Asymptotic Solutions to Weakly Coupled Stochastic Teams with Nonclassical Information

R. Srikant and Tamer Başar, Fellow, IEEE

Abstract—In this paper, we develop a new iterative approach toward the solution of a class of two-agent dynamic stochastic teams with nonclassical information when the coupling between the agents is weak, either through the state dynamics or through the information channel. In each case, the weak coupling is characterized in terms of a small (perturbation) parameter. When this parameter value (say, $\epsilon$) is set equal to zero, the original fairly complex dynamic team, with a nonclassical information pattern, is decomposed into or converted to relatively simple stochastic control or team problems, the solution of which makes up the zeroth-order approximation (in a function space) to the team-optimal solution of the original problem. The fact that the zeroth-order solution approximates the optimal cost up to at least $O(\epsilon)$ is shown by upper and lower bounding the optimal cost, and then proving that the zeroth-order terms of these bounds are identical. Using this zeroth-order term as the starting point for a policy iteration, we show that approximations of all orders can be obtained by solving a sequence of stochastic control and/or simpler team problems.

I. INTRODUCTION

One of the challenging issues in stochastic control and team theory has been the derivation of optimal policies for problems that feature nonclassical information. Such patterns arise in stochastic control problems when not all useful measurement information is transmitted to future stages (see, for example, [1]), and they arise in stochastic teams when decision makers who are coupled through the systems dynamics and/or the common performance index do not share the same information. When this information is shared with a delay of one time unit, then the information pattern falls somewhat between classical and nonclassical, and is called "quasiclassical" for which, in the LQG framework, the optimal cost up to at least $O(\epsilon)$ is shown by upper and lower bounding the optimal cost, and then proving that the zeroth-order terms of these bounds are identical. Using this zeroth-order term as the starting point for a policy iteration, we show that approximations of all orders can be obtained by solving a sequence of stochastic control and/or simpler team problems.

When this information is shared with a delay of one time unit, then the information pattern falls somewhat between classical and nonclassical, and is called "quasiclassical" for which, in the LQG framework, the optimal cost up to at least $O(\epsilon)$ is shown by upper and lower bounding the optimal cost, and then proving that the zeroth-order terms of these bounds are identical. Using this zeroth-order term as the starting point for a policy iteration, we show that approximations of all orders can be obtained by solving a sequence of stochastic control and/or simpler team problems.

The above ideas grew out of preliminary research on the role of weak coupling in solving LQG team problems, which was presented in [8]. Earlier work on weakly coupled LQ stochastic teams [9] deals only with those problems which are characterized by a complete sharing of information between the DM’s. Hence, it relies on the solvability of the perturbed problem (the problem with $\epsilon \neq 0$) to obtain approximate solutions. In [10], the...
situation where the perturbed problem may not be solvable has been addressed, but the analysis there is only applicable to the cases where the DM's have access to perfect state information. This is so because the approximate solution has been derived using either the stochastic Pontryagin's principle or the stochastic Hamilton-Jacobi-Bellman equation, neither of which applies to the case when the information pattern is nonclassical.

The rest of the present paper is organized as follows. Section II deals with the problem formulation, where we formulate the two types of weakly coupled team problems that were mentioned above. In Section III, we obtain the zeroth-order approximation to the optimal cost using FDLC policies for each problem, and we also interpret the zeroth-order solution as the solution to the problem with \( \epsilon = 0 \). In Section IV, we show that in the class of FDLC's, better approximations to the optimal cost can be obtained through a policy iteration approach. Each step of the policy iteration involves the solution of a one-person LQG stochastic control problem, and hence the policies at each step can be computed easily, albeit with an increase in the order of the controllers. We also show in Section IV that, if we are interested in an approximation, the order of the estimator can be reduced, compared to what we would have obtained through the policy iteration. Section V provides concluding remarks. The paper also includes three appendices and a notation/acronym list which precedes the appendixes.

II. THE PROBLEM STATEMENT

We formulate two problems in this section, both of which are weakly coupled, but one is weakly coupled through the state equation, whereas the other one is weakly coupled through the information channel.

A. Problem P1

Consider the stochastic system defined by the pair of Itô differential equations

\[
\begin{align*}
\dot{x}_1(t) &= \begin{bmatrix} A_1(t) & 0 \\ 0 & A_2(t) \end{bmatrix} x_1(t) + \begin{bmatrix} B_1(t) & B_2(t) \end{bmatrix} u(t) + F(t) w(t) \\
\dot{x}_2(t) &= \begin{bmatrix} A_2(t) & 0 \\ 0 & A_1(t) \end{bmatrix} x_2(t) + \begin{bmatrix} B_2(t) & B_1(t) \end{bmatrix} u(t) + F(t) w(t)
\end{align*}
\]

(2.1)

where \( x_1(t) \) and \( x_2(t) \) are stochastic processes with continuous sample paths of dimensions \( n_1 \) and \( n_2 \), respectively. Here, \( x_0 = (x_{10}, x_{20})' \) is a Gaussian random vector with mean \( \bar{x}_0 \) and covariance \( \Sigma_0 \) given by

\[
\begin{bmatrix} \bar{x}_0 \\ \Sigma_0 \end{bmatrix} = \begin{bmatrix} 0 \\ \Sigma_{01} \end{bmatrix} \begin{bmatrix} \Sigma_{02} \end{bmatrix}
\]

(2.2)

and \( \{w_1(t), t \geq 0\} \) and \( \{w_2(t), t \geq 0\} \) are standard Wiener processes independent of each other and of \( x_0 \). The matrices \( A_i(t), B_i(t), A_j(t), \) and \( F(t), i, j = 1, 2, \) have appropriate dimensions, with entries continuous in \( t \in [t_0, t_f] \). The stochastic processes \( \{u_1(t), t \geq t_0\} \) and \( \{u_2(t), t \geq t_0\} \) are \( r_1 \) and \( r_2 \)-dimensional, respectively, denoting the controls of decision makers 1 and 2, respectively, and \( \epsilon \) is a small coupling parameter. The observation processes \( \{y_1(t)\} \) and \( \{y_2(t)\} \), which are \( m_1 \) and \( m_2 \)-dimensional, respectively, are defined by the equations

\[
\dot{y}_i(t) = C_i(t) x_i(t) dt + dV_i(t); \quad y_i(t_0) = 0, \quad i = 1, 2
\]

(2.3)

where \( \{v_i(t), t \geq t_0\}, \quad i = 1, 2, \) are independent standard Wiener processes which are also independent of \( x_0 \), \( \{w_1(t)\} \), and \( \{w_2(t)\} \). An admissible policy for DM \( i \), \( i = 1, 2 \), is a mapping \( \gamma_i: \mathbb{R} \times \mathcal{M}_m \to \mathbb{R}^m \), where \( \mathcal{M}_m \) is the space of all \( m_i \)-dimensional continuous functions on \( [t_0, t_f] \), such that \( u_i(t) = \gamma_i(t, y) \) is adapted to the family of sigma-fields generated by \( \{y \in \mathcal{M}_m, y_s \in \mathcal{M}_m; r_0 \leq s \leq r_1\} \) for all \( y \in \mathcal{M}_m \), \( t_0 \leq r_0 \leq r_1 \leq t \) where \( \mathcal{M}_m \) is a Borel set in \( \mathbb{R}^m \). In other words, the information available to DM \( i \) is \( I_i' \), where

\[
I_i' = \{y(s); 0 \leq s \leq t\}.
\]

Let the space of all admissible policies for DM \( i \) be denoted by \( \Gamma_i \).

The underlying problem is to find team-optimal strategies \( \gamma_i^*(t, I_i') \in \Gamma_i, \quad i = 1, 2 \), which minimize the cost functional

\[
J_i(u_1, u_2) = \frac{1}{2} \mathbb{E} \left[ \int_{t_0}^{t_f} \left( x_1^2(t) + x_2^2(t) + u_1^2(t) + u_2^2(t) \right) dt \right]
\]

(2.4)

with \( \{u_i^*(\cdot) = \gamma_i^*(\cdot, I_i')\}_{i = 1, 2} \) and \( Q_i \geq 0, \quad Q_{ij} \geq 0, \quad i, j = 1, 2 \).

B. Problem P2

For the second problem, we consider the stochastic system defined by the Itô differential equation

\[
\dot{x}(t) = \left\{ \begin{array}{ll}
A(x(t)) x(t) + B_1(t) u_1(t) + B_2(t) u_2(t) \\
+F(t) w(t)
\end{array} \right\}, \quad t \geq t_0, \quad x(t_0) = x_0
\]

(2.5)

where \( x(t) \) is a stochastic process with continuous sample paths of dimension \( n \). Here, \( x_0 \) is a Gaussian random vector with mean \( \bar{x}_0 \) and covariance \( \Sigma_0 \), and \( \{w(t), t \geq 0\} \) is a standard Wiener processes independent of \( x_0 \). The \( m \)-dimensional observation processes \( \{y_1(t)\} \) and \( \{y_2(t)\} \) are defined by the equations

\[
\dot{y}_i(t) = C(t) x(t) dt + dV_i(t) + \epsilon dV_i(t), \quad i = 1, 2
\]

(2.6)

where \( \{v_i(t), t \geq t_0\} \) are independent standard Wiener processes which are also independent of \( x_0 \). The information available to DM \( i \) is \( I_i' \), where \( I_i' \) is as defined before. An admissible policy \( \gamma_i \) for DM \( i \), \( i = 1, 2 \), and the corresponding strategy spaces \( \Gamma_i \) and \( \Gamma_2 \) are defined in a manner similar to problem P1. The cost
functional is given by
\[ J_i(u_1, u_2) = \frac{1}{2} E \left[ x'(t)Q_i x(t) + \int_{t_0}^{t'} (x'(t)Q(t)x(t) + u_1'(t)u_1(t) + u_2'(t)u_2(t)) dt \right]. \] (2.7)

The main difference between problems P1 and P2 is that, when \( \epsilon = 0 \), in P1 we have two independent stochastic control problems, one for each DM, whereas, in P2 we have a (strongly coupled) team problem where both players use the same information. We call the former "spatial weak coupling" since with \( \epsilon = 0 \) we have two subsystems independently controlled by the two DM's, and we call the latter "informational weak coupling" since with \( \epsilon = 0 \), exchange of measurements does not provide any new information to either DM. Although both these zeroth-order problems are solvable using the standard LQG theory, there are conceptual differences between the two. This will be discussed later in Section III-B. Notice that problems of the type above, when \( \epsilon = 0 \), in P1 we have two subsystems independently controlled by the two DM's, whereas, in P2 we correspond to a situation where a small noise term makes the problem, which otherwise could have been viewed as centralized control problem, a decentralized one.

Earlier attempts to obtain approximate solutions to problems of the type above, when the coupling between the DM's is not weak, can be found in [11] and [12]. In [11], the order of the estimators that are used by the DM's has been fixed, and this has resulted in a set of necessary conditions in the form of a matrix minimum principle. But this solution is not satisfactory because the chosen candidate solution (one that satisfies the conditions of the minimum principle) does not necessarily yield the global minimum to the problem. Also, the set of equations that describe the solution are very complicated to solve. In [12], it was assumed that the control values are exchanged but, in order, it was also assumed that the DM's do not attempt to infer the value of each other's measurement from the control values. This is an arbitrary (unrealistic) assumption, which was made to assure the solvability of the resulting problem, but it is unlikely that the solution of this modified problem is a good approximation for the solution of the original problem. In view of these difficulties encountered in [11] and [12], instead of studying the general class of nonclassical information pattern problems, we have formulated above a class of weakly coupled systems for which we show that the team problem admits approximate solutions. Specifically, we develop an approach that exploits the presence of weak coupling in the problem and the fact that the problems arrived at by setting \( \epsilon = 0 \) are completely solvable. This recursive approach leads to successively better approximate solutions to the two classes of teams with nonclassical information patterns, as formulated above.

III. THE ZEROOTH-ORDER SOLUTION

A. Problem P1

As we mentioned earlier, when \( \epsilon \) is set equal to zero we have two independent, standard LQG stochastic control problems, which admit unique solutions. Let us denote these solutions by \( \gamma^{(0)}(t, I^1) \) and \( \gamma^{(0)}(t, I^2) \), respectively, which are given by [13]
\[
\gamma^{(0)}(t, I^i) = -B_i(t)'P^{(0)}(t)\hat{x}(t), \quad i = 1, 2
\]

where \( \hat{x}(t) = \hat{x}_{i0} \)

\[
\dot{P}^{(0)}_i + A_i'P^{(0)}_i + P_i A_i - P_i B_i C_i' P^{(0)}_i + Q_i = 0;
\]

\[
\dot{\Sigma}^{(0)}_i = A_i \Sigma^{(0)}_i C_i + A_i' \Sigma^{(0)}_i C_i' + F_i F_i';
\]

Suppose we use the above set of policies in our original problem. Then the question is whether these policies approximate the optimal cost \( J^*_i \) up to \( O(\epsilon) \), where

\[ J^*_i = \inf_{\gamma_1, \gamma_2} J_i(\gamma_1, \gamma_2). \]

Toward answering this question, we first state the following lemma.

**Lemma 3.1:** For any pair of strategies \( \{ \gamma_1, \gamma_2 \} \in \Gamma_1 \times \Gamma_2 \)

\[ J^*_i := \inf_{\gamma_1, \gamma_2 \in \Gamma_1 \times \Gamma_2} J_i(\gamma_1, \gamma_2) \leq J^{0}_i \leq J_i(\gamma_1, \gamma_2) \] (3.2)

where \( \Gamma^*_i \) is the space of mappings \( \gamma_i : \mathbb{R} \times \mathbb{R}^{m_0+m_2} \rightarrow \mathbb{R}^{q_i} \) such that \( \gamma(t, \eta) \) is adapted to the family of sigma-fields generated by \( \{ \gamma \in \mathbb{R}^{m_0+m_2}, \gamma_i \in \mathbb{R}^{m_0+m_2}; r_0 \leq s \leq r_1 \} \) for all \( \eta \in \mathbb{R}^{m_0+m_2}, r_0 \leq s \leq r_1 \leq t \), i.e., \( \Gamma^*_i \) is the space of all policies for DMi which use the combined information \( I^i = \{ I^1_i, I^2_i \} \).

**Proof:** The left inequality follows from the definition of \( \Gamma^*_i, i = 1, 2 \), which leads to \( \Gamma_1 \times \Gamma_2 \subseteq \Gamma^*_1 \times \Gamma^*_2 \). The right inequality follows from the definition of infimum.

Since the computation of \( J^*_i \) involves an LQG control problem, it readily follows from the standard theory [13] that

\[ J^*_i = \int_{t_0}^{t'} \left[ \dot{\Sigma}_c P_c (\hat{B}_1 \hat{B}_1' + \hat{B}_2 \hat{B}_2') P_c \right] dt 
+ \int_{t_0}^{t'} \left[ P_c F F' \right] dt + \dot{\Sigma}_c P_c \Sigma_c + \text{tr} \left[ \Sigma_0 P_c (t_0) \right] \] (3.3)

where

\[ \dot{\Sigma}_c = \Sigma_c A' + A \Sigma_c + FF' - \Sigma_c (C_1 C_1' + C_2 C_2') \Sigma_c ; \]

\[ \Sigma(t_0) = \Sigma_0 ; \] (3.4)

\[ \dot{P}_c = A' P_c + P_c A - P_c (\hat{B}_1 \hat{B}_1' + \hat{B}_2 \hat{B}_2') P_c + Q = 0; \]

\[ P_c(t_0) = Q_f ; \] (3.5)

\[ C_1 := (C_1, 0); \quad C_2 := (0, C_2); \quad \hat{B}_1 := (B_1', 0)'; \]

\[ \hat{B}_2 := (0, B_2'); \]

\[ A := \begin{bmatrix} A_1 & \epsilon A_{12} \\ \epsilon A_{21} & A_2 \end{bmatrix}. \] (3.6)

Now, application of the implicit function theorem (IFT) for ordinary differential equations [14, Theorems 7.1 and 7.2] to
the two Riccati equations (3.4) and (3.5) leads to the following relationships:
\[ \Sigma_{e}(t) = \Sigma_{0}(t) + \epsilon R_{1}(t, \epsilon), \quad t \in [t_0, t_f] \]
\[ P_{e}(t) = P_{0}(t) + \epsilon R_{2}(t, \epsilon), \quad t \in [t_0, t_f] \]  
(3.7)

with
\[ \Sigma^{(0)} := \begin{bmatrix} \Sigma_{1}^{(0)} & 0 \\ 0 & \Sigma_{2}^{(0)} \end{bmatrix}, \quad P^{(0)} := \begin{bmatrix} P_{1}^{(0)} & 0 \\ 0 & P_{2}^{(0)} \end{bmatrix} \]  
(3.8)

which are valid for all \( \epsilon \in [-\epsilon_0, \epsilon_0] \), for some \( \epsilon_0 > 0 \). Here, the functions \( R_{i}(\cdot, \cdot), \quad i = 1, 2 \), are continuous in their arguments, and \( \lim_{\epsilon \to 0} \epsilon R_{i}(t, \epsilon) = 0 \). Note that \( P_{0}(t) \) is the solution of the Riccati equation obtained from (3.5) by setting \( \epsilon = 0 \). From (3.7), and the fact that the integrals in (3.3) are over a compact interval \([t_0, t_f]\), it follows, using the dominated convergence theorem [15], that
\[ J_{e}^{*} = J^{(0)} + O(\epsilon) \]  
(3.9)

where
\[ J^{(0)} = \frac{1}{2} \sum_{i=1,2} \left[ \int_{t_0}^{t_f} \left[ \Sigma_{i}^{(0)} P_{i}^{(0)} B_{i} B_{i}^{T} P_{i}^{(0)} \right] dt + \bar{x}_{i0} P_{i}^{(0)} \bar{x}_{i0} \right. \]
\[ \left. + \int_{t_0}^{t_f} \left[ P_{i}^{(0)} F_{i} F_{i}^{T} P_{i}^{(0)} \right] dt \right] + \text{tr}\left[ \Sigma_{0}^{(0)} P_{0}^{(0)} (t_0) \right]. \]  
(3.10)

We can also show (see Appendix A) that
\[ J_{i}(\gamma_{1}^{(0)}, \gamma_{2}^{(0)}) = J^{(0)} + O(\epsilon). \]  
(3.11)

This leads to the following theorem.

**Theorem 3.1:** The pair of strategies \( \{\gamma_{i}^{(0)}(t, I_{i}')\}_{i=1,2} \) yield a cost that is \( O(\epsilon) \) close to the optimal cost, i.e.,
\[ J_{e}^{*} = J(\gamma_{1}^{(0)}, \gamma_{2}^{(0)}) + O(\epsilon). \]

**Proof:** The result follows by applying Lemma 3.1 to the pair of strategies \( \{\gamma_{1}^{(0)}, \gamma_{2}^{(0)}\} \), and using relations (3.9) and (3.11).

**Remark 3.1:** Notice that, to prove Theorem 3.1, we have not only used the fact that the zeroth-order problem is solvable but also the fact that it is the solution of the complete information exchange problem in the limit as \( \epsilon \) goes to zero. Hence, for instance, the proof for Theorem 3.1 will not hold when the initial states \( x_{10} \) and \( x_{20} \) are correlated, even though the zero-th-order problem would still be solvable. We can relax this condition somewhat, by allowing a crosscorrelation between \( x_{10} \) and \( x_{20} \) of \( O(\epsilon) \).

**B. Problem P2**

Let us first study the zeroth-order problem. The measurement processes, \( \{y_{i}(t)\}, \quad i = 1, 2 \), with \( \epsilon = 0 \), are given by
\[ dy_{i} = C(t) x(t) + dv(t) \]  
(3.12)

which shows that they are identical. As in problem P1, with \( \epsilon = 0 \), it does not make any difference to the optimal cost here whether the DM's exchange information or not. While in problem P1 the reason was that the problems faced by the DM's were completely decoupled, in problem P2 this is due to the fact that \( y_{1}(t) = y_{2}(t) \), in view of which the sigma-field generated by \( I' \) is the same as the one generated by either \( I_{1}' \) or \( I_{2}' \). Hence, for the zeroth-order problem, without loss of generality, we can assume that the information available to both players is \( I_{1}' \).

Now, again from standard LQG control theory, the solution to the zeroth-order problem is given by
\[ u_{i0}(t) = y_{i0}(t, I') = -B_{i} P_{0}^{-1} \dot{x}(t) \]  
(3.13)

where
\[ dx = \left[ A - B_{1} B_{1}^{T} P_{0} - B_{2} B_{2}^{T} P_{0} \right] \dot{x}(t) dt \]
\[ + \Sigma^{(0)} C'(dy_{1} - C \dot{x}(t)) dt; \quad \dot{x}(t_0) = \bar{x}_{0} \]
\[ \Sigma^{(0)} = \Sigma^{(0)} A' + A \Sigma^{(0)} + FF' - \Sigma^{(0)} C' \Sigma^{(0)}; \quad \Sigma^{(0)}(t_0) = \Sigma_{0} \]
\[ \dot{P}_{0} + A' P_{0}^{T} + P_{0} A - P_{0} B_{1} B_{1}^{T} P_{0} \]
\[ + Q = 0; \quad P_{0}(t_f) = Q_f. \]  
(3.14)

But this is not the only possible solution to the zeroth-order problem with the expanded policy space. There are other representations of the above solution if we allow the information to each DM to be \( I' = (I_{1}', I_{2}') \), and all the representations will yield the same minimum cost. For instance, in (3.14), if we replace \( y_{1} \) by \( \alpha_{1} y_{1} + (1 - \alpha_{1}) y_{2} \), and use the resulting policy for DM1, \( i = 1, 2 \), the resulting filters substituted in (3.13) will still yield the same globally minimum cost for the zeroth-order problem. To decide as to which of these to choose as the zeroth-order solution, we cannot use the idea behind the choice of the zeroth-order solution of problem P1 which was that, it corresponded to the limit of the complete information exchange problem as \( \epsilon \to 0 \), because here it would lead (in the representation above) to the parametric values \( \alpha_{1} = \alpha_{2} = \frac{1}{2} \). But, the resulting pair of policies does not belong to \( \Gamma_{1} \times \Gamma_{2} \), since the information used by DM1 should be \( I_{1}' \), and not \( I' \). Hence, it cannot be a candidate zeroth-order solution. The only pair of policies that belongs to \( \Gamma_{1} \times \Gamma_{2} \), and yields the globally minimum cost to the zeroth-order problem is
\[ y_{i}^{(0)}(t, I') = -B_{i} S_{i}^{(0)}(t), \quad i = 1, 2 \]
\[ d\dot{x}^{(0)} = \left[ A - B_{1} B_{1}^{T} P_{0} - B_{2} B_{2}^{T} P_{0} \right] \dot{x}(t) dt \]
\[ + \Sigma^{(0)} C'(dy_{1} - C \dot{x}^{(0)}(t)) dt; \quad \dot{x}^{(0)}(t_0) = \bar{x}_{0}, \]  
(3.15)

In problem P1, the zeroth-order solution approximated the optimal cost to \( O(\epsilon) \), whereas in problem P2, the zeroth-order solution approximates the optimal solution to \( O(\epsilon^2) \). This can be shown in a manner similar to the proof of Theorem 3.1. We will not go through the whole proof here, but will simply indicate the reason as to why we get a better approximation in problem P2. It is clear that \( \epsilon \) is going to enter the
costs on the right- and left-hand side inequalities of Lemma 3.1 only through the error covariance, and not through the control gain, because the weak coupling parameter $\epsilon$ appears only in the information, and not in the state equation. Hence, we now look at the covariance Riccati equation under the complete information exchange case, which is given by

$$
\Sigma = A \Sigma + \Sigma A^T - \frac{2}{2 + \epsilon^2} \Sigma C \Sigma + FF^T; \quad \Sigma(t_0) = \Sigma_0.
$$

(3.16)

Since the Taylor series expansion of $2/(2 + \epsilon^2)$ around $\epsilon = 0$ does not contain odd powers of $\epsilon$, clearly the coefficient of the $\epsilon$ term in the expansion of $\Sigma(t)$ will be zero, and likewise in the optimal cost. Therefore, if we obtain a policy that achieves the zeroth-order term in the optimal cost, it approximates the optimal cost to $O(\epsilon^2)$. This now leads to the following theorem.

**Theorem 3.2:** The pair of strategies $\{y_1^{(0)}(t, I_1^i), y_2^{(0)}(t, I_2^i)\} = \{y_1(0), y_2(0)\}$, given by (3.15), yield a cost that is $O(\epsilon^2)$ close to the optimal cost, i.e.,

$$
J^* = J(y_1^{(0)}, y_2^{(0)}) + O(\epsilon^2).
$$

**Proof:** Similar to the proof of Theorem 3.1, as discussed above.

IV. APPROXIMATION IN THE CLASS OF FINITE-DIMENSIONAL LINEAR CONTROLLERS

In this section, we will show that a policy iteration can be used to obtain good approximations, in terms of orders of $\epsilon$, as we discussed above.

By a FDLC for DM$i$, we mean a controller of the form

$$
\eta_i(t, I_i^i) = L_i(t) z_i(t)
$$

dz_i = G_i(t) z_i(t) dt + H_i(t) dy_i(t); \quad z_i(t_0) = M_i \bar{x}_0,
$$

(4.1)

where $L_i(t)$, $G_i(t)$, $H_i(t)$, and $M_i$ are finite-dimensional matrices of compatible dimensions, with the first three having piecewise continuous entries. The class of all such policies for DM$i$ is denoted by $\Gamma_i^f$.

The policy iteration that we will use is of the Gauss-Seidel (G-S) type, defined as

$$
\eta_{i(k+1)}^{(0)} = \arg \min_{\gamma_i \in \Gamma_i^f} J(y_{1(k)}^{(0)}, y_{2(k)}^{(0)})
$$

$$
\eta_{2(k+1)}^{(0)} = \arg \min_{\gamma_2 \in \Gamma_2^f} J(y_{1(k+1)}^{(0)}, y_2^{(0)})
$$

$$
k = 0, 1, \cdots; \quad \eta_{2(0)} \in \Gamma_2^f \mbox{ specified.}
$$

(4.2)

This corresponds to the case when DM2 starts the iteration first. A G-S policy iteration starting with DM1 can be defined in an analogous manner. Notice that, even though the G-S policy iteration may not converge, the corresponding costs at each step of the iteration will converge because of the fact that we are generating a nonincreasing sequence, lower bounded by zero.

Now, we state the following two lemmas, which will take us to the main result of this section.

**Lemma 4.1:** Let $\gamma_1(t, I_1^i) \in \Gamma_1^f$ be arbitrarily fixed, and $\gamma_2(t, I_2^i; y_2) := \arg \min_{y_2} J(y_1(t, y_2))$. Then

$$
dy_2(t, I_2^i; y_2) := \gamma_2(0)(t, I_2^i) = O(\epsilon^2)
$$

regardless of the choice for $y_2 \in \Gamma_2^f$.

**Proof:** See Appendix B.

**Lemma 4.2:** Introduce two policies for DM2: $\tilde{\gamma}_2(t, I_2^i):= \tilde{L}(t) \bar{z}(t)$, and $\tilde{\gamma}_2(t, I_2^i) := \tilde{L}(t) \tilde{z}(t)$ where

$$
d\tilde{z}(t) = \tilde{G}(t) \tilde{z}(t) dt + \tilde{H}(t) dy_2(t); \quad \tilde{z}(t_0) = \bar{M}_0 \bar{x}_0
$$

$$
d\tilde{z}(t) = \tilde{G}(t) \tilde{z}(t) dt + \tilde{H}(t) dy_2(t); \quad \tilde{z}(t_0) = \bar{M}_0 \bar{x}_0
$$

and

$$
\tilde{G}_i := \begin{bmatrix} \tilde{G}_i^{(1)}(\epsilon) & G_i^{(0)}(\epsilon) \\ G_i^{(0)}(\epsilon) & G_i^{(2)}(\epsilon) \end{bmatrix},
$$

$$
\tilde{H}(\epsilon) := (\tilde{H}_1(\epsilon), H_2^{(2)}(\epsilon))';
$$

$$
\tilde{L}(\epsilon) := (L_1^{(2)}(\epsilon), L_2^{(0)}(\epsilon)); \quad \bar{M} = (M', M_0')
$$

(4.3)

where all the matrices are of arbitrary but compatible dimensions and the matrices with a superscript "k" are arbitrary functions of $\epsilon$ in the order of $O(\epsilon^k)$.

Further, let $\gamma_2^{(0)}(t_1^i) := \arg \min_{y_2} J(y_1(t_1^i), y_2)$. Then

$$
\gamma_2^{(0)}(t, I_2^i) = \gamma_2^{(0)}(t_1^i) + O(\epsilon^{k+1})
$$

and

$$
J(\gamma_1^*, \gamma_2^*) = J(\gamma_1^{(0)}, \gamma_2^*) + O(\epsilon^{k+1}).
$$

(4.5)

**Proof:** See Appendix C.

Clearly, the above lemmas remain true if we reverse the roles of DM1 and DM2. Before we present the main theorem of this section, we first introduce the following notation: let $\{\gamma_{1(k)}(t, I_1^i), \gamma_{2(k)}(t, I_2^i; y_2)\}$ be the pair of policies obtained after $k$-steps of the G-S policy iteration starting with an initial guess of $\gamma_1(t, I_1^i)$ for DM$i$, $i = 1, 2$. Then, the set of all possible pairs of policies after $k$-steps of the G-S policy iteration is given by

$$
\Gamma_k = \{ \{\gamma_{1(k)}, \gamma_{2(k)}\}; \gamma_{2(k)}(t, I_2^i; y_2) \} \subseteq \Gamma_2^f,
$$

$$
i = 1, 2, \cdots; \quad \gamma_{2(0)} \in \Gamma_2^f \mbox{ specified.}
$$

(4.6)

Now, we are in a position to state the following theorem.

**Theorem 4.1:** Suppose $\{\gamma_{1(k)}, \gamma_{2(k)}(t, I_2^i)\} \in \Gamma_k$. Then

$$
J(\gamma_{1(k)^*}, \gamma_{2(k)^*}) \leq J^{(0)} + O(\epsilon^{2k-1})
$$

where $J^{(0)} := \inf_{y_2 \in \Gamma_2^f} J(\gamma_1^{(0)}, \gamma_2)$. Also, if $\{\gamma_{1(k)}(t, I_1^i), \gamma_{2(k)}(t, I_2^i)\} \in \Gamma_k$ is another pair of strategies, then

$$
\gamma_{1(k)^*}(t, I_1^i) = \gamma_{1(k)}(t, I_1^i) + O(\epsilon^{2k-1}), \quad i = 1, 2,
$$

(4.7)

2 We say that $\gamma_{1(k)^*}(t, \epsilon) \in \Gamma_k$, $i = 1, 2$, is $O(\epsilon^2)$ if the $\ell^2$ norm of $u(\cdot; \epsilon) = \gamma_{1(k)^*}(\cdot, \epsilon)$ is $O(\epsilon^2)$ a.s.
Proof: The proof is by induction. Without loss of generality, let us assume that the G-S iteration is started with DM2, i.e., $\gamma_{200}(t, I')$ is specified. Then, by Lemma 4.1, we have that $\gamma_{110}(t, I') = \gamma_{101}(t, I') + O(\epsilon)$. We also have the property that, if DM2 chooses a different starting policy $\gamma_{20k}(t, I')$, then again $\gamma_{11k}(t, I') = \gamma_{10k}(t, I') + O(\epsilon)$.

Hence, $\gamma_{11k}(t, I') = \gamma_{10k}(t, I') + O(\epsilon)$. Also, from Lemma 4.2, we have $J(y_{11}, \gamma_{20}) = J(y_{10}, \gamma_{20}) + O(\epsilon)$. Noting that Theorem 3.1 holds even if we replace $J^*_{1k}$ by $J_{1k}$, we have that $J(y_{11}, \gamma_{20}) = J^*_{1k} + O(\epsilon)$. Therefore, $J^*_{1k} = J(y_{10}, \gamma_{20}) + O(\epsilon)$. This proves the theorem for $k = 0$.

Now, let us assume that the result is true for $k = 1, k \geq 1$.

Notice that, from the proof of Lemma 3.1 (see Appendix B), it is clear that the matrices in the differential equation representation of $\gamma_{11k}(t, I')$ and $\gamma_{10k}(t, I')$ differ by $O(\epsilon)$. Hence, the assumption that the result holds for $k = 1$ is equivalent to the assumption that $\gamma_{11k-1}(t, I')$ and $\gamma_{10k-1}(t, I')$ have differential equation representations whose matrices differ by $O(\epsilon^{2k-1})$. Therefore Lemma 4.2 applies and it follows that $\gamma_{11k}(t, I') = \gamma_{10k}(t, I') + O(\epsilon^{2k-1})$, $i = 1, 2$. This proves (4.7). To prove (4.6), we first note, using the definition of $J_{1k}$ (i.e., the fact that $J_{1k}$ is the minimum of $J(y_k, \gamma_{2k})$ over $\Gamma_k^I \times \Gamma_k^I$), that there exists a pair of strategies $\{y_{1k}(t, I'), \gamma_{2k}(t, I')\} \in \Gamma_k^I \times \Gamma_k^I$ such that

$$J_{1k}(\gamma_{1k}, \gamma_{2k}) < J_{1k}^{k*} + \epsilon |\epsilon|^{2k-1}.$$ 

(4.8)

Let us start a G-S policy iteration with an initial guess $\gamma_{2k}(t, I')$. After $k$ steps, let the strategies resulting from this policy iteration be $\{y_{1k}^{(k)}, \gamma_{2k}^{(k)}\}, \gamma_{2k}^{(k)} = \gamma_{2k}(t, I')$. From (4.7) (which was proved earlier), it follows that

$$\gamma_{11k}^{(k)}(t, I') = \gamma_{10k}^{(k)}(t, I') + O(\epsilon^{2k-1}), \quad i = 1, 2.$$ 

Hence, from Lemma 4.2, we have that

$$J_{1k}(\gamma_{1k}, \gamma_{2k}) = J_{1k}(\gamma_{10k}, \gamma_{20k}) + O(\epsilon^{2k-1})$$

(4.9)

Now, using (4.8), (4.9), and the fact that $J_{1k}^{k*} \leq J_{1k}(\gamma_{1k}, \gamma_{2k}) \leq J_{1k}(\gamma_{1k}, \gamma_{2k})$, where the right inequality follows from the fact that $\{y_{1k}^{(k)}, \gamma_{2k}^{(k)}\}$ is obtained using a G-S policy iteration starting with $\gamma_{2k}$, we have that $J_{1k}^{k*} = J_{1k}(\gamma_{1k}, \gamma_{2k}) + O(\epsilon^{2k-1})$, which proves (4.6).

The above theorem established the following important fact: irrespective of the initial guess for the G-S policy iteration, after $k$ steps, we obtain a cost which is $O(\epsilon^{2k-1})$ close to the optimal solution in the class of $O(\epsilon^2)$-approximation to $J_{1k}^{k*}$.

Theorem 3.1 tells us that after two steps of the policy iteration, we can reduce the order of the estimators and still obtain $O(\epsilon^2)$ approximation to $J_{1k}^{k*}$. In fact, we can find an $O(\epsilon^2)$ policy for each DM, but we have only an $(n_1 + n_2)$th-order estimator.

Let us start the G-S policy iteration with DM2 using the initial guess $\gamma_{10}^{(0)}(t, I')$. The problem faced by DM1 at the first stage of the iteration is $\min_{u_{21}} J_{1k}(y_{12}, \gamma_{20})$, with the policy $\gamma_{20}^{(0)}(t, I')$ substituted for $u_{21}(t)$ in the state equation. This results in the stochastic control problem

$$dx_1 = \left[ A_1 x_1 + \epsilon A_{12} x_2 + B_1 u_1 \right] dt + F_1 dw_1; \quad x_1(t_0) = x_{10}$$

$$dx_2 = \left[ A_2 x_2 + \epsilon A_{21} x_1 - B_2 B_2^0 x_2 \right] dt + F_2 dw_2; \quad x_2(t_0) = x_{20}$$

$$d\bar{x}_2 = \left[ \begin{array}{c} A_2 - B_2 B_2^0 \gamma_{20}^{(0)} - \gamma_{20}^{(0)} C_2 \bar{x}_2 \end{array} \right] dt + \gamma_{20}^{(0)} C_2 dw_2; \quad \bar{x}_2(t_0) = \bar{x}_{20}$$

$$J(y_{12}, \gamma_{20}) = \frac{1}{2} E \left[ \sum_{t_0}^{t_1} \left( \gamma_{12}(t_f) Q_{12} x_1(t_f) + \gamma_{12}^2(t_f) Q_{12} \bar{x}_2(t_f) \right) + \gamma_{12}(t_f) Q_{12} x_2(t_f) + \gamma_{12}^2(t_f) Q_{12} \bar{x}_2(t_f) + u_1(t) u_1(t) \right].$$

(4.10)

Using standard LQG theory, the solution to the above problem is given by

$$u_{12}^{(0)}(t) = \gamma_{12}^{(0)}(t, I') = -B_1 P_1^1 q_1 - B_1 P_1^2 q_2 - B_1 P_1^3 q_3$$

$$dq = \left( A_1 - B_1 B_1^1 - \gamma_{20}^{(0)} \right) qt + \gamma \epsilon C_{12} (dt - C_1 q_1); \quad q(t_0) = (\bar{x}_{10}, \bar{x}_{20}, \bar{x}_{20})$$

$$\hat{P}_1 + \gamma_{12}^2 \hat{P}_1 + \gamma_{12}^1 \hat{P}_1 - \gamma_{12}^1 \hat{P}_1 B_1^1 \hat{P}_1 + \gamma_{12}^1 = 0; \quad P_1(t_f) = Q_1^1$$

$$\hat{S}_1 = A_1 \hat{S}_1 + \gamma_{12}^2 \hat{S}_1 + \gamma_{12}^1 \hat{S}_1 - \gamma_{12}^1 \hat{S}_1 C_1^1 C_1^2 \hat{S}_1; \quad \hat{S}_1(t_0) = \text{block diag} \{\Sigma_{10}, \Sigma_{10}, 0\}$$

where

$$A_1 = \begin{bmatrix} A_1 & 0 & 0 \\ \epsilon A_{12} & A_2 & 0 \\ 0 & 0 & A_2 - B_2 B_2^0 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} (B_1^1, 0, 0) \\ (B_1^2, 0, 0) \end{bmatrix}$$

$$F_1 = \text{block diag} \{F_1, F_2, \Sigma_{20}^2 C_2 \};$$

$$Q_1 = \text{block diag} \{Q_{12}, Q_{12}, 0\}; \quad C_1 = \begin{bmatrix} C_1, \epsilon, 0 \end{bmatrix}$$

and $P_1(t)$ is the block of $P_1$. By applying the IFT to the Riccati equation for $P_1(t)$ in (4.11), it readily follows that $P_1(t)$ and $P_1(t)$ are $O(\epsilon)$. Hence, if we can show that $q_3 - q_2 = O(\epsilon)$ as a function of $y_{12}(t)$, then we can replace $q_3$ by $q_2$ in the expressions for $\gamma_{12}^{(0)}(t, I')$. This will lead to a new policy which will be different from $\gamma_{12}^{(0)}(t, I')$ by only $O(\epsilon^2)$, and will use an
estimator of order \((n_1 + n_2)\). Now, to show that \(\hat{x}_2 := q_2 - q_3 = O(\epsilon)\), we first note from (4.11) that \(\hat{x}_2\) is given by the following stochastic differential equation.

\[
\begin{align*}
\dot{x}_2 &= (\epsilon A_2 q_1 + A_2 \hat{x}_2 - \Sigma_{12} c_i C_i \hat{x}_2) dt \\
&+ (\Sigma_{12} - \Sigma_{11}) C_i \xi_i dt; \\
\hat{x}_2(t_0) &= 0.
\end{align*}
\]

(4.13)

Studying the above expression, it is clear that \(\hat{x}_2 = O(\epsilon)\) as a function of \(y_2(t)\), if \((\Sigma_{12} - \Sigma_{11}) = O(\epsilon)\). By applying the IFT to the Riccati equation for \(E'(t)\) in (4.11), we can easily show that \(\Sigma_{12}(t) = O(\epsilon)\), and hence \((\Sigma_{12} - \Sigma_{11}) = O(\epsilon)\).

An intuitive interpretation of the above result is as follows. The vectors \(q_2\) and \(q_3\) are nothing but the conditional expectations \(E(x_2, I)\) and \(E(x_2, I)\), respectively. As \(\epsilon \to 0\), \(x_2\) and \(I_1\) become independent of \(I_1\); hence, \(E(x_2, I)\) becomes \(E(x_2)\), and \(E(x_2, I)\) becomes \(E(E(x_2, I))\), which, by the smoothing property of the conditional expectation, is again \(E(x_2)\). Before we summarize the above results formally (see Theorem 4.2 later) we note that if \(q_3\) is replaced by \(q_1\) in (4.11), then the following policy is obtained for DM1:

\[
\gamma^{e_1}(t, I_1) = -B_1'[P_{11}^{e_1}, \hat{x}^{e_1} + (P_{12}^{e_1} + P_{13}^{e_1}) \hat{x}^{e_1}]
\]

\[
\left(\begin{array}{c}
\dot{x}^{e_1} \\
\dot{x}^{e_2}
\end{array}\right) = \left(\begin{array}{cc}
A_1 & -B_1[P_{11}^{e_1} + (P_{12}^{e_1} + P_{13}^{e_1})]
\epsilon A_2 & A_2
\end{array}\right)
\left(\begin{array}{c}
\hat{x}^{e_1} \\
\hat{x}^{e_2}
\end{array}\right) dt
\]

\[
\left(\begin{array}{c}
\hat{x}^{e_1}(t_0) \\
\hat{x}^{e_2}(t_0)
\end{array}\right) = \left(\begin{array}{c}
\hat{x}_{10} \\
\hat{x}_{20}
\end{array}\right). 
\]

(4.14)

Similarly, by starting with \(\gamma^{e_0}(t, I_1)\), we could have obtained a similar expression for the first-order controller for DM2, which is given by

\[
\gamma^{e_2}(t, I_2) = -B_2'[P_{22}^{e_2}, \hat{x}^{e_2} + (P_{23}^{e_2} + P_{24}^{e_2}) \hat{x}^{e_2}]
\]

\[
\left(\begin{array}{c}
\dot{x}^{e_2} \\
\dot{x}^{e_2}
\end{array}\right) = \left(\begin{array}{cc}
A_1 & -B_2[P_{22}^{e_2} + (P_{23}^{e_2} + P_{24}^{e_2})]
\epsilon A_2 & A_2
\end{array}\right)
\left(\begin{array}{c}
\hat{x}^{e_2} \\
\hat{x}^{e_2}
\end{array}\right) dt
\]

\[
\left(\begin{array}{c}
\hat{x}^{e_2}(t_0) \\
\hat{x}^{e_2}(t_0)
\end{array}\right) = \left(\begin{array}{c}
\hat{x}_{10} \\
\hat{x}_{20}
\end{array}\right). 
\]

(4.15)

The result follows from the discussion preceding the theorem, and from Lemma 4.2 and Theorem 4.1. 

Note that to compute the pair of strategies \(\{\gamma^{e_1}(t, I_1), \gamma^{e_2}(t, I_2)\}\), one needs to compute the solution of two higher order Riccati equations to obtain \(P^{e_1}(t)\) and \(P^{e_2}(t)\). But, we do not really have to do this, if we are interested in an \(O(\epsilon^2)\) approximation to the optimal cost. Instead, by noting that the IFT applies to these Riccati equations, we can approximate \(P^{e_1}(t), i = 1, 2\), as

\[
P^{e_i}(t) = P^{e_0(t)} + \epsilon P^{e_1(t)} + O(\epsilon^2), \\
i = 1, 2
\]

(4.18)

where

\[
\begin{align*}
P^{e_0(t)} &= \text{block diag}\{P_{11}^{e_0(t)}, 0, 0\}; \\
P^{e_2(t)} &= \text{block diag}\{0, P_{22}^{e_2(t)}, 0\}
\end{align*}
\]

(4.19)

and \(P^{e_1(t)}, i = 1, 2\), is the solution of the following linear differential equation:

\[
\dot{P}^{e_i} + A^{e_0} P^{e_i} + A^{e_0} P^{e_i} A^{e_0} + P^{e_0} A^{e_0} - P^{e_0} B^2 B^2 P^{e_0} = 0.
\]

(4.20)

Hence, if we need an \(O(\epsilon^2)\) approximation to the optimal solution in the class of FDLC's, we just need to solve two lower order Riccati equations and two linear differential equations.
equations. A similar feature was observed earlier in [16], in the context of LQ deterministic optimal control problems.

Now, we state the following theorem for problem P2 which is the counterpart of Theorem 4.1.

**Theorem 4.3:** Suppose \( \{ \gamma_{1,k}(t, l'_1), \gamma_{2,k}(t, l'_2) \} \in \Gamma_{(k)} \), then

\[
J^*_k(\gamma^*_1(t), \gamma^*_2(t)) = J^*_k + O(e^{4k-2}) \quad (4.21)
\]

where \( J^*_k = \inf_{\gamma_1, \gamma_2 \in \Gamma_{(k)}} \int J(\gamma_1, \gamma_2) \). Also, if \( \{ \gamma_{1,k}(t, l'_1), \gamma_{2,k}(t, l'_2) \} \in \Gamma_{(k)} \) is another pair of strategies, then

\[
\gamma^*_k(t, l'_1) = \gamma^*_k(t, l'_2) = O(e^{4k-2}), \quad i = 1, 2. \quad (4.22)
\]

**Proof:** Similar to the proof of Theorem 4.1. □

It should be noted that, in problem P2, the expansion of the cost does not contain odd powers of \( e \) because of the fact that the power series expansion of the error covariance does not contain odd powers of \( e \) as explained in Section 3.3. Also note that in contrast to the result of Theorem 4.2, here the dimension of each of the estimators at the end of two iterations is larger than the dimension of the plant.

**V. CONCLUSIONS**

In this paper, we have studied continuous-time stochastic team problems, with two DM's who are weakly coupled through a small parameter \( e \). The systems are such that if \( e \) is set equal to zero, the team problems either decompose into two independent stochastic control problems with a classical performance index, or are converted into a tractable team problem. Using this as the starting point, we iteratively obtained strategies for the two DM's, which are (depending upon the type of weak coupling between the DM's) either \( O(e^{2k-1}) \) or \( O(e^{4k-2}) \) close to the optimal solution after the \( k \)th step of iteration. A major contribution of the paper is that, through this decomposition, the original problem, which features nonclassical information, and for which there is no generally available theory for obtaining the optimal solution, is converted into a sequence of more tractable problems. We also believe that the proof of the fact that the solution of the zeroth-order problem indeed provides the zeroth-order approximation to the cost is the first of its kind for stochastic control/team problems featuring nonclassical information.

We have also pointed out an interesting difference between problems that are weakly coupled through the state equation and those that are weakly coupled through the information. While in the former case the zeroth-order solution provides an \( O(e) \) approximation to the optimal cost, in the latter case we have an \( O(e^2) \) approximation to the optimal cost. This points out a quantitative difference in the significance of \( e \) in the two problems.

Here, we have dealt with systems in which the DM's are weakly coupled either through the state equations or through the information channel, but the techniques outlined in the paper apply even when the weak coupling is through the performance index, or is in the covariance of the initial state as was mentioned in Remark 3.1. Further, these techniques can be extended to situations where there are more than two DM's, some of which are weakly coupled, and others net, with the additional specification that the DM's who are not weakly coupled exchange information in such a way that the limiting problem arrived at by setting \( e = 0 \) is tractable. In such a case, we again expect that a proof using Lemma 3.1 can be used to show that the zeroth-order solution provides a cost that is \( O(e) \) close to the optimal solution. In extending the policy iteration result to the multiple DM case, it should be noted that to obtain successively better approximations after each step of the iteration, we have to ensure that each player has acted at least once during each step of the G-S iteration, although the order in which they act need not be fixed.

Problems featuring nonclassical information patterns arise not only in stochastic teams, but also in stochastic zero-sum and nonzero-sum games. For the latter class, however, since the players do not minimize a common performance index, one cannot immediately extend the results of this paper to obtain noncooperative equilibrium solutions in these situations. More tractable classes of problems are the counterparts of the above in the discrete-time, since then it is possible to write down more explicit necessary conditions for the existence of an optimal solution. These extensions will be discussed in a future publication.

**ACRONYMS AND NOTATIONS**

- **DM** Decision maker.
- **FDLC** Finite-dimensional linear controller.
- **G-S** Gauss–Seidel.
- **IFT** Implicit function theorem.
- **LQG** Linear quadratic Gaussian.
- **I_i** Information available to Player i at time \( t \).
- **P_i** Space of all admissible policies for Player i under a given information pattern.
- **P_i'** Space of all admissible policies for Player i under a complete sharing of information.
- **\Gamma_i** Space of all pairs of policies that can be obtained after \( k \)-steps of the G-S policy iteration.
- **\Gamma_i(k)** Space of all pairs of policies that can be obtained after \( k \)-steps of the G-S policy iteration.
- **\gamma_{i,k}(t, l'_1, l'_2)** The optimal policy for Player i after \( k \)-steps of the G-S iteration.
- **\gamma_{i,0}(t)** The zeroth-order solution for DM i.
- **\gamma_{i,k}(t)** The policy for Player i after \( k \)-steps of the G-S iteration.

**APPENDIX A**

Here we provide a proof for (3.11); that is, for problem P1, we prove that \( J(\gamma_{0,1}, \gamma_{0,2}) = J^{(0)} + O(e) \), where \( J^{(0)} \) is given by (3.10).

To compute \( J(\gamma_{0,1}, \gamma_{0,2}) \), we first note that \( J(u_1, u_2) \) can be written in the following manner, using the standard "completing the squares" technique [13]:

\[
J_k(u_1, u_2) = \frac{1}{2} \left\{ \int_{t_0}^{t_f} E(\| u_1 + \tilde{B}_i P_r x \|^2 + \| u_2 + \tilde{B}_j P_r x \|^2) \, dt \right\}
\]
\[ + x_0 P_c(t_0) x_0 + \text{tr} \left[ \Sigma_c(t_0) \right] + \int_{t_0}^{t_f} \text{tr} \left[ P F F' \right] dt \] (A.1)

Before proceeding further, let us introduce the following notation:

\[ \tilde{\epsilon}_i := x_i - \tilde{x}^{(0)}, \quad i = 1, 2; \quad x_\Sigma := (x_1', x_2', \tilde{x}_1', \tilde{x}_2' ), \quad w_\Sigma := (w_1', w_2', v_1', v_2' ). \]

Then, substituting \( \gamma_i(t', I'_i) \) for \( u_i(t) \) in (2.1), and using (3.1) yields

\[ dx_i = \tilde{A} x_i dt + \tilde{F} dw_i; \quad x_i(t_0) = (x_1, x_2, x_1' - \tilde{x}_1', x_2' - \tilde{x}_2'). \] (A.2)

Hence, the matrix \( \Sigma_i := E(x_i, x_i') \) is the solution of the following Riccati equation:

\[ \dot{\Sigma}_i = \Sigma_i \ddot{A} + \tilde{A} \Sigma_i + \tilde{F} \Sigma_i \tilde{F}' ; \quad \Sigma_i(t_0) = \Sigma_{i0} \] (A.4)

where

\[ \Sigma_{i0} = \begin{bmatrix} \Sigma_{10} & 0 & 0 & 0 \\ 0 & \Sigma_{20} & 0 & 0 \\ 0 & 0 & \Sigma_{20} & 0 \\ 0 & 0 & 0 & \Sigma_{20} \end{bmatrix}. \] (A.5)

Now, applying IFT to (A.4) yields

\[ \Sigma_{i33} = \Sigma_{i33}^{(0)} + O(\epsilon) \]
\[ \Sigma_{i44} = \Sigma_{i44}^{(0)} + O(\epsilon) \] (A.6)

where \( \Sigma_{i33} \) denotes the \( (i, j) \)th block of \( \Sigma_i \). Now, substituting \( \gamma_i(t', I'_i) \) for \( u_i(t) \), \( i = 1, 2 \), in (A.1), and using (3.7), (3.8), and (A.6), it follows by the dominated convergence theorem [15], that

\[ J(\gamma_{1'}, \gamma_{2'}) = J^{(0)} + O(\epsilon). \] (A.7)

**APPENDIX B**

**Proof of Lemma 4.1:** Let \( \gamma_1(t, I_1') = L(t) z(t) \), where

\[ dz = G(t) z dt + H(t) dy_2; \quad z(t_0) = M x_0. \]

Then, to obtain \( \gamma_i(t, I'_i) = \arg \min_{\gamma_i \in I'_i} J(\gamma_i, \gamma_2) \), the problem faced by DM1 is the following:

\[ \min_{\gamma_1 \in I_1'} J(\gamma_1(t, I_1')). \]

**APPENDIX C**

**Proof of Lemma 4.2:** If DM2 uses \( \tilde{L}_2(t, I'_2) \), the problem faced by DM1 is given by

\[ \min_{\gamma_2' \in I_2'} J(\gamma_1, \gamma_2') \]

\[ \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ \epsilon A_{21} & A_2 \end{bmatrix} x_\Sigma \] (A.3)

From standard LQG theory, the solution to the above problem is given by

\[ u_i(t) = \gamma_i(t, I'_i) = - \tilde{B}_i \tilde{P} \tilde{x} \]

\[ \dot{x} = \left[ \tilde{A} - \tilde{B}_i \tilde{P} \right] x + \Sigma \tilde{C}' (dy_1 - \tilde{C} x dt) \]

\[ \tilde{p} + \tilde{A} \tilde{P} + \tilde{P} \tilde{A} = \tilde{P} \tilde{B}_i \tilde{P} + \tilde{Q} = 0; \quad \tilde{P}(t_f) = \tilde{Q} \]

\[ \tilde{S} = \tilde{A} \Sigma + \Sigma \tilde{A} - \Sigma \tilde{C}' \tilde{C} \Sigma + \Sigma \tilde{F} \Sigma \]

where

\[ C = (C_1, 0, 0, 0); \quad \Sigma_0 = \text{block diag} \{ \Sigma_{10}, \Sigma_{20} \}. \] (B.4)

Now, using IFT, it is easy to show that

\[ \lim_{\epsilon \to 0} \Sigma_0 \tilde{C}' = (C \Sigma_0^{(0)}, 0, 0)' + O(\epsilon) \]

thus completing the proof.
where

\[ J(u_t) = \frac{1}{2} \mathbb{E} \left\{ x_t^T Q_{1, t} x_t + \tilde{x}_t^T \tilde{Q}_2 \tilde{x}_t + \int_{t_0}^t \left( x_{s}^T Q_{1} x_{s} + \tilde{x}_{s}^T \tilde{Q}_2 \tilde{x}_{s} + u_{s}^T u_{s} \right) ds \right\} \]

and

\[ \tilde{x}_2 := (x_2', \tilde{x})'; \quad \tilde{A}_{12} := (A_{12}, 0); \quad \tilde{A}_{31} := (A_{31}, 0)'; \]

\[ \tilde{A}_2 := \begin{pmatrix} A_2 & B_2 L \tilde{A} \\ \tilde{H} C_2 & \tilde{G} \end{pmatrix}; \quad \tilde{Q}_2 := \text{block diag} \{ \tilde{Q}_2, L; \tilde{L} \} \]

\[ \tilde{P}_2 := \text{block diag} \{ \tilde{P}_2, \tilde{F}_2 \} = \tilde{Q}_2 + \tilde{F}_2 \tilde{P}_2 \tilde{F}_2^T \]

If, instead of \( \tilde{G}(t), \tilde{H}(t), \) and \( \tilde{L}(t) \), we had used

\[ \text{block diag} \{ \tilde{G}, 0 \}, \{ \tilde{H', 0} \}, \text{and} \ (0, 0) \],

respectively, then only the terms \( \tilde{A}_2, \tilde{Q}_2, \) and \( \tilde{F}_2 \) would have changed by \( O(\varepsilon^2) \). This corresponds to using \( \tilde{G}_2(t, \tilde{I}_2) \) instead of \( \tilde{G}_2(t, \tilde{I}_2) \). Referring back to (C.4) and (C.5), it is obvious that this will result only in an \( O(\varepsilon^{k+1}) \) change in \( P_{11}, P_{12}, \Sigma_{11}, \) and \( \Sigma_{12} \) because the terms \( \tilde{A}_2, \tilde{Q}_2, \) and \( \tilde{F}_2 \) enter the equations for \( P_{11}, P_{12}, \Sigma_{11}, \) and \( \Sigma_{12} \) through an \( \varepsilon \) term. Also, since \( \tilde{P}_{11}, \tilde{P}_{12}, \tilde{\Sigma}_{11}, \) and \( \tilde{\Sigma}_{12} \) determine the optimal cost for this problem, the optimal cost will change also by an order of \( O(\varepsilon^{k+1}) \). This completes the proof of Lemma 4.2.

REFERENCES


R. Srikant was born in Arani, India, in 1964. He received the B.Tech. degree from the Indian Institute of Technology, Madras, in 1985, and the M.S. and Ph.D. degrees from the University of Illinois, Urbana-Champaign, in 1988 and 1991, respectively, all in electrical engineering.

From August 1985 to July 1991, he was a Research Assistant at the Coordinated Science Laboratory at the University of Illinois. Since August 1991, he has been working at AT&T Bell Laboratories, Holmdel, NJ. His research interests include stochastic control, decision theory, and the application of perturbation techniques to decentralized control problems.
Tamer Başar (S'71–M'73–SM'79–F'83) was born in Istanbul, Turkey in 1946. He received the B.S.E.E. from Robert College, Istanbul, Turkey, and the M.S., M.Phil, and Ph.D. degrees in engineering and applied science from Yale University, New Haven, CT.

After being at Harvard University, Marmara Research Institute, and Bogaziçi University, he joined the University of Illinois, Urbana-Champaign in 1981, where he is currently a Professor of Electrical and Computer Engineering. He has spent two sabbatical years (1978-1979 and 1987-1988) at Twente University of Technology, The Netherlands, and INRIA, France, respectively.

Prof. Başar has authored or co-authored over one hundred journal articles and book chapters, and numerous conference publications, in the general areas of optimal control, dynamic games, stochastic control, estimation theory, stochastic processes, information theory, and mathematical economics. He is the co-author of the text Dynamic Noncooperative Game Theory (New York: Academic, 1982; 2nd printing 1989), Editor of the volume Dynamic Games and Applications in Economics (New York: Springer-Verlag, 1986), co-editor of Differential Games and Applications (New York: Springer-Verlag, 1988), and co-author of the text $H^\infty$-Optimal Control and Related Minimax Design Problems (Cambridge, MA: Birkhäuser, 1991). He carries memberships in several scientific organizations, among which are Sigma Xi, SIAM, SEDC, ISDG, and the IEEE. He has been active in the IEEE Control Systems Society in various capacities, most recently as an Associate Editor at Large for its TRANSACTIONS, as the Program Chairman of the Conference on Decision and Control in 1989, and as General Chairman in 1992. Currently, he is also the President of the International Society of Dynamic Games, and Associate Editor of two international journals.