest to note that the virtual play described here has a resemblance in spirit to the so-called tâtonnement in economic theory which was first proposed by Walras in 1874 and formalized in a modern version in terms of ordinary differential equations by Samuelson in 1947 (for relevant literature, the reader is referred to [13] (pp. 620-626) and references therein)[1].

C. Explicit solution with uniform agents

For a system of uniform agents (i.e., $a_i \equiv a$) with $\Phi(z) = \hat{\gamma}(z + \eta)$, a solution to the state aggregation equation system may be explicitly calculated. In this case, (17) is no longer required. Since $\overline{z}$ coincides with $\overline{z}_a$, we simply specify it by (16). The equation system (15)-(18) specializes to

$$\rho s_a = \frac{ds_a}{dt} + as_a - \frac{b^2}{r}\Pi_a s_a + \alpha\Pi_a\overline{z} - z^*, \quad (19)$$

$$\frac{d\overline{z}}{dt} = (a - \frac{b^2}{r}\Pi_a)\overline{z} + \alpha\overline{z} - \frac{b^2}{r}s_a, \quad (20)$$

$$z^* = \Phi(\overline{z}) = \hat{\gamma}(\overline{z} + \eta). \quad (21)$$

Here we shall compute a solution with a general initial condition $\overline{z}(0)$ for (20), which is not necessarily zero. Write a set of steady state equations as follows

$$\begin{cases} \beta_2(a)s_a(\infty) - \alpha\Pi_a\overline{z}(\infty) + z^*(\infty) = 0 \\ -\frac{b^2}{r}s_a(\infty) + (\alpha - \beta_1(a))\overline{z}(\infty) = 0 \\ \hat{\gamma}\overline{z}(\infty) - z^*(\infty) = -\hat{\gamma}\eta. \end{cases} \quad (22)$$

Proposition 5: Under (H2)-(H3), there is a unique solution $(s_a(\infty), \overline{z}(\infty), z^*(\infty))$ to (22), and the unique bounded

[1]In price tâtonnement, given an initial non-equilibrium price, the economic agents will each dynamically adjust its price in a trial and error process where the ensemble of all excess demands is assumed to be announced to all agents by a certain central planner. Such a process is continuously carried out in fictional time and is said to possess tâtonnement stability if it converges to an (Walrasian) equilibrium price.
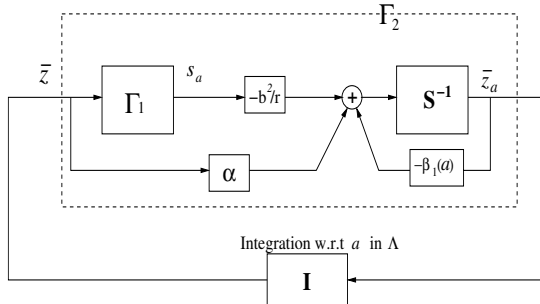

Fig. 1. The loop for individual and mass interaction.

solution $(\overline{z}, s_a)$ in (19)-(20) is given by

$$\overline{z}(t) = \overline{z}(\infty) + (\overline{z}(0) - \overline{z}(\infty))e^{\lambda_1 t},$$

$$s_a(t) = s_a(\infty) + \frac{\hat{\gamma} - \alpha\Pi_a}{\beta_2(a) - \lambda_1}(\overline{z}(\infty) - \overline{z}(0))e^{\lambda_1 t},$$

where $\lambda_1 = \frac{\rho + \alpha - \sqrt{(\rho+\alpha)^2 + 4\Theta}}{2} < 0$ and $\Theta \triangleq \beta_2(a)(\beta_1(a) - \alpha) + \frac{b^2}{r}(\alpha\Pi_a - \hat{\gamma}) > 0$. □

V. THE DECENTRALIZED ε-NASH EQUILIBRIUM

We continue to consider the system of n agents and rewrite the dynamics in Section II as follows:

$$dz_i = (a_i z_i + bu_i)dt + \alpha z^{(n)}dt + \sigma_i dw_i, \ 1 \le i \le n, \ t \ge 0.$$

For the individual cost given by (2), to indicate its dependence on u_i and the set of controls of all other agents, we write it as $J_i(u_i, u_{-i})$ where u_{-i} denotes the row $(u_1, \cdots, u_n)$ with u_i deleted, so that

$$J_i(u_i, u_{-i}) \triangleq E\int_0^\infty e^{-\rho t}\{[z_i - \Phi(\frac{1}{n}\sum_{k=1}^n z_k)]^2 + ru_i^2\}dt. \quad (23)$$

The admissible control set for each agent will be specified later. Let u_i^0 denote the optimal tracking based control law,

$$u_i^0 = -\frac{b}{r}(\Pi_i z_i + s_i), \quad (24)$$

where s_i is derived from (15)-(18) by matching a_i to a, and s_i implicitly depends on $\overline{z}$ therein. It should be emphasized that in the following asymptotic analysis the control law u_i^0 for the ith agent among a population of n agents is constructed using the limit empirical distribution $F(a)$.

For the performance analysis of this section, we need to restrict $\{a_i, i \ge 1\}$ to be bounded.

(H5) The set $\mathcal{A}$ in (H3) is the union of a finite number of disjoint compact intervals and $\hat{\varepsilon} > 0$ is a constant such that $\beta_1(a) \ge \hat{\varepsilon}$ for all $a \in \mathcal{A}$ □.

Notice that under the positivity assumption of $\beta_1(a)$ in (H3), the compactness of $\mathcal{A}$ and continuity of $\beta_1(a)$ ensure that $\hat{\varepsilon}$ specified above always exists.

Concerning notation in this section, we make the important convention as follows. $\overline{z}_a$, given by (16), denotes the individual mean computed in the large population limit context, and $\overline{z}_a^{(n)}$ stands for the mean of the agents with $a_i = a$ in a population of n agents and it is computed using the closed-loop dynamics associated with the control laws u_i^0, $1 \le i \le n$. Also, each of s_a, $\overline{z}$ and z^* is computed using (15)-(18) for the large population limit.

A. Stability guarantees for closed-loop systems

In order to analyze the closed-loop behaviour when u_i^0 is applied by the ith agent, we define the diagonal matrix

$$B_n = \text{Diag}(-\beta_1(a_i))_{n\times n}$$

and denote by $1_{n\times n}$ the $n \times n$ matrix with each entry being one, i.e., $1_{n\times n} = e(n) \times e^T(n)$, where $e^T(n) = [1, \cdots, 1]$. Let $\overline{B}_n = B_n + \frac{\alpha}{n}1_{n\times n}$. Hence $\overline{B}_n$ is a real symmetric matrix

which has n real eigenvalues. We introduce the following property related to closed-loop stability when the control laws u_i^0, $1 \leq i \leq n$, are applied.

(P1) There exist $\mu^* < 0$ and integer $N_0 > 0$ such that for all $n \geq N_0$, $\overline{B}_n \leq \mu^* I_n$. □

In the case $\alpha \leq 0$ and $\inf_{i \geq 1} \beta_1(a_i) = \beta^* > 0$, it is easy to verify **(P1)**. We give a sufficient condition to ensure **(P1)** for the case $\alpha > 0$.

Proposition 6: Assume (i) $\alpha > 0$, (ii) $\beta_1(a_i) \geq \beta^* > 0$ for $i \geq 1$, and (iii) there exists $N_0 > 0$ such that $\sup_{n \geq N_0} \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha}{\beta_1(a_i)} < 1$. Then **(P1)** holds for all $n \geq N_0$. □

Lemma 7: **(H1)**-**(H3)** and **(H5)** imply **(P1)**. □

To prove Lemma 7, it suffices to verify condition (iii) in Proposition 6 for the case $\alpha > 0$.

Lemma 8: Assuming **(H1)**-**(H5)**, we have the estimate

$$\sup_{t \geq 0} E[\sum_{i=1}^{n}(z_i - Ez_i)(t)]^2 = O(n),$$

where the set of states z_i, $1 \leq i \leq n$, corresponds to the control laws u_i^0, $1 \leq i \leq n$, given by (24). □

B. The asymptotic equilibrium analysis

Within the context of a population of n agents, for any $1 \leq k \leq n$, the kth agent's admissible control set $\mathcal{U}_k$ consists of all feedback controls u_k adapted to the σ-algebra $\sigma(z_i(\tau), \tau \leq t, 1 \leq i \leq n)$ (i.e., $u_k(t)$ is a function of $(t, z_1(t), \cdots, z_n(t))$) such that a unique strong solution to the closed-loop system of the n agents exists on $[0, \infty)$. Note that $\mathcal{U}_k$ itself is not restricted to be decentralized.

Definition 9: 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 $J_k, 1 \leq k \leq n$, if there exists $\varepsilon \geq 0$ such that for any fixed $1 \leq i \leq n$, we have

$$J_i(u_i, u_{-i}) \leq J_i(u_i', u_{-i}) + \varepsilon,$$

when any alternative $u_i' \in \mathcal{U}_i$ is applied by the i-th player. □

Now for $t \geq 0$, we define

$$\epsilon_n(t) = |\int_{\mathcal{A}} \overline{z}_a(t) dF_n(a) - \int_{\mathcal{A}} \overline{z}_a(t) dF(a)|, \quad (25)$$

$$\epsilon_n'(t) = |\int_{\mathcal{A}} \overline{z}_a(t) dF_n(a) - \int_{\mathcal{A}} \overline{z}_a^{(n)}(t) dF_n(a)|. \quad (26)$$

Proposition 10: Under **(H1)**-**(H5)**, we have

$$\lim_{n \to \infty} \overline{\epsilon}_n \triangleq \lim_{n \to \infty} \sup_{t \geq 0} \epsilon_n(t) = 0,$$

$$\lim_{n \to \infty} \overline{\epsilon}_n' \triangleq \lim_{n \to \infty} \sup_{t \geq 0} \epsilon_n'(t) = 0,$$

where $\epsilon_n(t)$ and $\epsilon_n'(t)$ are defined by (25)-(26). □

In Theorems 11 and 12, u_i^0 is given by (24); and z^* is computed using (15)-(18) for the large population limit.

Theorem 11: Under **(H1)**-**(H5)**, we have

$$|J_i(u_i^0, u_{-i}^0) - J_i^0(u_i^0, z^*)| = O(\overline{\epsilon}_n + \overline{\epsilon}_n' + \frac{1}{\sqrt{n}}),$$

as $n \to \infty$, where $J_i^0(u_i^0, z^*)$ is the cost introduced in section III w.r.t. z^* and $f = \overline{z}$, $J_i(u_i^0, u_{-i}^0)$ is determined by (23). □

Theorem 12: Under **(H1)**-**(H5)**, the set of controls $u_i^0, 1 \leq i \leq n$, for the n players is an ε-Nash equilibrium with respect to the costs $J_i(u_i, u_{-i})$, $1 \leq i \leq n$, i.e.,

$$J_i(u_i^0, u_{-i}^0) - \varepsilon \leq \inf_{u_i} J_i(u_i, u_{-i}^0) \leq J_i(u_i^0, u_{-i}^0)$$

where $0 < \varepsilon = O(\overline{\epsilon}_n + \overline{\epsilon}_n' + \frac{1}{\sqrt{n}})$ (hence $\varepsilon \to 0$) as $n \to \infty$, and $u_i \in \mathcal{U}_i$ is any alternative control depending on $(t, z_1, \cdots, z_n)$. □

In proving Theorem 12, we need to consider the control perturbation of a given agent. When the control laws change from (u_i^0, u_{-i}^0) to (u_i, u_{-i}^0) for the n agents, a change will accordingly take place for each of the n state components since z_k, $k \neq i$, is coupled with $\frac{1}{n}z_i$ even if u_{-i}^0 remains the same. This makes the performance estimate significantly more difficult than in [6], [7]; see [9] for proof details.

VI. STRATEGY ADAPTATION WITH UNKNOWN INDIVIDUAL PARAMETERS

In this section we formulate a paradigm for strategy adaptation when the agents have unknown parameters in their dynamics. For simplicity, we assume all agents are described by the same set of parameters with a unity coefficient for both the control and the noise, and the cost coupling is given by the linear function $\Phi(z) = \hat{\gamma}(z + \eta)$ as in Section IV-C. Then the individual dynamics reduce to

$$dz_i = az_i dt + u_i dt + \alpha z^{(n)} dt + dw_i, \quad (27)$$

where a and α need to be estimated.

We assume the parameters a and α are contained in the compact intervals $I_1 = [\underline{a}, \overline{a}]$ with a prior distribution μ_1, and $I_2 = [\underline{\alpha}, \overline{\alpha}]$ with a prior distribution μ_2, respectively. Thus $\mu_1(I_1) = 1$ and $\mu_2(I_2) = 1$. Notice that each of the two true parameters is a fixed constant. Furthermore, we assume for any $(a, \alpha) \in I_1 \times I_2$, assumption **(H3)** holds with respect to the dirac measure (for F) at a.

Our underlying assumption is that each agent can measure its own state and the population average $z^{(n)}$.

A. The parameter estimation

Letting $\theta = [a, \alpha]^T$, $\psi_t = [z_i, z^{(n)}]^T(t)$, we write the dynamics (27) in the form

$$dz_i = \theta^T \psi_t dt + u_i dt + dw_i. \quad (28)$$

Let the ith agent's estimate be denoted by $\hat{\theta}_i(t) = [\hat{a}_i, \hat{\alpha}_i]^T(t)$ which is computed by the continuous time recursive least squares algorithm as follows:

$$d\hat{\theta}_i(t) = \Psi_t[dz_i(t) - \hat{\theta}_i(t)^T \psi_t dt], \quad (29)$$

$$d\Psi_t = -\Psi_t \psi_t \psi_t^T \Psi_t dt, \quad (30)$$

where $\Psi_0 > 0$. In fact $\Psi_t = (\int_0^t \psi_s \psi_s ds + \Psi_0^{-1})^{-1}$ for $t \geq 0$. We fix $\overline{a} \in I_1$ and $\overline{\alpha} \in I_2$. The initial condition is taken as $\hat{\theta}_i(0) = [\overline{a}, \overline{\alpha}]^T$ for all i.

It should be noted that although all agents share the same true parameter θ, the estimates $\hat{\theta}_i(t)$, $1 \leq i \leq n$, in general

differ from each other since each agent combines its own state with the population average $z^{(n)}$ to get the estimate.

Since $(\hat{a}_i(t), \hat{\alpha}_i(t))$ is not necessarily in the set $I_1 \times I_2$, we carry out the modification:

$$\hat{a}_i^*(t) = \begin{cases} \hat{a}_i(t) & \text{if } \hat{a}_i(t) \in I_1 \\ \bar{a} & \text{otherwise,} \end{cases}$$

$$\hat{\alpha}_i^*(t) = \begin{cases} \hat{\alpha}_i(t) & \text{if } \hat{\alpha}_i(t) \in I_2 \\ \bar{\alpha} & \text{otherwise.} \end{cases}$$

We denote $\hat{\theta}_i^*(t) = [\hat{a}_i^*(t), \hat{\alpha}_i^*(t)]^T$.

B. The certainty equivalence control law

Now, at time t the ith agent solves the Riccati equation (5) and the state aggregation equation system (19)-(21) (for uniform agents) using $\hat{\theta}_i^*(t)$ obtained by itself to get $\Pi_i^* \triangleq \Pi(\hat{a}_i^*(t))$ and $s_i^*(t) \triangleq s(t, \hat{a}_i^*(t), \hat{\alpha}_i^*(t))$. Here $s_i^*(t)$ denotes the value of the function $s_a(\cdot)$ at time t when the pair (a, α) is substituted by $(\hat{a}_i^*(t), \hat{\alpha}_i^*(t))$. We write the certainty equivalence control law

$$u_i = -\frac{1}{r}(\Pi_i^* z_i + s_i^*), \tag{31}$$

where we have $b = 1$. In future work, it is of interest to analyze the asymptotic property of the parameter estimates obtained by all agents.

C. Transient performance and dither signal selection

We note that under the discounted cost and the associated Nash strategy adaptation, it is generally inadequate to consider only attaining asymptotic consistency for the agents' parameter estimates due to the memory effect of the cost. For the transient phase, from the localized optimal tracking point of view, in order to have efficient reaction to the population average, it is necessary to have good parameter estimates at the early stage so that all of the agents can collectively maintain good prediction of the mass effect.

For improving the transient estimation, it is potentially useful to consider introducing an appropriately diminishing dither signal for each agent. This is effective for producing excitation for improving parameter convergence [2], [11]. However, on the other hand it also tends to incur a higher control energy. These may be termed as the dual effect. The interesting issue is to seek a certain trade-off for the above two effects.

Now based on (31) we may formally write the following modified adaptive control law

$$u_i(t) = -\frac{1}{r}(\Pi_i^* z_i + s_i^*) + \epsilon_k[W_i(t) - W_i(k\Delta)]$$

for $t \in [k\Delta, (k+1)\Delta)$, where $W_i(t)$ is a standard Wiener process independent of $\{w_i, z_i(0), 1 \leq i \leq n\}$ and is called the dither signal. $W_i, 1 \leq i \leq n$, are also mutually independent of each other. $\Delta > 0$ is a selected step size, and $\epsilon_k > 0$, $k \geq 0$, is a vanishing sequence. We may call $\{\epsilon_k, k \geq 0\}$ the dither profile which needs to be designed for transient performance optimization.

We note that for the special case of agents with only cost-coupling, i.e., $\alpha = 0$, it is easy to establish parameter

estimation convergence under very mild conditions, and the dither signal is inessential for convergence insurance. In this case, the role of dither signal largely lies in affecting transient behaviour, and a practical approach is then to optimize the dither profile on a finite window employing the prior distribution information for (a, α). Such a profile selection is essentially related to optimization in function spaces [5]. After a dither profile is computed off-line, one can run adaptive control for the agents. The associated computation methods will be investigated in future work.

REFERENCES

[1] A. Bensoussan. *Perturbation Methods in Optimal Control*. Wiley, New York, 1988.
[2] H.-F. Chen and L. Guo. *Identification and Stochastic Adaptive Control*. Birkhäuser, Boston, 1991.
[3] G. M. Erickson. Differential game models of advertsing competition. *Europ. J. Oper. Res.*, vol. 83, pp. 431-438, 1995.
[4] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamic features of escape panic. *Nature*, vol. 407, pp. 487-490, Sept. 2000.
[5] Y.-C. Ho and J. T. Lee. Granular optimisation: an approach to function optimization. *Proc. 39th IEEE Conf. Decision Control*, Sydney, Australia, pp. 103-111, Dec. 2000.
[6] 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*, Maui, HI, pp. 98-103, Dec. 2003.
[7] M. Huang, P. E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems: generalizations to non-uniform individuals. *Proc. 43rd IEEE Conf. Decision Contr.*, Bahamas, pp. 3453-3458, Dec. 2004. Long version under revision for *IEEE Trans. Automat. Contr.*.
[8] M. Huang, P. E. Caines, and R. P. Malhamé. Uplink power adjustment in wireless communication systems: a stochastic control analysis. *IEEE Trans. Automat. Contr.*, vol. 49, pp. 1693-1708, Oct. 2004.
[9] 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* (GERAD 25th Anniveary Series), E. K. Boukas and R. P. Malhamé Eds., Chapter 9, pp. 215-252, Springer, New York, 2005.
[10] V. E. Lambson. Self-forcing collusion in large dynamic markets. *J. Econ. Theory*, vol. 34, pp. 282-291, 1984.
[11] D. Levanony and P. E. Caines. On persistent excitation for linear systems with stochastic coefficients. *SIAM J. Control and Optimization*, vol. 40, no. 3, pp. 882 - 897, 2002.
[12] Y. Liu and K. M. Passino. Stable social foraging swarms in a noisy environment. *IEEE Trans. Automat. Contr.*, vol. 49, pp. 30-44, Jan. 2004.
[13] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, New York, 1995.
[14] G. P. Papavassilopoulos. On the linear-quadratic-Gaussian Nash game with one-step delay observation sharing pattern. *IEEE Trans. Automat. Control*, vol. 27, no.5, pp. 1065-1071, 1982.
[15] B. Petrovic and Z. Gajic. The recursive solution of linear quadratic Nash games for weakly interconnected systems. *J. Optim. Theory Appl.*, vol. 56, no. 3, pp. 463-477, Mar. 1988.
[16] R. G. Phillips and P. V. Kokotovic. A singular perturbation approach to modelling and control of Markov chains. *IEEE Trans. Automat. Control*, vol. 26, pp. 1087-1094, 1981.
[17] W. G. Phoel and M. L. Honig. Performance of coded DS-CDMA with pilot-assisted channel estimation and linear interference suppression. *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 822-832, 2002.
[18] S. P. Sethi and Q. Zhang. *Hierarchical Decision Making in Stochastic Manufacturing Systems*. Birkhäuser, Boston, 1994.
[19] R. Srikant and T. Basar. Iterative computation of noncooperative equilibria in nonzero-sum differential games with weakly coupled players. *J. Optim. Theory Appl.*, vol. 71, pp. 137-168, Oct. 1991.
[20] H. G. Tanner, A. Jadbabaie, and G. J. Pappas. Stable flocking of mobile agents, Part I: fixed topology. *Proc. 42nd IEEE Conf. Decision Control*, Maui, HI, pp. 2010-2015, Dec. 2003.