2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Optimal Decentralized Control of a Stochastically Switched System with Local Parameter Knowledge

Anshuman Mishra, Cédric Langbort, Geir E. Dullerud

Abstract— **This paper considers an optimal decentralized control problem for a linear system with stochastically switched input/output matrices depending on local parameters. These stochastic parameters are assumed independent in time and available instantaneously to the local controller but with a one time step delay to the other. We first solve this problem for the case of a single time step and show that the optimal policy can be reached through iterating the best responses of each player. For the optimal multiple step problem, a dynamic programming approach is employed while using the result of the one step control at each step.**

I. INTRODUCTION

In decentralized control, uncertainties entering the system are typically modeled as additive noise unknown to the agents. However in numerous settings, it seems practical to assume that, individual controllers have access to localized disturbances which are not relayed instantaneously to the other agents. In this paper we consider a scenario where controllers can react instantaneously to stochastic variations in local actuation and sensing, but can acquire knowledge about the global system after a small but non-zero delay. We borrow ideas from a closely related field called Team decision theory in order to tackle this problem.

Team decision theory was introduced more than half a century ago for application in Economic theory (see [1], [2]) and was subsequently adopted in Decentralized Control (for e.g. [3]–[6]). In [2], the author considered a class of problems with quadratic cost and under specific conditions the uniqueness of the solution was established. Further, a stationarity condition which also serves as the necessary condition of team optimality was provided. When the matrices associated with the quadratic cost function are independent of the random state of the system and the players' measurements are normally distributed, these necessary conditions admit a solution that is affine in the measurements. When applied to the static problem encountered in each step of decentralized Linear Quadratic Gaussian (LQG) problem, this result implies that linear strategies are optimal (see [3]–[5]). For the problem considered in this paper, however, the structure of optimal strategies may not be guessed so readily, and trying to directly solve the stationarity conditions could thus prove vexing. Motivated by [5], [7], we thus instead obtain the optimal solution by applying the best response mappings iteratively while making no assumption on the structure of the strategies. Similar approaches have been used in Game theory literature (see [8]) to compute Nash equilibria. The resulting strategies which solve the one step problem were found to be affine in the measurements with the stochastic parameters appearing multiplicatively.

For the finite horizon multistep problem, we assume a one step delayed sharing information pattern which implies that each player shares his knowledge of measurements and parameters with the other player after a delay of one time step. For this information structure, authors in [9] prove a separability result which says that the players can restrict their strategies at any time to functions of local information and the distribution of the most recent state conditioned on the common information. This allows us to use dynamic programming for the multistep problem. At each step of dynamic program, the minimization problem given by the Bellman equation matches with that of the one step problem. We can thus use the corresponding optimal policy and propagate the cost function through a backwards recursion. The information structure also implies that each player has access to all past measurements and parameters allowing it to use a centralized Kalman filter to obtain an estimate of the state. The optimal strategy obtained was found to be affine in the local measurements and estimated state with the corresponding coefficient matrices obtained through solving a set of linear equations.

II. NOTATIONAL PRELIMINARIES

For an indexed variable of the form θ_{it} (or $\theta_{i,t}$), the indices i and t would correspond to the player and time respectively. Since we consider only two player scenarios, a player index (subscript) of $-i$ would correspond to the other player and the subscript $-i$ can be identified with the integer $3-i$. The spectral radius of a matrix M is denoted by $\rho(M)$ and its induced 2-norm by $\|M\|$. We denote the set of symmetric positive positive definite and positive semi-definite matrices by $\mathbb{S}_+$ and $\bar{\mathbb{S}}_+$ respectively; the set of positive and nonnegative integers by $\mathbb{Z}_+$ and $\mathbb{Z}_0$ respectively. For a matrix A, $\mathrm{Tr}(A)$ represents its trace. As a shorthand notation we represent a block diagonal matrix by $\mathrm{diag}(D_1, \ldots, D_K)$ with D_i being its diagonal blocks. We represent a multivariate Gaussian random variable with mean μ and covariance matrix H as $\mathcal{N}(\mu, H)$. The corresponding joint density is defined by a Gaussian function parameterized by $\mu \in \mathbb{R}^d$ and $H \in \mathbb{S}_+$ as

$$\mathcal{G}_{\mu,H}(z) := \frac{1}{(2\pi)^{\frac{d}{2}}|H|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(z-\mu)^T H^{-1}(z-\mu)\right\}$$

The authors are with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801, USA (email: mishra10@illinois.edu).

The authors acknowledge support from AFOSR under grant FA9550-10-1-0573.

978-3-952-41734-8/©2013 EUCA

where $|\cdot|$ denotes the determinant. For a sequence of variables $\phi = \phi_1, \dots, \phi_r$, we will use the following notation for integration

$$\int_\phi F(\phi) := \int \dots \int F(\phi)\, d\phi_1 \dots d\phi_r$$

The domain of integration is typically the entire Euclidean space. We use similar notation for summation when ϕ_i takes discrete values.

III. TEAM THEORETIC BACKGROUND

In team decision theory, multiple players are faced with the problem of finding feedback strategies in order to minimize a common cost function. These strategies are functions of local information available to each player. For a two player problem, consider a cost function $J(\nu, u_1, u_2)$ where u_i is the action of i-th player and ν is a random state of the system. Let the information available to player i be denoted by $\mathbb{I}_i$, which depends on ν. The objective then is to find strategies $\gamma_i : \mathbb{I}_i \mapsto u_i$ which minimize the expected cost as

$$J^* = \min_{\gamma_1, \gamma_2} \mathbb{E}_\nu[J(\nu, \gamma_1(\mathbb{I}_1), \gamma_2(\mathbb{I}_2))]. \tag{1}$$

The minimizing solution (γ_1^*, γ_2^*) is called the *team optimal* strategy. The *best response* of a player is defined as a function of strategy $\gamma_{\text{-}i}$ of the other player as

$$\Gamma_i(\gamma_{\text{-}i}) \in \operatorname*{argmin}_{\gamma_i} \mathbb{E}[J(\nu, \gamma_1(\mathbb{I}_1), \gamma_2(\mathbb{I}_2))|\gamma_{\text{-}i}] \tag{2}$$

When $\gamma_{\text{-}i} = \gamma_{\text{-}i}^*$, the above best response yields the optimal strategy γ_i^*, i.e. $\Gamma_i(\gamma_{\text{-}i}^*) = \gamma_i^*$. When a tuple of strategies (γ_1, γ_2) satisfy

$$\gamma_i = \Gamma_i(\gamma_{\text{-}i}) \quad \text{for } i = 1, 2$$

they are called *person-by-person* optimal strategies. Note that while the optimal strategy is a person-by-person optimal, the converse may not hold in general. We can rewrite the best response (2), point-wise as

$$(\Gamma_i(\gamma_{\text{-}i}))(\mathbb{I}_i) \in \operatorname*{argmin}_{u_i} \mathbb{E}[J(\nu, u_i, \gamma_{\text{-}i}(\mathbb{I}_{-i}))|\mathbb{I}_i, \gamma_{\text{-}i}] \tag{3}$$

Radner in [2] considers the static team decision problem with quadratic cost function given by

$$J(\nu, u_1, u_2) = u^T \tilde{Q}(\nu) u + \delta^T(\nu) u + \beta(\nu) \tag{4}$$

Under convexity assumptions of the cost function, he showed that the person-by-person optimal strategy is unique (hence the team optimal) and stationarity condition leads to

$$\mathbb{E}[\tilde{Q}_{ii}|\mathbb{I}_i]\gamma_i(\mathbb{I}_i) + \mathbb{E}[\tilde{Q}_{i,\text{-}i}\gamma_{\text{-}i}(\mathbb{I}_{\text{-}i})|\mathbb{I}_i] + \mathbb{E}[\delta_i|\mathbb{I}_i] = 0 \quad \text{for } i = 1, 2 \tag{5}$$

Further when $\tilde{Q}$ is independent of ν, and the observations $\mathbb{I}_1$ and $\mathbb{I}_2$ have a joint normal distribution, the strategies γ_i are linear in $\mathbb{I}_i$. This result has been used in [3]–[5] to solve the decentralized LQG problem with one time step delayed sharing information pattern. In such situation, knowing that the optimal strategy is linear allows one to substitute this structure into (5) and solve for the corresponding unknown scaling matrices.

IV. PROBLEM STATEMENT

We consider a linear time varying system controlled by two players having the following dynamics

$$\begin{aligned} x_{t+1} &= A_t x_t + B_{1t}(\theta_{1t}) u_{1t} + B_{2t}(\theta_{2t}) u_{2t} + w_t \\ y_{it} &= C_{it}(\theta_{it}) x_t + v_{it}, \quad \text{for } i = 1, 2 \end{aligned} \tag{6}$$

Here $x_t \in \mathbb{R}^n$ is the state of the system at time t, $u_{it} \in \mathbb{R}^{m_i}$ and $y_{it} \in \mathbb{R}^{l_i}$ are respectively the control input and measurement of the i-th player. We define $y_t = \begin{bmatrix} y_{1t}^T \, y_{2t}^T \end{bmatrix}^T$, and similarly introduce u_t and v_t. The system matrices are functions of random variables $\theta_t = \{\theta_{1t}, \theta_{2t}\}$ which we call the *type* of the agents. We assume that θ_{it} takes value in a finite set Θ_i, it is i.i.d. in time and the types of the all agents are independent of each other. Further, each agent has complete knowledge of mappings B_{it} and C_{it} for all t, and the distributions of all the player types. The process noise (w_t) and players' measurement noise (v_{it}) are assumed to be i.i.d. Gaussian random variables as $w_t \sim \mathcal{N}(0, W)$ and $v_{it} \sim \mathcal{N}(0, V_i)$. Further the initial state is assumed to have a normal distribution, $x_0 \sim \mathcal{N}(\bar{x}_0, X_0)$. We assume that player i has the following information set at time t

$$\mathbb{I}_{it} = \mathbb{I}_{t-1}^c \cup \{y_{it}, \theta_{it}\} \tag{7}$$

where $\mathbb{I}_t^c = \{y_0, \dots, y_t, \theta_0, \dots, \theta_t, u_1, \dots, u_t\}$ is the common information. Let us denote the decentralized information structure as $\mathbb{I}_t^d = (\mathbb{I}_{1t}, \mathbb{I}_{2t})$.

A decentralized control strategy is the tuple (γ_1, γ_2) of sequences $\gamma_i = \{\gamma_{it}\}_{t=0}^{N-1}$ which map the information sets of individual players at each time to their control inputs as

$$u_{it} = \gamma_{it}(\mathbb{I}_{it}), \quad \text{for } i = 1, 2. \tag{8}$$

We will consider the following finite N-step horizon quadratic cost function

$$J\left(\{u_{1t}, u_{2t}\}_{t=0}^{N\text{-}1}\right) = x_N^T Q_N x_N + \sum_{t=0}^{N-1} \{x_t^T Q_t x_t + u_t^T R_t u_t\} \tag{9}$$

where R_t is structured as $\text{diag}(R_{1t}, R_{2t})$. We assume that $R_{it} \in \mathbb{S}_+$ and $Q_t \in \bar{\mathbb{S}}_+$ for all t.

The main objective of the decentralized control problem is to find a decentralized control strategy $(\gamma_{1t}^*, \gamma_{2t}^*)$ which solves the following minimization problem

$$\min_{\gamma_1, \gamma_2} \mathbb{E}\left[J\left(\{\gamma_{1t}(\mathbb{I}_{1t}), \gamma_{2t}(\mathbb{I}_{2t})\}_{t=0}^{N-1}\right)\right]$$

Here the expectation is taken over $\{\theta_{it}\}_{i=1,2;t=0,\dots,N-1}$, $\{w_t\}_{t=0,\dots,N-1}$, $\{v_{it}\}_{i=1,2;t=0,\dots,N-1}$ and x_0.

Having emphasized how B_{it} and C_{it} depend on the types θ_t, for convenience we will explicitly show this dependence only when required. We will often use the compact representation $B_t = \begin{bmatrix} B_{1t} \, B_{2t} \end{bmatrix}$, $C_t = \begin{bmatrix} C_{1t}^T \, C_{2t}^T \end{bmatrix}^T$ and $V = \text{diag}(V_1, V_2)$.

V. ONE STEP PROBLEM

A. Setup

We now consider the one time step ($N = 1$) version of the problem described in the previous section with two players. The dynamics evolve as

$$\begin{aligned} x_+ &= Ax + B_1(\theta_1)\, u_1 + B_2(\theta_2)\, u_2 + w \qquad (10) \\ y_i &= C_i(\theta_i)\, x + v_i, \text{ for } i = 1, 2 \end{aligned}$$

Here B_i and C_i depends on the local type θ_i as described earlier. The information available to player i is

$$\mathbb{I}_i = \{\theta_i, y_i\} \qquad (11)$$

Further the players have knowledge of the mean and covariance of $x \sim \mathcal{N}(\bar{x}, X)$. We consider a cost function

$$\begin{aligned} J(u_1, u_2) &= u_1^T R_1 u_1 + u_2^T R_2 u_2 + x_+^T Q x_+ \qquad (12) \\ &= \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}^T \begin{bmatrix} R_1 + B_1^T Q B_1 & B_1^T Q B_2 \\ B_2^T Q B_1 & R_2 + B_2^T Q B_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad (13) \\ &+ 2(Ax+w)^T Q \begin{bmatrix} B_1 \; B_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + (Ax+w)^T Q (Ax+w) \end{aligned}$$

In the above cost function, random variables x, θ_1, θ_2, v_1, v_2 and w play the role of ν as used in Section III.

The following lemma lists some useful definitions and properties making use of linear estimation theory

Lemma 1: For the dynamics in (10) and information structure (11), we have

(a) $\hat{x}_i := \mathbb{E}[x|\mathbb{I}_i] = \bar{x} + L_i(y_i - C_i\bar{x})$ and $\hat{X}_i := \mathbb{E}[(x - \hat{x}_i)(x - \hat{x}_i)^T|\mathbb{I}_i] = (I - L_iC_i)X$ where $L_i = XC_i^T(V_i + C_iXC_i^T)^{-1}$

(b) $\tilde{y}_{-i} = \mathbb{E}[y_{-i}|\mathbb{I}_i, \theta_{-i}] = C_{-i}(\theta_{-i})\hat{x}_i$

$$\mathbb{E}[(y_{-i} - \tilde{y}_{-i})(y_{-i} - \tilde{y}_{-i})^T|\mathbb{I}_i, \theta_{-i}] = V_{-i} + C_{-i}(\theta_{-i})\hat{X}_i C_{-i}^T(\theta_{-i})$$

(c) $\hat{e}_i := \hat{x}_i - \bar{x} = L_i(y_i - C_i\bar{x}) = L_iC_i(x - \bar{x}) + L_iv_i$, $\mathbb{E}[\hat{e}_i] = 0$, $\mathbb{E}[\hat{e}_i\hat{e}_i^T] = \mathbb{E}_{\theta_i}[L_i(C_iXC_i^T + V_i)L_i^T]$

(d) $\mathbb{E}[xx^T|\mathbb{I}_i] = \hat{X}_i + \hat{x}_i\hat{x}_i^T$, $\mathbb{E}[(x - \bar{x})(x - \bar{x})^T|\mathbb{I}_i] = \hat{X}_i + \hat{e}_i\hat{e}_i^T$

(e) For some $G_{-i}(\theta_{-i})$,
$\mathbb{E}[G_{-i}(\theta_{-i})\hat{e}_{-i}|\mathbb{I}_i] = \mathbb{E}_{\theta_{-i}}[G_{-i}(\theta_{-i})L_{-i}C_{-i}]\hat{e}_i$.

Note that L_i being a function of C_i, is also dependent on θ_i. However for simplicity, we choose to suppress this dependence. The above lemma uses standard properties (see for example [10]), and the proof is skipped.

B. Best Response Mappings

To compute a player's best response from (3), we first find the expected cost for player i conditioned on the other player's strategy as

$$\begin{aligned} J_i &:= \mathbb{E}[J(u_1, u_2)|\mathbb{I}_i, \gamma_{-i}] = u_i^T(R_i + B_i^T Q B_i)u_i \\ &+ \mathbb{E}\left[x^T A^T Q A x|\mathbb{I}_i\right] + 2u_i^T\{B_i^T Q \mathbb{E}[B_{-i}u_{-i} + Ax|\gamma_{-i}, \mathbb{I}_i]\} \\ &+ \mathbb{E}[2x^T A^T Q B_{-i}u_{-i} + u_{-i}^T(R_{-i} + B_{-i}^T Q B_{-i})u_{-i}|\gamma_{-i}, \mathbb{I}_i] \end{aligned}$$

Here we have used u_{-i} instead of $\gamma_{-i}(\mathbb{I}_{-i})$ for convenience. Noting that $R_i \succ 0$ ensures the convexity of the minimization problem in (3), we can set $\partial J_i/\partial u_i = 0$ to find the strategy

$$\begin{aligned} \big(\Gamma_i(\gamma_{-i})\big)(\mathbb{I}_i) &= -\mathbf{N}_i^Q \left\{\mathbb{E}[B_{-i}\gamma_{-i}(\mathbb{I}_{-i})|\gamma_{-i}, \mathbb{I}_i] + A\mathbb{E}[x|\mathbb{I}_i]\right\} \\ &= -\mathbf{N}_i^Q \left\{\mathbb{E}[B_{-i}\gamma_{-i}(\mathbb{I}_{-i})|\gamma_{-i}, \mathbb{I}_i] + A\,(\bar{x} + \hat{e}_i)\right\} \qquad (14) \end{aligned}$$

where $\mathbf{N}_i^Q(\theta_i) = -(R_i + B_i^T Q B_i)^{-1} B_i^T Q$.

We do not make any assumption on the structure of the strategies. However we restrict our search to the space described by the following assumption.

Assumption 2: The strategies γ_i satisfy

$$\sum_{\theta_i \in \Theta_i} \int_{y_i \in \mathbb{R}^{l_i}} \|B_i\gamma_i(\theta_i, y_i)\| \mathcal{G}_{\mu_i(\theta_i), \Sigma_i(\theta_i)}(y_i) < \infty \qquad (15)$$

where $\mu_i(\theta_i) = C_i(\theta_i)\bar{x}$ and $\Sigma_i(\theta_i) = V_i + C_i(\theta_i)XC_i^T(\theta_i)$. We can argue that the above assumption is not restrictive because the term on the left hand side is the expected value $\mathbb{E}[\|B_i\gamma_i(\mathbb{I}_i)\|]$ under no information availability and has to be finite.

C. Solution Through Iterated Best Response

In this section we will iteratively apply the best response mapping (14) and show that the resulting strategy converges to the optimal one. We define the mapping $\Psi_i^{Q,X} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ as

$$\Psi_i^{Q,X}(G) := \mathbb{E}[\,\mathbf{B}_i^Q(\theta_i)\, G\, \mathbf{C}_i^X(\theta_i)\,]$$

where $\mathbf{B}_i^Q(\theta_i) = B_i(R_i + B_i^T Q B_i)^{-1} B_i^T Q$ and $\mathbf{C}_i^X(\theta_i) = L_iC_i = XC_i^T(V_i + C_iXC_i^T)^{-1}C_i$. For this section, we will ignore superscripts Q and X as this is the only possibility. Let us also define $\bar{\mathbf{B}}_i = \mathbb{E}[\mathbf{B}_i(\theta_i)]$. We now present the main result of the section.

Theorem 3: For the system (10) and information set $\mathbb{I}_i$, the solution to the one step problem of minimizing the expected value of (12) is given by

$$\begin{bmatrix} \gamma_1^*(\mathbb{I}_1) \\ \gamma_2^*(\mathbb{I}_2) \end{bmatrix} = -\mathbf{N}^Q \left(\mathbf{M}_Q^{-1} \begin{bmatrix} I \\ I \end{bmatrix} A\bar{x} + SL(y - C\bar{x}) \right) \qquad (16)$$

where $\mathbf{N}^Q = \text{diag}\big(\mathbf{N}_1^Q, \mathbf{N}_2^Q\big)$, $\mathbf{M}_Q = \begin{bmatrix} I & \bar{\mathbf{B}}_2 \\ \bar{\mathbf{B}}_1 & I \end{bmatrix}$ and $L = \text{diag}(L_1, L_2)$. Here $S = \text{diag}(S_1, S_2)$ with $S_i \in \mathbb{R}^{n \times n}$ being the solution to the linear equations

$$S_i + \Psi_{-i}(S_{-i}) = A \text{ for } i = 1, 2 \qquad (17)$$

The resulting optimal expected cost is given by

$$\begin{aligned} J^* &= \text{Tr}(\mathbb{E}\big[L^T S^T (\mathbf{N}^Q)^T (R + B^T Q B)\mathbf{N}^Q S L\big] V) + \\ &\text{Tr}(\mathbb{E}\big[\mathbf{C}^T S^T (\mathbf{N}^Q)^T (R + B^T Q B)\mathbf{N}^Q S \mathbf{C}\big] X) + \text{Tr}(QW) \\ &-2\text{Tr}(A^T Q(\Psi_1(S_1) + \Psi_2(S_2))X) + \text{Tr}(A^T Q A X) + \\ &\bar{x}^T A^T (Q(I - \bar{\mathbf{B}}_1)(I - \bar{\mathbf{B}}_2\bar{\mathbf{B}}_1)^{-1}(I - \bar{\mathbf{B}}_2))A\bar{x} \end{aligned} \qquad (18)$$

□

The proof of the above theorem will be developed through a series of Lemmas. But first we continue with few more definitions. We introduce the mapping $\mathcal{E}_i$ acting on functions of the other player's information set defined point-wise as

$$(\mathcal{E}_i(q_{\text{-}i}))(\mathbb{I}_i) := \mathbb{E}[\mathbf{B}^Q_{\text{-}i} q_{\text{-}i}(\mathbb{I}_{\text{-}i})|\mathbb{I}_i]$$

Assuming that each player's strategy is a best response to some strategy (say $\tilde{\gamma}_{\text{-}i}$) of the other player, then (14) suggests that we can simply use $q_i(\mathbb{I}_i) := \mathbb{E}[B_{\text{-}i}\tilde{\gamma}_{\text{-}i}(\mathbb{I}_{\text{-}i})|\mathbb{I}_i]$ as a representation of player i's strategy such that

$$\gamma_i(\mathbb{I}_i) = -\mathbf{N}^Q_i \Big\{ q_i(\mathbb{I}_i) + A\left(\bar{x} + \hat{e}_i\right) \Big\} \tag{19}$$

Note that q_i also satisfies the following

$$\begin{aligned} \|q_i(\mathbb{I}_i)\| = \|\mathbb{E}[B_{\text{-}i}\tilde{\gamma}_{\text{-}i}(\mathbb{I}_{\text{-}i})|\mathbb{I}_i]\| \leq \mathbb{E}[\|B_{\text{-}i}\tilde{\gamma}_{\text{-}i}(\mathbb{I}_{\text{-}i})\||\mathbb{I}_i] \\ \Rightarrow \mathbb{E}[\|q_i(\mathbb{I}_i)\|] \leq \mathbb{E}[\|B_{\text{-}i}\tilde{\gamma}_{\text{-}i}(\mathbb{I}_{\text{-}i})\|] < \infty \end{aligned} \tag{20}$$

with the last inequality being a result of Assumption 2.

Using Lemma 1(b), we define the conditional density

$$g_i(y_{\text{-}i}|\theta_{\text{-}i}, \mathbb{I}_i) = \hat{\mathcal{G}}_{C_{\text{-}i}L_i,\, C_{\text{-}i}(I-L_iC_i)\bar{x},\, V_{\text{-}i}+C_{\text{-}i}\hat{X}_iC^T_{\text{-}i}}(y_{\text{-}i}, y_i). \tag{21}$$

This enables us to define the conditional distribution $f_{\text{-}i,i}(\mathbb{I}_{\text{-}i}|\mathbb{I}_i) := p_{\text{-}i}(\theta_{\text{-}i})g_i(y_{\text{-}i}|\theta_{\text{-}i}, \mathbb{I}_i)$, which describes the belief of the player $-i$'s information given player i's. We can thus evaluate the conditional expectation

$$\mathbb{E}[q_{\text{-}i}(\mathbb{I}_{\text{-}i})|\mathbb{I}_i] = \sum_{\theta_{\text{-}i}} \int_{y_{\text{-}i}} q_{\text{-}i}(\mathbb{I}_{\text{-}i}) f_{\text{-}i,i}(\mathbb{I}_{\text{-}i}|\mathbb{I}_i).$$

We further define

$$h_i(y^{(1)}_i|\theta^{(1)}_i, \theta_{\text{-}i}, \mathbb{I}_i) := \int_{y_{\text{-}i}} g_{\text{-}i}(y^{(1)}_i|\theta^{(1)}_i, \mathbb{I}_{\text{-}i}) g_i(y_{\text{-}i}|\theta_{\text{-}i}, \mathbb{I}_i)$$

which allows us to define $f_i(\theta^{(1)}_i, y^{(1)}_i|\theta_i, y_i) := \sum_{\theta_{\text{-}i}} p_i(\theta^{(1)}_i)p_{\text{-}i}(\theta_{\text{-}i})g_i(y^{(1)}_{\text{-}i}|\theta_{\text{-}i}, \mathbb{I}_i)$, the distribution of what the player i believes is player $-i$'s belief of player i's information, based on the actual information of player i. In the above definitions, in order to keep the notation compact, we do not make the distinction between a random variable and the value it takes. We can recursively define deeper beliefs of player i, with $\mathbb{I}^{(k)}_i = (\theta^{(k)}_i, y^{(k)}_i)$ and $\mathbb{I}^{(k-1)}_{\text{-}i}$ being the corresponding information variables at the k-th stage, which starts with $\mathbb{I}^{(0)}_i = \mathbb{I}_i$. We denote the corresponding sequence of player types as

$$\phi^{(k)}_i = (\theta^{(k-1)}_{\text{-}i}, \theta^{(k-1)}_i, \ldots, \theta^{(1)}_{\text{-}i}, \theta^{(1)}_i, \theta^{(0)}_{\text{-}i}). \tag{22}$$

With $h^{(1)}_i = h_i$, we recursively define

$$\begin{aligned} &h^{(k)}_i(y^{(k)}_i|\theta^{(k)}_i, \phi^{(k)}_i, \mathbb{I}_i) = \\ &\int_{y^{(k-1)}_i} h_i\big(y^{(k)}_i|\theta^{(k)}_i, \theta^{(k-1)}_{\text{-}i}, \mathbb{I}^{(k-1)}_i\big) h^{(k-1)}_i\big(y^{(k-1)}_i|\theta^{(k-1)}_i, \phi^{(k-1)}_i, \mathbb{I}_i\big) \end{aligned}$$

which is a Gaussian function whose exact expression is presented in Lemma 7 of Appendix. We can thus obtain the expression of k-level deep belief held by player i as

$$f_i(\mathbb{I}^{(k)}_i|\mathbb{I}_i) = \sum_{\phi^{(k)}_i} p_i(\theta^{(k)}_i)\hat{p}_i(\phi^{(k)}_i)h^{(k)}_i(y^{(k)}_i|\theta^{(k)}_i, \phi^{(k)}_i, \mathbb{I}_i)$$

where $\hat{p}_i(\phi^{(k)}_i) = p_{\text{-}i}(\theta^{(k-1)}_{\text{-}i})p_i(\theta^{(k-1)}_i)\ldots p_{\text{-}i}(\theta^{(0)}_{\text{-}i})$.

Lemma 4: For any representation of strategy q_i satisfying (20), we have

$$\lim_{k\to\infty} \|\big((\mathcal{E}_i\mathcal{E}_{\text{-}i})^k(q_i)\big)(\mathbb{I}_i)\| = 0.$$

Proof: We have

$$\begin{aligned} &\big((\mathcal{E}_i\mathcal{E}_{\text{-}i})(q_i)\big)(\mathbb{I}_i) = \mathbb{E}\Big[\mathbf{B}_{\text{-}i}(\theta^{(0)}_{\text{-}i})\mathbb{E}\Big[\mathbf{B}_i(\theta^{(1)}_i)q_i(\mathbb{I}^{(1)}_i)\Big|\mathbb{I}^{(0)}_{\text{-}i}\Big]\Big|\mathbb{I}_i\Big] \\ &= \sum_{\theta^{(1)}_i, \theta^{(0)}_{\text{-}i}} \int_{y^{(1)}_i} \mathbf{B}_{\text{-}i}(\theta^{(0)}_{\text{-}i})\mathbf{B}_i(\theta^{(1)}_i)q_i(\mathbb{I}^{(1)}_i)p_i(\theta^{(1)}_i) \\ &\qquad\qquad p_{\text{-}i}(\theta^{(0)}_{\text{-}i})p_i(\theta_i)h_i(y^{(1)}_i|\theta^{(1)}_i, \theta^{(0)}_{\text{-}i}, \mathbb{I}_i) \end{aligned}$$

Similarly $((\mathcal{E}_i\mathcal{E}_{\text{-}i})^k(q_i))(\mathbb{I}_i) =$

$$\sum_{\phi^{(k)}_i} \int_{y^{(k)}_i} \hat{\mathbf{B}}(\phi^{(k)}_i)q_i(\mathbb{I}^{(k)}_{\text{-}i})p_i(\theta^{(k)}_i)\hat{p}_i(\phi^{(k)}_i)h^{(k)}_i(y^{(k)}_i|\theta^{(k)}_i, \phi^{(k)}_i, \mathbb{I}_i)$$

where $\hat{\mathbf{B}}(\phi^{(k)}_i) := \mathbf{B}_{\text{-}i}(\theta^{(0)}_{\text{-}i})\ldots\mathbf{B}_i(\theta^{(k-1)}_i)\mathbf{B}_{\text{-}i}(\theta^{(k-1)}_{\text{-}i})$. Thus

$$\begin{aligned} \|((\mathcal{E}_i\mathcal{E}_{\text{-}i})^k(q_i))(\mathbb{I}_i)\| \leq \tilde{\alpha}\alpha^{2k-1} \max_{\phi^{(k)}} \sum_{\theta^{(k)}_i} \int_{y^{(k)}_i} \|q_i(\mathbb{I}^{(k)}_i)\| \\ p_i(\theta^{(k)}_i)h^{(k)}_i(y^{(k)}_i|\theta^{(k)}_i, \phi^{(k)}_i, \mathbb{I}_i) \end{aligned}$$

where $\tilde{\alpha}$ is a positive constant and α satisfies $\rho(\mathbf{B}_i(\hat{\theta}_i)) \leq \alpha < 1$ for all $\hat{\theta}_i \in \Theta_i$ and $i = 1, 2$. Due to the structure of $\mathbf{B}_i(\hat{\theta}_i)$ (described in [7], Equation (13)) it can be shown that there exist a constant $\tilde{\alpha}$ such that $\|\hat{\mathbf{B}}(\phi^{(k)}_i)\| \leq \tilde{\alpha}\alpha^{2k-1}$. We can also find a scalar $\delta > 0$ such that for each $\theta^{(k)}_i \in \Theta_i$

$$h^{(k)}_i(y^{(k)}_i|\theta^{(k)}_i, \phi^{(k)}_i, \mathbb{I}_i) \leq \delta\mathcal{G}_{\mu(\theta^{(k)}_i),\Sigma(\theta^{(k)}_i)}(y^{(k)}_i)$$

for all $y^{(k)}_i \in \mathbb{R}^{l_i}$ and $k \in \mathbb{Z}_+$ with $\mu_i(\theta^{(k)}_i) = C_i(\theta^{(k)}_i)\bar{x}$ and $\Sigma(\theta^{(k)}_i) = V_i - C_i(\theta^{(k)}_i)XC^T_i(\theta^{(k)}_i)$. To see why this is true, note that using Lemma 7 we know that any function $h^{(k)}_i$ above is a Gaussian with its variance (say $\tilde{\Sigma}(\theta^{(k)}_i)$) satisfying $V_i \preceq \tilde{\Sigma}(\theta^{(k)}_i) \preceq V_i + C_i(\theta^{(k)}_i)XC^T_i(\theta^{(k)}_i)$. The mean of corresponding functions lie in a bounded set, and hence the choice of $\mu(\theta^{(k)}_i)$ need not be the tightest in terms of the above inequality. Thus we have

$$\begin{aligned} &\|((\mathcal{E}_i\mathcal{E}_{\text{-}i})^{k-1}q_i)(\mathbb{I}_i)\| \leq \\ &\tilde{\alpha}\alpha^{2k-1}\delta \sum_{\theta^{(k)}_i} \int_{y^{(k)}_i} \|q_i(\mathbb{I}^{(k)}_i)\| p_i(\theta^{(k)}_i)\mathcal{G}_{\mu(\theta^{(k)}_i),\Sigma(\theta^{(k)}_i)}(y^{(k)}_i) \end{aligned}$$

Since the summation term on the right is just $\mathbb{E}[\|q_i(\mathbb{I}_i)\|]$ we know that it satisfies (20). Thus, setting $k \to \infty$ the right hand side converges to 0. ■

We now prove the main result of the section. For simplicity, in the following proof we will not make any distinction between $\mathbb{I}^{(j)}_i$ and $\mathbb{I}_i$.

Proof of Theorem 3: We can recursively apply the best response operations to obtain the following for all $k \in \mathbb{Z}_+$

$$\begin{aligned}((\Gamma_i\Gamma_{\text{-}i})^k(\gamma_i^0))(\mathbb{I}_i) = -\mathbf{N}_i^Q\{ \\ -(\mathcal{E}_i\mathcal{E}_{\text{-}i})^{k-1}\mathcal{E}_i(\mathbb{E}[B_i\gamma_i^0(\mathbb{I}_i)|\mathbb{I}_{-i}])(\mathbb{I}_i) && \text{(23a)} \\ +\sum_{j=0}^{k-1}(\bar{\mathbf{B}}_{\text{-}i}\bar{\mathbf{B}}_i)^j(I-\bar{\mathbf{B}}_{\text{-}i})A\bar{x} && \text{(23b)} \\ +\sum_{j=0}^{k-1}(\Psi_{\text{-}i}\Psi_i)^j(I-\Psi_{\text{-}i})(A)\hat{e}_i\} && \text{(23c)}\end{aligned}$$

Using Lemma 4 we know that setting $k \to \infty$ leads to (23a) converging to 0. Also since $\|\bar{\mathbf{B}}_i\| < 1$ and Ψ_i is contractive, (23b) converges to $(I-\bar{\mathbf{B}}_{\text{-}i}\bar{\mathbf{B}}_i)^{-1}(I-\bar{\mathbf{B}}_{\text{-}i})A\bar{x}$ and (23c) can be obtained by solving the (17) for S_i. Thus setting $k \to \infty$ results in (23) converging to

$$\gamma_i^*(\mathbb{I}_i) = \text{-}\mathbf{N}_i^Q\left\{(I\text{-}\bar{\mathbf{B}}_{\text{-}i}\bar{\mathbf{B}}_i)^{-1}(I\text{-}\bar{\mathbf{B}}_{\text{-}i})A\bar{x}\text{+}S_i\hat{e}_i\right\} \tag{24}$$

Since the best response iterations converge to a unique person-by-person equilibrium, this also establishes the global optimality of the strategies for the team problem. When written in a matrix form the above is the same as (16). In order to compute the expected cost, let us split the expression of optimal strategy as $u^* = \left[(\gamma_1^*(\mathbb{I}_1))^T\,(\gamma_2^*(\mathbb{I}_2))^T\right]^T = u^s + u^c$ which is explained in Remark 5. Plugging the above strategy into (13) (identified here with the quadratic form in (4)), we have

$$\begin{aligned}&\mathbb{E}[J(\gamma_1^*(\mathbb{I}_1),\gamma_2^*(\mathbb{I}_2))] = \mathbb{E}[(u^*)^T\tilde{Q}u^* + \delta^T u^* + \beta] \\ &= \mathbb{E}[(u^s)^T\tilde{Q}u^s+\delta^T u^s+\bar{x}^T A^T QA\bar{x}] + \mathbb{E}[(u^c)^T\tilde{Q}u^c+\delta^T u^c] \\ &\quad +2\mathbb{E}[(u^s)^T\tilde{Q}u^c] + \mathbb{E}[(x-\bar{x})^T A^T QA(x-\bar{x})] + \mathbb{E}[wQw^T] \\ &= \bar{x}^T A^T(Q(I-\bar{\mathbf{B}}_1)(I-\bar{\mathbf{B}}_2\bar{\mathbf{B}}_1)^{-1}(I-\bar{\mathbf{B}}_2))A\bar{x} \\ &\qquad + \left(\mathbb{E}[\hat{e}^T S^T(R+B^T QB)S\hat{e}] - 2\mathbb{E}[x^T A^T Q\mathbf{B}S\hat{e}]\right) \\ &\qquad +0 + \text{Tr}(A^T QAX) + \text{Tr}(QW)\end{aligned}$$

The first term above is the cost corresponding to the full state feedback and was derived in [7]. Other terms can be further expanded to obtain the expression in (18). ■

Remark 5: The optimal strategy (16) can be written as the following sum $\gamma_i(\mathbb{I}_i) = \gamma_i^s(\theta_i) + \gamma_i^c(\mathbb{I}_i)$ with $\begin{bmatrix}\gamma_1^s(\theta_1)\\ \gamma_2^s(\theta_2)\end{bmatrix} = -\mathbf{N}^Q\mathbf{M}_Q^{-1}\begin{bmatrix}I\\ I\end{bmatrix}A\bar{x}$ being the state feedback strategy as obtained in [7] applied to the expected value of the state and $\gamma_i^c(\mathbb{I}_i) = -(R_i + B_i^T QB_i)^{-1}B_i^T QS_iL_i(y_i - C_i\bar{x})$ is a corrective term based on local measurement.

VI. FINITE HORIZON MULTISTEP PROBLEM

In this section, we solve the multistep problem described in Section IV using a dynamic programming approach. For this, we denote the cost-to-go at time step t as $\mathcal{V}_t(\mathbb{I}_t^c)$. We will see in the next theorem that the optimal cost-to-go has a quadratic (plus constant) form in $\mathbb{E}[x_t]$ with the cost matrix (P_t) obtained through a backwards recursive equation. This allows us to use the one step result obtained in Theorem 3 to solve the minimization problem at each step of dynamic programming. Further the information structure (7) implies that the controllers have access to all measurements and types until the previous step. This allows each controller to use a centralized Kalman filter and have a common estimate of the state for the current time step. The covariance matrix (X_t) of the state corresponding to the Kalman filter is obtained through a forward propagating Riccati equation and depends on the past history of player types. The controller which is linear in the measurements uses coefficient matrices obtained by solving a set of linear equations involving both P_{t+1} and X_t.

Theorem 6: For the system described by (6) and information structure (7), the optimal control policy which minimizes the expected value of (9) is given by

$$\begin{bmatrix}\gamma_{1t}(\mathbb{I}_{1t})\\ \gamma_{2t}(\mathbb{I}_{2t})\end{bmatrix} = -\mathbf{N}^{P_{t+1}}\left(\mathbf{M}_{P_{t+1}}^{-1}\begin{bmatrix}I\\ I\end{bmatrix}A_t\bar{x}_t+S_tL_t(y_t-C_t\bar{x}_t)\right) \tag{25}$$

Here P_t is obtained by the backward recursion

$$\begin{aligned}P_t = Q_t + A_t^T(P_{t+1}(I - \bar{\mathbf{B}}_{1t}^{P_{t+1}}) \\ (I - \bar{\mathbf{B}}_{2t}^{P_{t+1}}\bar{\mathbf{B}}_{1t}^{P_{t+1}})^{-1}(I - \bar{\mathbf{B}}_{2t}^{P_{t+1}}))A_t\end{aligned} \tag{26}$$

with the terminal condition $P_N = Q_N$. The mean and covariance of the state is obtained by the following equations

$$\bar{x}_t = A_{t-1}\bar{x}_{t-1}+B_{t-1}u_{t-1}+A_{t-1}L_{t-1}^c(y_{t-1}-C_{t-1}\bar{x}_{t-1}) \tag{27}$$

$$X_{t+1} = A_t^T(X_t-X_tC_t^T(C_tX_tC_t^T+V)^{-1}C_tX_t)A_t+W \tag{28}$$

initialized with $\bar{x}_0 = x_0$ and X_0. Further

$$L_t = \text{diag}(L_{1t}, L_{2t}),\ L_{it} = X_tC_{it}^T(V_i + C_{it}X_tC_{it}^T)^{-1}$$

$L_t^c = X_tC_t^T(V + C_tX_tC_t^T)^{-1}$ and $S_t = \text{diag}(S_{1t}, S_{2t})$ where S_{it} is the solution to the linear equation

$$S_{it} + \Psi_{-i,t}^{P_{t+1},X_t}(S_{-i,t}) = A_t \text{ for } i = 1, 2 \tag{29}$$

The resulting optimal expected cost is given by

$$J^* = \text{Tr}(P_0X_0) + c_0 \tag{30}$$

where c_t is obtained as below starting with $c_N = 0$

$$\begin{aligned}c_t = \mathbb{E}_{\theta_t}[c_{t+1}]+\text{Tr}(P_{t+1}W)+\text{Tr}\left((A_t^T P_{t+1}A_t+Q_t-P_t)X_t\right) \\ +\text{Tr}(\mathbb{E}\left[L_t^T S_t^T(\mathbf{N}^{P_{t+1}})^T(R_t + B_t^T P_{t+1}B_t)\mathbf{N}^{P_{t+1}}S_tL_t\right]V)+ \\ \text{Tr}(\mathbb{E}\left[(\mathbf{C}_t^{X_t})^T S_t^T(\mathbf{N}^{P_{t+1}})^T(R_t+B_t^T P_{t+1}B_t)\mathbf{N}^{P_{t+1}}S_t\mathbf{C}_t^{X_t}\right]X_t) \\ -2\text{Tr}(A_t^T P_{t+1}\mathbb{E}\left[\mathbf{B}_t^{P_{t+1}}S_t\mathbf{C}_t^{X_t}\right]X_t)\end{aligned} \tag{31}$$

Proof: Due to the one step delayed information sharing, players have knowledge of all past system matrices and inputs. Thus the distribution of x_t conditioned on the common information is Gaussian and can be obtained through a centralized Kalman filter (after prediction but before update step). This yields the mean $\bar{x}_t$ as in (27) and the error covariance X_t through a forward Riccati equation as in (28). For a given set of strategies $u_{ik} = \gamma_{ik}(\mathbb{I}_{ik})$, the expected cost-to-go is defined as

$$\mathcal{V}_t(\mathbb{I}_t^c) = \mathbb{E}\left[\sum_{k=t}^{N-1}(x_k^T Q_kx_k + u_k^T R_ku_k) + x_N^T Q_Nx_N|\mathbb{I}_t^c\right].$$

We will show that the optimal expected cost-to-go has the form $\mathcal{V}_t^*(\mathbb{I}_t^c) = \mathbb{E}[x_t^T P_t x_t | \mathbb{I}_t^c] + c_t$ (or alternatively $\bar{x}_t^T P_t \bar{x}_t + \mathrm{Tr}(P_t X_t) + c_t$). We start with the terminal time $t = N$, the cost-to-go here is $\mathcal{V}_N^*(\mathbb{I}_N^c) = \mathbb{E}[x_N^T Q_N x_N | \mathbb{I}_N^c]$ which can be written as $\mathbb{E}[x_N^T P_N x_N] + c_N$ with $P_N = Q_N$ and $c_N = 0$. At time $t = N - 1$, we solve

$$\begin{aligned}
&\mathcal{V}_{N-1}^*(\mathbb{I}_{N-1}^c) = \mathbb{E}[x_{N-1}^T Q_{N-1} x_{N-1} | \mathbb{I}_{N-1}^c] + \\
&\quad \inf_{\gamma_{1,N-1}, \gamma_{2,N-1}} \mathbb{E}[u_{1,N-1}^T R_{1,N-1} u_{1,N-1} + u_{2,N-1}^T R_{2,N-1} u_{2,N-1} \\
&\qquad\qquad\qquad\qquad + x_N^T Q_N x_N | \mathbb{I}_{N-1}^c] \\
&= \mathbb{E}[x_{N-1}^T P_{N-1} x_{N-1} | \mathbb{I}_{N-1}^c] + c_{N-1}
\end{aligned}$$

where the minimization is solved by applying Theorem 3 resulting in strategies (25) with $t = N - 1$. The associated cost, also obtained from Theorem 3 gives us the last equality above with P_{N-1} and c_{N-1} obtained by substituting $t = N - 1$ in (26) and (31) respectively. This establishes the quadratic structure of the optimal cost-to-go for $t = N - 1$. We can follow the same procedure for each time t and write the Bellman equation for dynamic programming as

$$\begin{aligned}
&\mathcal{V}_t^*(\mathbb{I}_t^c) = \min_{\gamma_{1t}, \gamma_{2t}} \mathbb{E}\left[x_t^T Q_t x_t + u_t R_t u_t + \mathcal{V}_{t+1}^*(\mathbb{I}_{t+1}^c) | \mathbb{I}_t^c\right] \\
&= \mathbb{E}[x_t^T Q_t x_t + c_{t+1} | \mathbb{I}_t^c] + \min_{\gamma_{1t}, \gamma_{2t}} \mathbb{E}\left[u_t R_t u_t + x_{t+1}^T P_{t+1} x_{t+1} | \mathbb{I}_t^c\right] \\
&= \mathbb{E}[x_t^T P_t x_t | \mathbb{I}_t^c] + c_t.
\end{aligned}$$

As in traditional LQG theory, the use of a quadratic plus constant cost here for $\mathcal{V}_t^*(\mathbb{I}_t^c)$ can be proved by induction. Note that we start with such a form at $t = N$ and at every step of the backward recursion we recover the quadratic cost-to-go for the previous step. Continuing in this manner until $t = 0$, we obtain the optimal cost for the entire horizon. Note that the distribution of $\mathbb{I}_{t+1}^c$ conditioned on $\mathbb{I}_t^c$ does not depend on the past strategies [9], which is crucial for this dynamic program to work. ■

VII. CONCLUSIONS

This paper discusses a decentralized control problem with system matrices dependent on stochastic parameters available locally to each player. Under a one step delayed information sharing pattern we obtained optimal controllers which minimize a finite horizon quadratic cost function. The solution technique relies on ideas in team decision theory to construct an iterated best response to obtain strategies for the one step optimization problem.

For the multiple-step problem, the one step delayed information sharing pattern enables individual controllers to have access to all measurements and types until the previous step. This allows us to use a centralized Kalman filter at each controller and have a common estimate of the state for the current time step. With this estimate the solution to the one-step problem is directly applicable to minimize the cost-to-go at each stage of the dynamic programming.

Subsequent work on the problem considers a system where the state transition and the cost matrices also depend on the stochastic parameters. Further the stochastic parameters are allowed to be independent Markov chains. Updated results with refined proofs will appear in a journal version of the paper [11].

APPENDIX

The following lemma allows us to obtain an expression for the player's beliefs at each stage of the best response iterations.

Lemma 7: For a sequence $\phi_i^{(k)}$ as defined in (22), we have

$$h_i^{(k)}(y_i^{(k)} | \theta_i^{(k)}, \phi_i^{(k)}, \theta_i, y_i) = \mathcal{G}_{R_i^{(k)} y_i + r_i^{(k)}, \Xi_i^{(k)}}(y_i^{(k)})$$

$$\begin{aligned}
\text{where} \quad R_i^{(k)} &= C_i(\theta_i^{(k)}) \hat{\mathbf{C}}(\phi_i^{(k)}) L_i(\theta_i) \\
r_i^{(k)} &= C_i(\theta_i^{(k)}) \left(I - \hat{\mathbf{C}}(\phi_i^{(k)}) \mathbf{C}(\theta_i)\right) \bar{x} \\
\Xi^{(k)} &= V_i + C_i(\theta_i^{(k)}) X C_i^T(\theta_i^{(k)}) - \\
&\qquad R_i^{(k)} \left(V_i + C_i(\theta_i) X C_i^T(\theta_i)\right) (R_i^{(k)})^T \\
\hat{\mathbf{C}}(\phi_i^{(k)}) &= \mathbf{C}_{-i}(\theta_{-i}^{(k-1)}) \mathbf{C}_i(\theta_i^{(k-1)}) \ldots \mathbf{C}_{-i}(\theta_{-i}^{(0)})
\end{aligned}$$

As $k \to \infty$, we have $h_i^{(k)}(y_i^{(k)} | \theta_i^{(k)}, \phi_i^{(k)}, \theta_i, y_i)$ converges pointwise to $\mathcal{G}_{\mu_i(\theta_i^{(k)}), \Sigma_i(\theta_i^{(k)})}(y_i^{(k)})$ where $\mu_i(\theta_i) = C_i(\theta_i) \bar{x}$ and $\Sigma_i(\theta_i) = V_i + C_i(\theta_i) X C_i^T(\theta_i)$. □

The proof uses some properties of the Gaussian function and is skipped due the lack of space. The above lemma conveys that the function $h^{(k)}$ becomes less dependent on the measurement y_i and the choice of sequence $\phi_i^{(k)}$ as k increases. As $k \to \infty$, the distribution $f_i(\mathbb{I}_i^{(k)} | \mathbb{I}_i)$ converges pointwise to $p_i(\theta_i^{(k)}) \mathcal{G}_{\mu_i(\theta_i^{(k)}), \Sigma_i(\theta_i^{(k)})}(y_i^{(k)})$ which represents the distribution of the measurement under no available information.

REFERENCES

[1] J. Marschak, "Elements for a theory of teams," *Management Science*, vol. 1, no. 2, pp. 127–137, 1955.

[2] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.

[3] B. Kurtaran and R. Sivan, "Linear-quadratic-gaussian control with one-step-delay sharing pattern," *Automatic Control, IEEE Transactions on*, vol. 19, no. 5, pp. 571–574, 1974.

[4] J. Sandell, N. and M. Athans, "Solution of some nonclassical lqg stochastic decision problems," *Automatic Control, IEEE Transactions on*, vol. 19, no. 2, pp. 108 – 116, apr 1974.

[5] M. Toda and M. Aoki, "Second-guessing technique for stochastic linear regulator problems with delayed information sharing," *Automatic Control, IEEE Transactions on*, vol. 20, no. 2, pp. 260 – 262, apr 1975.

[6] Y. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[7] A. Mishra, C. Langbort, and G. E. Dullerud, "A team theoretic approach to decentralized control of systems with stochastic parameters," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 2116–2121.

[8] D. Fudenberg and D. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

[9] P. Varaiya and J. Walrand, "On delayed sharing patterns," *Automatic Control, IEEE Transactions on*, vol. 23, no. 3, pp. 443–445, 1978.

[10] B. Anderson and J. Moore, "Optimal filtering," *Prentice-Hall Information and System Sciences Series, Englewood Cliffs: Prentice-Hall*, vol. 1, 1979.

[11] A. Mishra, C. Langbort, and G. Dullerud, "Optimal decentralized control of a stochastically switched system with decentralized policies and local parameter knowledge," *Submitted*, 2013.