Stochastic Adaptive Nash Certainty Equivalence Control: Self-Identification Case

Arman C. Kizilkale and Peter E. Caines

Abstract— **For noncooperative games the Nash Certainty Equivalence (NCE), or Mean Field (MF) methodology developed in previous work provides decentralized strategies which asymptotically yield Nash equilibria. The NCE (MF) control laws use only the local information of each agent on its own state evolution and knowledge of its own dynamical parameters, while the behaviour of the mass is precomputable from knowledge of the distribution of dynamical parameters throughout the mass population.**

Relaxing the a priori information condition introduces the methods of parameter estimation and stochastic adaptive control (SAC) into MF control theory. In particular one may consider incrementally the problems where the agents must estimate: (i) its own dynamical parameters, (ii) the distribution of the population's dynamical parameters [1], and (iii) the distribution of the population's cost function parameters [2]. In this paper we treat the first problem.

Each agent estimates its own dynamical parameters via the recursive weighted least squares (RWLS) algorithm. Under reasonable conditions on the population dynamical parameter distribution, we establish: (i) the strong consistency of the self-parameter estimates; and that (ii) all agent systems are long run average L^2 stable; (iii) the set of controls yields a (strong) ϵ-Nash equilibrium for all ϵ; and (iv) in the population limit the long run average cost obtained is equal to the non-adaptive long run average cost.

I. Introduction

Overview:

The control and optimization of large-scale complex systems is evidently of importance due to their ubiquitous appearance in engineering, industrial, social and economic settings. In many social, economic, and engineering models, the agents involved have conflicting objectives and it is more appropriate to consider optimization based upon individual payoffs or costs. Game theoretic approaches are intended to capture the individual interest seeking nature of agents in many social, economic and manmade systems; however, in a large-scale dynamic model this approach results in an analytic complexity which is in general prohibitively high, and correspondingly leads to few implementable results on dynamic optimization.

The optimization of large-scale linear control systems wherein many agents are each coupled with others via the individual dynamics and the costs in a particular form was studied in [3]. The study of such large-scale weakly coupled systems is motivated by a variety of scenarios, for instance, see, e.g., [4], [5], [6], [7].

Arman C. Kizilkale and Peter E. Caines are with the Department of Electrical and Computer Engineering and the Centre for Intelligent Machines, McGill University, Montreal, Canada. {arman;peterc}@cim.mcgill.ca

In the literature, studies of stochastic dynamic games and team problems go back to 1960s [8], [9], [10]. Within the optimal control context weakly interconnected systems were studied in [11], and in a two player noncooperative nonlinear dynamic game setting, the Nash equilibria was analysed in [12] where the coefficients for the coupling terms in the dynamics and costs are restricted to be sufficiently small. In contrast to these, [3] concentrates on games with large populations.

For noncooperative games with mean field coupling the Nash Certainty Equivalence (NCE), or Mean Field (MF) methodology developed in [13], [3], [14], [15], [16], [17] provides decentralized strategies which asymptotically yield Nash equilibria. The key idea of this methodology is to specify a certain consistency relationship between the individual strategies and the mass effect in the population limit, and each decision-maker can ignore the fine details of the behaviour of any other individual player by only focusing on the overall impact of the population. This procedure leads to decentralized strategies for the individual players in a large but finite population. For this class of game problems, a related approach has been independently developed in [18], [19] where the notion of oblivious equilibrium by use of a mean field approximation for models of many firm industry dynamics is proposed. Another related work has been presented in [20], [21] subject to the assumed existence of a system Nash equilibrium under independent local agent feedback controls.

Stochastic Adaptive Control:

The mean square sample path stability for continuous time linear stochastic adaptive systems via the recursive least squares algorithm for adaptation under a persistent excitation hypothesis was established in [22]. In order to relax the persistent excitation hypothesis, the WLS scheme introduced in [23] was shown to be convergent without stability and excitation assumptions in [24], and a complete solution to the continuous time adaptive LQG control problem under controllability and observability assumptions using the WLS scheme for estimation was subsequently obtained in [25].

MF Stochastic Adaptive Control:

It is important to note that the NCE (MF) control laws in [13], [3], [14], [16], [26], [15] use only the local information of each agent on its own state evolution and knowledge of its own dynamical parameters, while the behaviour of the mass is precomputable from knowledge of the distribution of dynamical parameters throughout the mass population. All this information is assumed known to each agent in the basic non-adaptive NCE (MF) theory.

ISBN 978-963-311-370-7

The relaxation of the requirements of a priori known information above naturally leads to the use of the methods of stochastic adaptive control in the MF stochastic control context.

An initial problem on this path of adaptive MF stochastic systems is that where each agent needs to estimate its own dynamical parameters, while its control actions are permitted to be explicit functions of the parameter distribution of the entire population of competing agents; this is the problem we treat in this paper. A subsequent problem is the generalization where each agent also needs to estimate the distribution of the population's dynamical parameters [1]; and a further generalization is the case where the cost function parameters also vary over the population and this distribution is unknown to each agent and hence must be estimated [2].

In this paper we present a mean field stochastic adaptive control algorithm which for each agent results in (i) the strong consistency of the self-parameter estimates; and when applied by all agents in the system, gives rise to the following properties: (ii) all agent systems are long run average L^2 stable; (iii) the set of controls yields a (strong) ϵ-Nash equilibrium for all ϵ; and (iv) in the population limit the long run average cost obtained is equal to the non-adaptive cost.

II. Problem Formulation

A. Review of Non-Adaptive NCE Stochastic Control

We consider a large population of N stochastic dynamic agents which (subject to independent controls) are stochastically independent, but which shall be cost coupled, where the individual dynamics are defined by

$$dx_i = (\mathbf{A}_i x_i + \mathbf{B}_i u_i)dt + \mathbf{D}dw_i, \quad 1 \le i \le N, \quad t \ge 0, \tag{1}$$

where $x_i \in \mathbb{R}^n$ is the state, $u_i \in \mathbb{R}^m$ is the control input, $\{w_i, 1 \le i \le N\}$ denotes N independent standard Wiener processes in $\mathbb{R}^r$ on a sufficiently large underlying probability space $(\Omega, \mathcal{F}, P)$ such that w is progressively measurable with respect to $\mathcal{F}^w := \{\mathcal{F}_t^w; t \ge 0\}$. The initial states $\{x_i(0), 1 \le i \le N\}$ are mutually independent and also independent of $\mathcal{F}_\infty^w$; $\mathbb{E}w_i^2 = \Sigma_i$ and $\mathbb{E}|x_i(0)|^2 < \infty$. We denote the state configuration by $x = (x_1, \cdots, x_N)^{\mathrm{T}}$, and the population average state by $x^N = (1/N)\sum_{i=1}^N x_i$. The pair of coefficients $\theta_i^{\mathrm{T}} \triangleq [\mathbf{A}_i, \mathbf{B}_i] \in \mathbf{\Theta} \subset \mathbb{R}^{n(n+m)}$, will be called the *dynamical parameters*. The variability of θ_i from agent to agent is used to model a heterogeneous population of agents.

The long run average (LRA) cost function for the agent $A_i, 1 \le i \le N$, is given by

$$J_i^N(u_i, u_{-i}) = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \{\|x_i - m_i^N\|_Q^2 + \|u_i\|_R^2\}dt \quad w.p.1, \tag{2}$$

where $\|x - m^N\|_Q^2 := (x - m^N)^{\mathrm{T}}\mathbf{Q}(x - m^N)$ and $\|u\|_R^2 := u_i^{\mathrm{T}}\mathbf{R}u$. We assume the cost-coupling to be of the form $m^N(\cdot) = m(x^N(t) + \eta), \eta \in \mathbb{R}^n$. The function $u_i(\cdot)$ is the control input of agent A_i and u_{-i} denotes the control inputs of the complementary set of agents $A_{-i} = \{A_j, j \ne i, 1 \le j \le N\}$. The two cost matrices $\mathbf{Q}$ and $\mathbf{R}$ satisfy $\mathbf{Q} = \mathbf{Q}^{\mathrm{T}} \ge 0, \mathbf{R} = \mathbf{R}^{\mathrm{T}} > 0$.

For the basic MF control problem, the following assumptions are adopted:

H 1: All agents have mutually independently distributed initial conditions; $\{w_i, 1 \le i \le N\}$, are mutually independent and independent of the initial conditions, and $\sup_{i\ge1}[\mathrm{Tr}\Sigma_i + \mathbb{E}\|x_i(0)\|^2] < \infty$.

H 2: $\tilde{\mathbf{\Theta}}$ is such that for each $\theta^{\mathrm{T}} = [\mathbf{A}_\theta, \mathbf{B}_\theta] \in \tilde{\mathbf{\Theta}}$, $[\mathbf{A}_\theta, \mathbf{B}_\theta]$ is controllable and $[\mathbf{Q}^{1/2}, \mathbf{A}_\theta]$ is observable.

H 3: Let the parameter set $\mathbf{\Theta}$ be a compact set such that

$$\mathbf{\Theta} \subset \tilde{\mathbf{\Theta}} \subset \mathbb{R}^{n(n+m)}.$$

H 4: The cost-coupling is of the form: $m(1/N\sum_{k=1}^N x_k + \eta)$, $\eta \in \mathbb{R}^n$, where the function $m(\cdot)$ is Lipschitz continuous on $\mathbb{R}^n$ with a Lipschitz constant $\gamma > 0$, i.e. $\|m(x) - m(y)\| \le \gamma\|x - y\|$ for all $x, y \in \mathbb{R}^n$.

The cost coupling function $m^N(\cdot)$ is estimated by a deterministic function $x^*(t), t \ge 0$, and the problem is solved in [3] for the cost function (2) when $m^N(\cdot)$ is substituted for $x^*(t), t \ge 0$. The positive solution is obtained for the following algebraic Riccati equation

$$\mathbf{A}_i^{\mathrm{T}}\mathbf{\Pi}_i + \mathbf{\Pi}_i\mathbf{A}_i - \mathbf{\Pi}_i\mathbf{B}_i\mathbf{R}^{-1}\mathbf{B}_i^{\mathrm{T}}\mathbf{\Pi}_i + \mathbf{Q} = 0; \tag{3}$$

the mass offset function is calculated solving the differential equation

$$\frac{ds_i(t)}{dt} = -\mathbf{A}_i^{\mathrm{T}}s_i(t) + \mathbf{\Pi}_i\mathbf{B}_i\mathbf{R}^{-1}\mathbf{B}_i^{\mathrm{T}}s_i(t) + \mathbf{Q}x^*, \tag{4}$$

Then, the optimal tracking control solution [27] is given by

$$u_i(t) = -\mathbf{R}^{-1}\mathbf{B}_i^{\mathrm{T}}(\mathbf{\Pi}_i x_i(t) + s_i(t)), \quad t \ge 0. \tag{5}$$

We first define the empirical distribution associated with the first N agents:

$$F_\zeta^N(\theta) = \frac{1}{N}\sum_{i=1}^N \mathbb{I}_{(\theta_i<\theta)}, \quad \theta \in \mathbb{R}^{n(n+m)},$$

where $\zeta \in P$ is the population dynamical distribution parameter. Then we employ the following assumption:

H 5: There exists a distribution function F_ζ on $\mathbb{R}^{n(n+m)}$ such that $F_\zeta^N \to F_\zeta$ weakly as $N \to \infty$, i.e., $\lim_{N\to\infty} F_\zeta^N(\theta) = F_\zeta(\theta)$ if F_ζ is continuous at $\theta \in \mathbf{\Theta} \subset \mathbb{R}^{n(n+m)}$.

Definition 2.1: Nash Certainty Equivalence (Mean Field) (NCE) Equation System on $[t, \infty)$:

$$\begin{aligned}
\frac{ds_\theta}{d\tau} &= (-\mathbf{A}_\theta^{\mathrm{T}} + \mathbf{\Pi}_\theta\mathbf{B}_\theta\mathbf{R}^{-1}\mathbf{B}_\theta^{\mathrm{T}})s_\theta + \mathbf{Q}x^*(\tau, \zeta),\\
\frac{d\bar{x}_\theta}{d\tau} &= (\mathbf{A}_\theta - \mathbf{B}_\theta\mathbf{R}^{-1}\mathbf{B}_\theta^{\mathrm{T}}\mathbf{\Pi}_\theta)\bar{x}_\theta - \mathbf{B}_\theta\mathbf{R}^{-1}\mathbf{B}_\theta^{\mathrm{T}}s_\theta,\\
\bar{x}(\tau, \zeta) &= \int_{\mathbf{\Theta}} \bar{x}_\theta dF_\zeta(\theta),\\
x^*(\tau, \zeta) &= m(\bar{x}(\tau, \zeta) + \eta), \quad \tau \le t < \infty.
\end{aligned} \tag{6}$$

■

The convergence of the NCE Equation System using empirical $F_\zeta^N(.)$ to a unique solution x^* is established in [3] under the following technical assumption:

***H** 6:*

$$\|\mathbf{R}^{-1}\|\|\mathbf{Q}\|\gamma \int_{\theta\in\mathbf{\Theta}} \|\mathbf{B}(\theta)\|^2 \left(\int_0^\infty \|e^{\mathbf{A}_*(\theta)\tau}\| d\tau\right)^2 dF_\zeta(\theta) < 1. \quad (7)$$

For the optimality analysis, we first introduce two admissible control sets. The set of control inputs $\mathcal{U}_{g,i}$, the global observation control set, consists of all feedback controls adapted to $\{\theta_1^N;\, F_\zeta(\theta);\, \mathcal{F}_t^N, t \geq 0;\, \mathbf{Q}, \mathbf{R}\}$ and the set of control inputs $\mathcal{U}_{l,i}$, the local observation control set of agent A_i, consists of the feedback controls adapted to the set $\{\theta_i;\, F_\zeta(\theta);\, \mathcal{F}_{i,t}, t \geq 0;\, \mathbf{Q}, \mathbf{R}\}$. The σ-field $\mathcal{F}_{i,t}$ is an increasing family of the σ-field generated by the set of $\{x_i(\tau);\, 0 \leq \tau \leq t\}$, and $\mathcal{F}_t^N$ is an increasing family of the σ-field generated by the set of $\{x_j(\tau);\, 0 \leq \tau \leq t, 1 \leq j \leq N\}$.

Definition 2.2: Given $\epsilon > 0$, the set of controls $\mathcal{U}^0 = \{u_i^0; 1 \leq i \leq N\}$ generates an *ϵ-Nash Equilibrium* w.r.t. the costs $\{J_i; 1 \leq i \leq N\}$ if for each i, $1 \leq i \leq N$,

$$J_i^N(u_i^0, u_{-i}^0) - \epsilon \leq \inf_{u_i \in \mathcal{U}_{g,i}} J_i^N(u_i, u_{-i}^0) \leq J_i^N(u_i^0, u_{-i}^0).$$

■

Theorem 2.1: Non-Adaptive NCE Theorem [28]

Let **H**1-**H**5 hold. The NCE Control Law generates a set of controls $\mathcal{U}_{nce}^N = \{u_i^0; 1 \leq i \leq N\},\ 1 \leq N < \infty$, with

$$u_i^0(t) = -\mathbf{R}^{-1}\mathbf{B}_i^{\mathrm{T}}(\mathbf{\Pi}_i x_i(t) + s_i(t)), \quad t \geq 0, \quad (8)$$

s.t.

(i) The NCE Equations (6) have a unique solution.

(ii) All agent systems $S(A_i),\ 1 \leq i \leq N,$ are second order stable.

(iii) $\{\mathcal{U}_{nce}^N; 1 \leq N < \infty\}$ yields an ϵ-Nash equilibrium for all ϵ,
i.e. $\forall \epsilon > 0\ \exists N(\epsilon)$ s.t. $\forall N \geq N(\epsilon)$

$$J_i^N(u_i^0, u_{-i}^0) - \epsilon \leq \inf_{u_i \in \mathcal{U}_{g,i}} J_i^N(u_i, u_{-i}^0) \leq J_i^N(u_i^0, u_{-i}^0). \quad (9)$$

■

Conceptually, Theorem 2.1 may be paraphrased to say that individual competitive actions against the mass effect collectively produce the mass behavior, and hence the equilibrium is stable in the Nash game theoretic sense. In the proof of Theorem 2.1, the results are first established for an infinite population and then are shown to be approximated by a large finite population; it is this which gives the ϵ-Nash property.

B. NCE Stochastic Adaptive Control

In this section, each agent estimates self dynamical parameters; in other words, the analysis concerns with a family of agents $A_i,\ 1 \leq i \leq N$, whose control action at any instant is not permitted to be an explicit function of the system parameter θ_i. The dynamical parameter θ_i is estimated from the input output sample path $\{x_i(\tau), u_i(\tau); 0 \leq \tau \leq t\}$ of A_i; in other words, each agent A_i performs the estimation based upon observations of its own trajectory.

For definiteness in this work, the estimation algorithm chosen in the dynamical parameter estimation scheme is the RWLS algorithm, However, any estimation scheme which generates consistent estimates w.p.1 (subject to the given hypotheses) will also yield the system asymptotic equilibrium properties to be established.

1) Parameter Estimation: We denote the estimate of θ_i by $\hat{\theta}_{i,t} = [\hat{\mathbf{A}}_{i,t}, \hat{\mathbf{B}}_{i,t}]$, $t \geq 0$, $1 \leq N < \infty$, and assume $\hat{\theta}_{i,t}$ is generated at each $t \geq 0$ by the estimation algorithm. We adopt the notation $\theta^0 \triangleq \theta$ for the true parameters in the system. At time t, using $\hat{\theta}_{i,t}$ agent A_i solves the Riccati equation (3), obtains $\hat{\mathbf{\Pi}}_{i,t} \triangleq \mathbf{\Pi}(\hat{\theta}_{i,t})$ and solves the mass offset differential equation (4) to obtain $\hat{s}_i(t) \triangleq s(t; \hat{\theta}_{i,t})$. The certainty equivalence adaptive control for the admissible control set $\mathcal{U}_{l,i}$ is then given by $\hat{u}_i(t) \triangleq u(t; \hat{\theta}_{i,t})$, where

$$\hat{u}_i(t) = -\mathbf{R}^{-1}\hat{\mathbf{B}}_i^{\mathrm{T}}(\hat{\mathbf{\Pi}}_i x_i(t) + \hat{s}_i(t)),\ \ t \geq 0. \quad (10)$$

We observe that the solution of the control law is based on estimates of local parameters obtained from the agent's own trajectory and the distribution of the population dynamical parameters. To obtain the main (NCE)SAC result stated in Theorem 4.1, we first establish the strong consistency for the family of estimates $\{\hat{\theta}_t,\ t \geq 0\}$. In order to generate consistent estimates $\hat{\theta}$ w.p.1, a diminishing excitation is added to the adaptive control (10) given by,

$$\hat{u}_i(t) = -\mathbf{R}^{-1}\hat{\mathbf{B}}_i^{\mathrm{T}}(\hat{\mathbf{\Pi}}_i x_i(t) + \hat{s}_i(t)) + \xi_k\left[\epsilon_i(t) - \epsilon_i(k)\right], \quad t \in (k, k+1], \quad k \in \mathbb{N}, \quad 1 \leq i \leq N, \quad (11)$$

where $\{\xi_k^2 = \log k/\sqrt{k},\ \ k \geq 1\}$ and the process $(\epsilon(t), t \geq 0)$ is an $\mathbb{R}^m$-valued standard Wiener process that is independent of $\{w(t), t \geq 0\}$. The countable set of random processes $\{(\epsilon(t+k) - \epsilon(k), t \in (0,1]); k \in \mathbb{N}\}$ is assumed to be mutually independent and all members of the set have the same probability law on $(0, 1]$. Since the sequence $(\xi_k, k \in \mathbb{N})$ converges to zero at a suitable rate, it will be established following [25] that the diminishing control excitation $\{\xi_k[\epsilon(t) - \epsilon(k)], t \in [0,1); k \in \mathbb{N}\}$ provides sufficient excitation for almost sure consistent identification and decreases sufficiently rapidly enough not to affect the limiting performance of the system w.r.t. $\hat{\theta}_t = \theta^o, t \geq 0$, i.e. the non-adaptive case. In other words, the asymptotic performance achieved is equal to the one obtained in the non-adaptive case almost surely. The diminishing control excitation (11) was introduced in [29], and it was shown in [25] to generate strongly consistent RWLS parameter estimates for dynamical parameters of the system (1) under certainty equivalence adaptive control.

C. The NCE Stochastic Adaptive Control Law

In this section we present the NCE Stochastic Adaptive Control Law which generates the ϵ-Nash behaviour of the entire agent population. We observe that the solution of the

control law (11) has three terms based on local information and population distribution parameter and can be written for $1 \leq i \leq N;\ \ t \geq 0$, in the form of

$$u(\hat{\theta}_t) = u^{loc}(\hat{\theta}_t) + u^{pop}(\hat{\theta}_t) + u^{dit}(t),$$

$t \geq 0$, where $u^{loc}(\cdot)$ is the LQG feedback for the system of agent A_i; $u^{pop}(\cdot)$ is the mass offset term; and $u^{dit}(\cdot)$ is the locally generated dither input.

Simulateously for all $t \geq 0$ the solution to the following set of ODEs and algorithm equations generate the feedback control law $\hat{u}(t) \triangleq u(t; \hat{\theta}_t),\ \ t \geq 0$.

The continuous time NCE SAC control law for agent $A_i(\theta)$ with parameter $\theta \in \mathbf{\Theta}$, is summarized in 3 steps below:

NCE SAC Law

For $t \geq 0$:

(i) Solve the NCE Equations generating $x^*(t; F_\zeta)$.

(ii) Solve the RWLS equations:

$$\begin{aligned} \hat{\theta}_t^{\mathrm{T}} &= [\hat{\mathbf{A}}_t, \hat{\mathbf{B}}_t], \qquad \psi_t^{\mathrm{T}} = [x_t^{\mathrm{T}}, u_t^{\mathrm{T}}], \\ d\hat{\theta}_t &= a(t)\Psi_t\psi(t)[dx^{\mathrm{T}}(t) - \psi_t^{\mathrm{T}}\hat{\theta}_t dt], \\ d\Psi_t &= -a(t)\Psi_t\psi_t\psi_t^{\mathrm{T}}\Psi_t dt, \end{aligned} \tag{12}$$

and calculate $\hat{\theta}_t^{pr}$:

$$\hat{\theta}_t^{pr} = \arg\min_{\psi \in \theta} \|\hat{\theta}_t - \psi\|. \tag{13}$$

(iii) The NCE Control Law Equation at $\hat{\theta}_t^{pr}$:

a) $\hat{\mathbf{\Pi}}_t$: Solve the Riccati Equation at $\hat{\theta}_t^{pr}$:

$$\hat{\mathbf{A}}_t^{\mathrm{T}}\hat{\mathbf{\Pi}}_t + \hat{\mathbf{\Pi}}_t\hat{\mathbf{A}}_t - \hat{\mathbf{\Pi}}_t\hat{\mathbf{B}}_t\mathbf{R}^{-1}\hat{\mathbf{B}}_t^{\mathrm{T}}\hat{\mathbf{\Pi}}_t + \mathbf{Q} = 0. \tag{14}$$

b) $\hat{s}(t) \triangleq s(\hat{\theta}_t^{pr})$: Solve the mass offset differential equation at $\hat{\theta}_t^{pr}$:

$$\frac{d\hat{s}(\tau)}{d\tau} = (-\hat{\mathbf{A}}_t^{\mathrm{T}} + \hat{\mathbf{\Pi}}_t\hat{\mathbf{B}}_t\mathbf{R}^{-1}\hat{\mathbf{B}}_t^{\mathrm{T}})\times \hat{s}(\tau) + \mathbf{Q}x^*(\tau). \tag{15}$$

c) Obtain the control law from Certainty Equivalence Adaptive Control at $\hat{\theta}_t^{pr}$:

$$\hat{u}^{nce}(t) = -\mathbf{R}^{-1}\hat{\mathbf{B}}_t^{\mathrm{T}}\left(\hat{\mathbf{\Pi}}_t\hat{x}(t) + \hat{s}(t)\right) + \xi_k\left[\epsilon(t) - \epsilon(k)\right]. \tag{16}$$

The function $a(t), t \geq 0$, in (12) is in the form of $a(t) = 1/f(r(t))$, where $r(t) = \|\Psi_0^{-1}\| + \int_0^t |\psi(s)|^2 ds$, and $f \in \{f : \mathbb{R}_+ \to \mathbb{R}_+$, f is slowly increasing and $\int_c^\infty 1/(xf(x))dx < \infty;\ c \geq 0\}$. The function $f(.)$ is slowly increasing if it is increasing and satisfies $f(.) \geq 1$ and $f(x^2) = O(f(x))$ [25].

Note that solutions to (14) exists as $\hat{\theta}_t^{pr} \in \mathbf{\Theta} \subset \tilde{\mathbf{\Theta}}$, the set of controllable and observable dynamical parameters in $\mathbb{R}^{n(n+m)}$.

The main theorem in Section IV shows that the cost obtained through the NCE SAC Law is almost surely equal to the cost obtained with the non-adaptive control. Note that each agent starts with no prior information on its self-parameter, and the state aggregation integration in (6) is performed by use of the distribution $F_\zeta(\cdot)$.

The analysis is divided into two parts: In Section III we present the properties of the RWLS parameter estimation scheme used in the NCE SAC Law. In Section IV analysis of the behaviour of the LRA cost functions of each agent under the NCE SAC Law is presented. This section establishes the key convergence and Nash Equilibrium properties.

III. Convergence Properties of Parameter Estimates

In this section we show that the RWLS equations (12) with the projection method (13) provide strongly consistent, uniformly controllable and observable estimates of the individual dynamical parameters.

A. Asymptotic Convergence of RWLS Parameter Estimation

The RWLS algorithm is self-convergent [24], but there is no guarantee that the estimated values will be controllable. To ensure that the family of estimated models is uniformly controllable and observable we use the *projection method* of [22].

Specifically, the dynamic parameter estimate $\hat{\theta}_t \in \mathbb{R}^{n(n+m)}, t \geq 0$, is projected into the compact subset $\mathbf{\Theta}$ of the set of controllable and observable parameters $\tilde{\mathbf{\Theta}}$, for which the optimal control law generated by (14) will necessarily exist and be asymptotically stabilizing.

Lemma 3.1: (see Appendix I) Let $\mathbf{\Theta}$ be a compact set such that $\theta^0 \in \mathbf{\Theta} \subset \tilde{\mathbf{\Theta}} \subset \mathbb{R}^{n(n+m)}$. Let $\hat{\theta}_t$ be the estimate of $\theta^0 \in \mathbf{\Theta}$ obtained by the RWLS equations (12). Then, $\hat{\theta}_t^{pr} \triangleq \arg\min_{\psi \in \mathbf{\Theta}} \|\hat{\theta}_t - \psi\|$ (together with a co-ordinate ordering measurable tie breaking rule), satisfies $\hat{\theta}_t^{pr} \to \theta^0$ w.p.1 as $t \to \infty$.

■

Now, given the projection method lemma, we present the theorem that shows that the RWLS equations (12) generate strongly consistent estimates.

Theorem 3.1: Let $x_0^t \triangleq \{x(\tau),\ 0 \leq \tau \leq t\}$, and hypotheses **H**1, **H**2, **H**3 hold, and let $\{\hat{\theta}_t(x_0^t),\ t \geq 0\}$ be the process of estimates obtained by the RWLS equations (12) along the control trajectory $(x_t, \hat{u}_t^{nce}),\ \ 0 \leq t < \infty$, generated by the control $\hat{u}_t^{nce}$ in (16); and $\{\hat{\theta}_t^{pr}(x_0^t); t \geq 0\}$, be the projected estimates according to Lemma 3.1. Then,

(i) The input process given in (16) is well defined and is given by

$$\hat{u}^{nce}(t; \hat{\theta}_t^{pr}) = -\mathbf{R}^{-1}\hat{\mathbf{B}}_t^{\mathrm{T}}(\hat{\mathbf{\Pi}}_t x_t + s(t; \hat{\theta}_t^{pr}) + \xi_k\left[\epsilon(t) - \epsilon(k)\right], \tag{17}$$

(ii) $\hat{\theta}_t(x_0^t) \to \theta^0$ w.p.1, as $t \to \infty$.

■

The theorem is proved in detail in [30] using the methodology of [25], which established the convergence of the RWLS estimates (12) with diminishing excitation in the controls (16). The required uniform controllability and observability of the estimates is given here by Lemma 3.1.

B. Asymptotic Behaviour of the NCE Equations

In this section we present two results under the hypotheses of convergent individual dynamical parameters and the population distribution parameters. The convergence results for the mass offset function and the control law function are shown below.

Proposition 3.1: [30] Let **H**2-**H**4 hold and $\hat{\theta}_t \to \theta^0$ w.p.1 as $t \to \infty$. Let $s(t;\theta^0)$ be the solution to the mass offset function differential equation (4) and $s(t;\hat{\theta}_t)$ be the certainty equivalence function. Then,

$$s(t;\hat{\theta}_t) \to s(t;\theta^0) \text{ w.p.1 as } t \to \infty.$$

■

Recall that $\mathcal{U}^N_{nce} = \{u^0_i; 1 \le i \le N\}$ is the set of controls generated by the non-adaptive NCE Law, while $\hat{\mathcal{U}}^N_{nce} = \{\hat{u}_i; 1 \le i \le N\}$ is the set of controls generated by the NCE SAC Law (Sec:II-C).

Lemma 3.2: [30] For the system (1), under **H**1-**H**3, let $\hat{u}^{nce}_i \triangleq u^{nce}_i(t;\hat{\theta}_{i,t}) \in \hat{\mathcal{U}}^N_{nce}$, and $u^0_i \triangleq u^0_i(t;\theta^0_i) \in \mathcal{U}^N_{nce}$. Then, the following holds:

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T \|\hat{u}^{nce}_i - u^0_i\|^2 dt = 0 \text{ w.p.1,}$$
$$1 \le i \le N, \text{ as } N \to \infty. \quad (18)$$

■

C. Asymptotic (LRA) Behaviour of System Trajectories

In this section we show that under the hypothesis that the individual dynamical parameters converge to their true values, the trajectories of adaptive individual agents also converge to the trajectories obtained when the non-adaptive control law is applied.

Let $\hat{x}^{nce}_i \triangleq x_i(t;\hat{\theta}^{(0,t)}_i)$ be the state trajectory of agent A_i, $1 \le i \le N$, under the control law $u^{nce}_i(t;\hat{\theta}_{i,t}) \in \hat{\mathcal{U}}^N_{nce}$, and $x^0_i \triangleq x_i(t;\theta^o_i)$ be the state trajectory of agent A_i under the control law $u^0_i \triangleq u^0_i(t;\theta^0_i) \in \mathcal{U}^N_{nce}$.

Theorem 3.2: [30] Let $\hat{\theta}_{i,t} \to \theta^0_i$ w.p.1 as $t \to \infty$, $1 \le i \le N$. Then, for system (1), under **H**1-**H**3, the following holds:

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T \|\hat{x}^{nce}_i - x^0_i\|^2 dt = 0 \text{ w.p.1, } 1 \le i \le N.$$

■

IV. The Main Theorem

First, we show that the asymptotic cost function of an agent performing the NCE SAC Law (Sec:II-C) in a system of all adaptive agents is almost surely equal to the cost that would occur in a system of agents all performing the non-adaptive NCE Law. Then, we present the main result of the paper. The NCE Theorem (2.1) holds for the system of all adaptive agents when all agents apply the NCE SAC Law.

We set

$$J^\infty_i(u_i,u_{-i}) = \limsup_{T\to\infty}\frac{1}{T}\int_0^T \left\{\|x_i - x^*\|^2_Q + \|u_i\|^2_R\right\}$$
$$\text{w.p.1}, \quad 1 \le i \le N, \quad (19)$$

where x^* is given by (6) in the SAC $(\hat{u}_i)$ case, and by (6) in the non-SAC (u^0_i) case.

Lemma 4.1: [30] For the system (1), under **H**1-**H**4, **H**6, $u^0_i \in \mathcal{U}^N_{nce}$, $\hat{u}_i \in \hat{\mathcal{U}}^N_{nce}$, $1 \le i \le N$, the following holds:

$$\lim_{N\to\infty} J^N_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) = J^\infty_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) = J^\infty_i(u^0_i,u^0_{-i})$$
$$\text{w.p.1}, \quad 1 \le i \le N. \quad (20)$$

■

Lemma 4.2: [30] For the system (1), under **H**1-**H**4, **H**6, $u_i \in \mathcal{U}_{g,i}$, $u^0_i \in \mathcal{U}^N_{nce}$, $\hat{u}_i \in \hat{\mathcal{U}}^N_{nce}$, $1 \le i \le N$, the following holds:

$$\inf_{u_i\in\mathcal{U}_{g,i}} J^\infty_i(u_i,\hat{u}^{nce}_{-i}) = \inf_{u_i\in\mathcal{U}_{g,i}} J^\infty_i(u_i,u^0_{-i})$$
$$\text{w.p.1}, \quad 1 \le i \le N. \quad (21)$$

■

Theorem 4.1: NCE SAC Theorem (see Appendix II)
Let **H**1-**H**6 hold. Then, assume each agent A_i

(i) estimates its own parameter $\hat{\theta}_{i,t}$ via RWLS (12);
(ii) computes $u^{nce}_i(t;\hat{\theta}_{i,t})$ via the NCE equations plus dither.

Then,

(a) $\hat{\theta}_{i,t} \to \theta^0_i$ w.p.1 as $t \to \infty$, $1 \le i \le N$ (strong consistency);
The NCE SAC Law generates a set of controls $\hat{\mathcal{U}}^N_{nce} = \{\hat{u}_i;\ 1 \le i \le N\}$, $1 \le N < \infty$, such that:
(b) all agent systems $S(A_i)$, $1 \le i \le N$, are $LRA - L^2$ stable w.p.1;
(c) $\{\hat{\mathcal{U}}^N_{nce};\ 1 \le N < \infty\}$ yields an ϵ-Nash Equilibrium for all ϵ, i.e., $\forall\epsilon > 0$, $\exists N(\epsilon)$ s.t. $\forall N \ge N(\epsilon)$

$$J^N_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) - \epsilon \le \inf_{u_i\in\mathcal{U}_{g,i}} J^N_i(u_i,\hat{u}^{nce}_{-i}) \le$$
$$J^N_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) \text{ w.p.1, } 1 \le i \le N; \quad (22)$$

(d)

$$\lim_{N\to\infty} J^N_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) = J^\infty_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) =$$
$$J^\infty_i(u^0_i,u^0_{-i}) \text{ w.p.1, } 1 \le i \le N; \quad (23)$$

(e)

$$J^\infty_i(\hat{u}^{nce}_i,\hat{u}^{nce}_{-i}) = \inf_{u_i\in\mathcal{U}_{g,i}} J^\infty_i(u_i,\hat{u}^{nce}_{-i}) \text{ w.p.1,}$$
$$1 \le i \le N. \quad (24)$$

■

V. Simulation

Consider a system of 100 agents. The system matrices $\{A_k\},\{B_k\}$, $1 \le k \le 100$ are defined as

$$A \triangleq \begin{bmatrix} -0.2 + a_{11} & -2 + a_{12} \\ 1 + a_{21} & 0 + a_{22} \end{bmatrix} \quad B \triangleq \begin{bmatrix} 1 + b_1 \\ 0 + b_2 \end{bmatrix}.$$

The population dynamical parameter distribution a_{ij}'s and b_i's are mutually independent and distributed according to $a_{ij} \sim N(0, 0.2)$ and $b_i \sim N(0, 0.2)$. All agents apply the

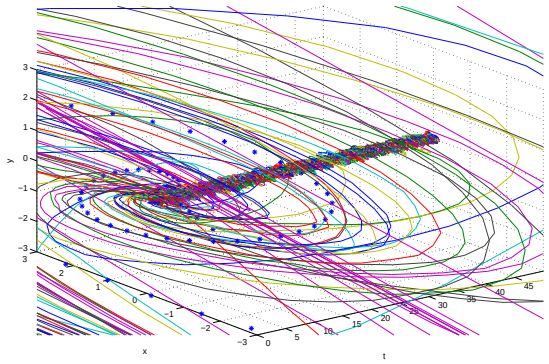

Fig. 1. State Trajectories

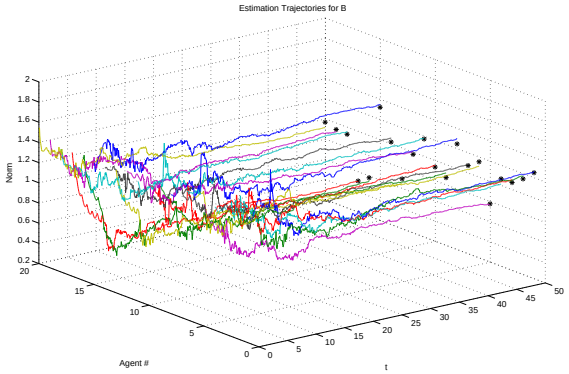

Fig. 2. Self Parameter Estimation - B

NCE SAC Law (Sec:II-C). Rapid convergence of the state trajectories to the steady state values can be seen in Fig. 1. In order to plot the convergence of the self estimation of dynamical parameter B, we plot the norm trajectories of the estimates in Fig. 2. The symbol '*' denotes the true value of the parameter for each agent. We only show 20 randomly chosen agents for visibility. In Fig. 3, the histogram of the norm of true values of B and the histogram of the norm of the estimated values at final evaluated instant are shown.

VI. Concluding Remarks

We study a stochastic adaptive optimal control problem where the costs of the agents in a population are coupled and each agent estimates its own dynamical parameters. The strong consistency of the self-parameter estimates, the stability of the system, and an ϵ-Nash Equilibrium property are established. Investigating the case where each agent in the system estimates population dynamical parameter distribution via observation on a random subset of agents in the system, general rates of convergence, relaxing the prior information on the cost function parameters are some of the possible future research directions.

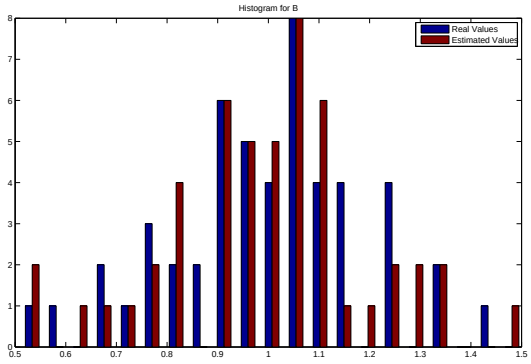

Fig. 3. True Parameter and Estimated Parameter Histograms - B

Appendix I
Proof of Lemma 3.1

Proof:

At each instant for the solution $\hat{\theta}_t$ to RWLS equations (12):

$$\hat{\theta}_t^{pr} \triangleq \arg\min_{\psi\in\mathbf{\Theta}} \|\hat{\theta}_t - \psi\|,$$

employing a co-ordinate ordering measurable tie breaking rule, if necessary.

Since $\hat{\theta}_t \in \mathbb{R}^{n(n+m)}$, $\hat{\theta}_t^{pr} \in \mathbf{\Theta}$ and $\theta^0 \in \mathbf{\Theta}$, the definition of $\hat{\theta}_t^{pr}$ gives

$$\|\hat{\theta}_t - \hat{\theta}_t^{pr}\| \leq \|\hat{\theta}_t - \theta^0\|.$$

But by Theorem 3.1, $\|\hat{\theta}_t - \theta^0\| \to 0$ w.p.1 as $t \to \infty$; therefore,

$$\hat{\theta}_t^{pr} \to \theta^0 \text{ w.p.1 as } t \to \infty.$$

■

Appendix II
Proof of Theorem 4.1

Proof: The sketch of the proof is below:

Theorem 3.1 implies (a), and Theorem 3.2 implies (b).

(c) Using a technique similar to the one used in [28, Theorem 6.2], we first show

$$J_i^N(\hat{u}_i^{nce}, \hat{u}_{-i}^{nce}) \leq J_i^\infty(u_i^0, u_{-i}^0) + O(\varepsilon_1(N)) \quad \text{w.p.1, } 1 \leq i \leq N. \quad (25)$$

Then we show

$$J_i^\infty(u_i^0, u_{-i}^0) \leq \inf_{u_i\in\mathcal{U}_{g,i}} J_i^N(u_i, \hat{u}_{-i}^{nce}) + O(\varepsilon_2(N)) + O(N^{-1}) \text{ w.p.1, } 1 \leq i \leq N, \quad (26)$$

where $\varepsilon_1(N) \to 0$ and $\varepsilon_2(N) \to 0$ as $N \to \infty$.

Applying (25) and (26) together, we get

$$J_i^N(\hat{u}_i^{nce}, \hat{u}_{-i}^{nce}) - \epsilon_N \leq \inf_{u_i\in\mathcal{U}_{g,i}} J_i^N(u_i, \hat{u}_{-i}^{nce}) \leq J_i^N(\hat{u}_i^{nce}, \hat{u}_{-i}^{nce}) \text{ w.p.1, } 1 \leq i \leq N, \quad (27)$$

where $\epsilon_N = O\left(\varepsilon_1(N) + \varepsilon_2(N) + N^{-1}\right)$.

Lemma 4.1 implies (d), and Lemma 4.2 implies (e).

■

REFERENCES

[1] A. C. Kizilkale and P. E. Caines, "Stochastic adaptive Nash certainty equivalence control: Population parameter distribution estimation," 2010, submitted to the 49th IEEE Conference on Decision and Control (CDC 2010).

[2] ——, "Stochastic adaptive nash certainty equivalence control with population dynamical and cost parameter estimation," 2010, submitted to the 19th Latin American Congress of Automatic Control (ACCA 2010).

[3] M. Huang, P. E. Caines, and R. P. Malhamé, "Large population cost-coupled LQG problems with non-uniform agents: Individual-mass behaviour and decentralized ϵ - Nash equilibria," *IEEE Trans. on Automatic Control*, vol. 52, no. 9, pp. 1560–1571, Sep 2007.

[4] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamic features of escape panic," *Nature*, vol. 407, pp. 487–490, September 2000.

[5] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Stable flocking of mobile agents, part i: Fixed topology," in *Proc. the 42nd IEEE Conference on Decision and Control*, Maui, Hawaii, 2003, pp. 2010–2015.

[6] Y. Liu and K. M. Passino, "Stable social foraging swarms in a noisy environment," *IEEE Trans. on Automatic Control*, vol. 49, pp. 30–44, Jan. 2004.

[7] D. J. Low, "Following the crowd," *Nature*, vol. 407, pp. 465–466, Sept. 2000.

[8] Y. C. Ho, A. E. B. Jr., and S. Baron, "Differential games and optimal pursuit-evasion strategies," *IEEE Trans. on Automatic Control*, vol. 10, pp. 385–389, 1965.

[9] P. P. Varaiya, "The existence of solutions to a differential game," *SIAM Journal on Control*, vol. 5, pp. 153–162, 1967.

[10] H. S. Witsenhausen, "Alternatives to the tree model for extensive games," in *The Theory and Applications of Differential Games*. The Netherlands: Reidel Publishing Company, 1975, pp. 77–84.

[11] A. Bensoussan, *Perturbation methods in optimal control*. New York: Wiley, 1988.

[12] R. Srikant and T. Basar, "Iterative computation of noncooperative equilibria in nonzero-sum differential games with weakly coupled players," *J. Optimization Theory Appl.*, vol. 71, no. 1, pp. 137–168, Oct. 1991.

[13] M. Huang, P. E. Caines, and R. P. Malhamé, "Individual and mass behaviour in large population stochastic wireless power control problems: Centralized and Nash equilibrum solutions," in *Proc. of the 42nd IEEE Conference on Decision and Control*, Maui, Hawaii, December 2003, pp. 98–103.

[14] ——, "The Nash certainty equivalence principle and McKean-Vlasov systems: an invariance principle and entry adaptation," in *Proc. of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, December 2007, pp. 121–126.

[15] M. Huang, R. P. Malhamé, and P. E. Caines, "Nash certainty equivalence in large population stochastic dynamic games: connection with the physics of interacting particle systems," in *Proc. of the 45th IEEE Conference on Decision and Control*, San Diego, CA, December 2006, pp. 4921–4926.

[16] ——, "Chapter 9 - Nash equilibria for large-population linear stochastic systems of weakly coupled agents," in *Analysis, Control and Optimization of Complex Dynamic Systems*, ser. GERAD 25th Annivesary Series, E. K. Boukas and R. P. Malhame, Eds. Springer, New York, 2005, pp. 215–252.

[17] P. E. Caines, *Mean Field Stochastic Control*. Bode Lecture at the 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Dec. 2009. [Online]. Available: http://www.cim.mcgill.ca/˜arman/ Shanghai2009/Shanghai2009.zip

[18] G. Y. Weintraub, C. L. Benkard, and B. V. Roy, "Oblivious equilibrum: A mean field approximation for large-scale dynamic games," in *Advances in Neural Information Processing Systems*. MIT Press, 2005.

[19] ——, "Markov perfect industry dynamics with many firms," *Econometrica*, vol. 76, no. 6, pp. 1375–1411, 2008.

[20] J. M. Lasry and P.-L. Lions, "Jeux á champ moyen. i - le cas stationnaire," *C. R. Acad. Sci. Paris, Ser. I*, vol. 343, pp. 619–625, 2006.

[21] ——, "Mean field games," *Japan J. Math.*, vol. 2, no. 1, pp. 229–260, 2007.

[22] P. E. Caines, "Continuous time stochastic adaptive control: non-explosion, ϵ-consistency and stability," *Syst. Contr. Lett.*, vol. 19, pp. 169–176, 1991.

[23] B. Bercu, "Weighted estimation and tracking for ARMAX models," *SIAM Journal on Control and Optimization*, vol. 33, pp. 89–106, 1995.

[24] L. Guo, "Self-convergence of weighted least-squares with applications to stochastic adaptive control," *IEEE Trans. on Automatic Control*, vol. 41, pp. 79–89, 1996.

[25] T. E. Duncan, L. Guo, and B. Pasik-Duncan, "Adaptive continuous-time Linear Quadratic Gaussian control," *IEEE Trans. on Automatic Control*, vol. 44, pp. 1653–1662, September 1999.

[26] 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," *Special issue in honour of the 65th birthday of Tyrone Duncan, Communications in Information and Systems*, vol. 6, pp. 221–252, November 2006.

[27] A. Bensoussan, *Stochastic Cotnrol of Partially Observable Systems*. U. K.: Cambridge Univ. Press, 1992.

[28] T. Li and J.-F. Zhang, "Asymptotically optimal decentralized control for large population stochastic multiagent systems," *IEEE Trans. on Automatic Control*, vol. 53, no. 7, pp. 1643–1660, August 2008.

[29] H.-F. Chen and L. Guo, "Optimal stochastic adaptive control with quadratic index," *International Journal of Control*, vol. 43, no. 3, pp. 869–881, 1986.

[30] A. C. Kizilkale and P. E. Caines, "Nash certainty equivalence stochastic adaptive control systems," Department of Electrical and Computer Engineering, McGill University, Tech. Rep., March 2010.