Structured Equilibria for Dynamic Games with Asymmetric Information and Dependent Types

Nasimeh Heydaribeni and Achilleas Anastasopoulos

Abstract—We consider a dynamic game with asymmetric information where each player observes privately a noisy version of a (hidden) state of the world $V$, resulting in dependent private observations. We study structured perfect Bayesian equilibria that use private beliefs in their strategies as sufficient statistics for summarizing their observation history. The main difficulty in finding the appropriate sufficient statistic (state) for the structured strategies arises from the fact that players need to construct (private) beliefs on other players’ private beliefs on $V$, which in turn would imply that an infinite hierarchy of beliefs on beliefs needs to be constructed, rendering the problem unsolvable. We show that this is not the case: each player’s belief on other players’ beliefs on $V$ can be characterized by her own belief on $V$ and some appropriately defined public belief. We then specialize this setting to the case of a Linear Quadratic Gaussian (LQG) non-zero-sum game and we characterize linear structured PBE that can be found through a backward/forward algorithm akin to dynamic programming for the standard LQG control problem. Unlike the standard LQG problem, however, some of the required quantities for the Kalman filter are observation-dependent and thus cannot be evaluated off-line through a forward recursion.

I. INTRODUCTION

Dynamic games with asymmetric information play an important role in decision and control problems, yet there is no general framework to study such games in a tractable manner. The appropriate solution concept for these games is some notion of equilibrium such as Bayesian Nash equilibrium, perfect Bayesian equilibrium (PBE), sequential equilibrium, etc. [2]–[4]. Due to the dynamic nature of such games, the players’ histories expand with time and therefore the corresponding strategies have an expanding domain. To mitigate this problem, researchers have introduced equilibrium concepts that summarize the time expanding histories into sufficient statistics. For symmetric information games Markov perfect equilibria [5] have been introduced, in which the players’ strategies depend only on payoff-relevant past events and not the whole history. For asymmetric information games or control problems, finding the appropriate sufficient statistic is a challenging task and various information structures and corresponding statistics have been considered in the literature [6]–[10]. A quantity commonly used as a sufficient statistic, is a belief over some unknown part of the system. The main challenge in this context is the emergence of private beliefs in the sufficient statistics, i.e., the fact that different agents in the system may have different (private) observations about the same quantity.

One way to avoid this problem is to consider models in which private beliefs either do not exist (symmetric information games, or asymmetric but independent observations [6], [7], [11]) or, if they exist, they are not taken into account in agents’ strategies (see for example the concept of “public perfect equilibrium” [12]). In order to intuitively explain the conceptual difficulty arising from having private beliefs in the sufficient statistics, consider the following thought process. If a player acts according to her private belief $\xi^i_t$ of a hidden variable and she expects other players to behave in the same way, she needs to form a belief over other players’ beliefs to interpret and predict their actions and she has to take that belief into account when acting. In other words, she has to form a belief over (at least) $\xi^j_t$ for all other users $j \neq i$. This is a belief on beliefs which is also a private information of user $i$ and it has to be taken into account in her strategies. Due to symmetry of the information structure, all other players should do the same. But now, it is clear that user $i$ needs to form beliefs over beliefs over beliefs of other players. This chain continues as long as this hierarchy of beliefs are private. It stops whenever the beliefs in one step are public or public functions of previous step beliefs.

In this paper, we study a dynamic game with asymmetric information. We consider a model with an unknown state of the world $V$, where each player $i$ has a private noisy observation $X^i_t$ of it at each time $t$. The private observations of players are conditionally independent given $V$. We then specialize this setting to the case of a Linear Quadratic Gaussian (LQG) non-zero-sum game where $V$ is a Gaussian random variable and players’ observations are generated through a linear Gaussian model from $V$. Our LQG model closely follows that of [11] with one important difference: the private observations of players in [11] are independent where in our case, they are dependent through $V$; in particular they are conditionally independent given $V$. Our model can also be thought of as a generalization of the one in [13] where $V$ models the value of a product (or a technology) and agents receive a noisy private signal about it and decide whether to adopt it or not, with the important difference that we allow multiple agents to act simultaneously and, unlike [13], we also allow them to return to the marketplace at each time instance and receive a new observation on $V$.

One of the contributions of this paper is to show that, due to the conditional independence of the private signals given $V$, the private belief chain stops at the second step and players beliefs over others’ beliefs are public functions of their own beliefs (the first step beliefs). In the LQG model, we further show that the beliefs are Gaussian and hence, are characterized...
by their mean and covariance matrix. Furthermore, the players estimation over others’ estimations are public linear functions of their own estimations. We hypothesize (and eventually prove) structured PBE with strategies for user $i$ being linear in $\hat{V}_i$, the private estimate of $V$ by user $i$, generated by a (private) Kalman filter. This is the second contribution of this work.

We show that the equilibrium strategies can be characterized by an appropriate backward sequential decomposition algorithm akin to dynamic programming. In the LQG model, the main difference of our work from the standard stochastic control LQG framework is that the forward recursion that evaluates covariance matrices cannot be performed separately as it depends on the equilibrium strategies. This was also the case in [11]. A unique feature of our development is the requirement to update in a forward manner additional quantities that are observation dependent (public actions). This precludes off-line evaluation of these forward-updated quantities and necessitates their inclusion as part of the state of the above mentioned backward sequential decomposition. This is the third contribution of this work.

A. Literature Review

In this section we give an overview of the related literature with a focus on the information structures. In [14], a framework, called precedence diagram, was introduced to characterize the information structures in team problems with asymmetric information. The evolving (dynamic) information of the decision makers is modeled by a different (new) controller making a decision at each time with the specific information corresponding to that time available to her. The authors have also provided some examples of the dynamic team problems, one of which is LQG team problem with nested information structure and have proved optimality of linear controllers. The specific information structure considered, nested information, allows the authors to form an equivalent static team problem for the dynamic model considered and hence, avoiding further challenges of dealing with dynamic models.

LQG models have been studied extensively for decision and control problems. In the simplest instance of a single centralized controller it is well known that there is separation of estimation and control, posterior beliefs of the state are Gaussian, a sufficient statistic for control is the state estimate evaluated by the Kalman filter, the optimal control is linear in the state estimate, and the required covariance matrices can be calculated offline [15]. Although it is known that, in general, linear controllers are not optimal in LQG team problems [16], as we mentioned, some information structures have been identified for which linear controllers are shown to be optimal such as the works with nested information structure [14], stochastically nested information structure [9] and partial history sharing information structure [8]. Private beliefs do not emerge in these models because of the specific information structure considered. In the nested information structure, there is no need to form beliefs to interpret the action of the predecessors because the decision maker already knows their information. In the model considered in [8], the decision makers have local memory (not perfect memory) and the authors have not defined any summaries for the history and therefore, beliefs and hence, private beliefs are not introduced.

In order to capture the strategic behavior of agents, dynamic decision problems have also been considered in the context of dynamic games and there is extensive literature on dynamic games with asymmetric information. In [17], the author considers a delayed observation sharing model where all of the previous private observations are shared with all of the players and the asymmetry of the information is only due to the private observations at current time. This specific information structure avoids the private beliefs in the sufficient statistics because they can be formed by augmenting the public belief by the current private observation. One-step delayed information sharing is also used in [18]. Similarly, in [13], [19]–[21], there is a public belief that can be augmented by the players’ static private signals, to form the private beliefs.

Authors in [22] have used the common information approach, which breaks the history into the common and private parts and similarly, two partial strategies are introduced. One is applied to the private part of the history and the other one generates the first one based on the public part of the history. Finding the strategy that is generated based on the public part of the history does not have the challenges of asymmetric games because the public part of the history is common between all players. The solution concept used is called common information based Markov perfect equilibria. Note that in [22], the private part of the history is not summarized into any other quantity, and therefore, no private beliefs had to be defined. A similar approach is used in [6].

In [11], authors have considered a multi-stage LQG game and characterized a signaling equilibrium which is linear in agents’ private observations. In addition, a backward sequential decomposition was presented for the construction of the equilibrium, based on the general development in [7]. In this work, the private observations are independent across agents and therefore there are no private beliefs in the game. This is because a player’s belief over others’ private observations is independent of her private observation and hence, the belief is public.

A number of works consider LQG games where information available to some players is affected by the decision of others. The works of [23] on strategic information transmission, and [24] on Gaussian cheap talk consider two-stage games and focus on Bayesian Nash equilibria. These works, however, consider games that are not dynamic. This implies that there is no need to search for the sufficient statistics and no private belief will be defined. The classic work on Bayesian persuasion [25], and the related one on strategic deception [26] consider two-stage and multi-stage games, respectively, and focus on (sender preferred) subgame perfect equilibria owing to the fact that strategies (as opposed to only the actions) of the sender are observed. Although the authors of [26] consider a dynamic game, they do not summarize the history into time invariant quantities and they search for the strategies over the whole time horizon. Therefore, although the problem becomes intractable for large time horizons, the issue of private beliefs
does not appear.

The unique feature of this work is that we consider dependent private observations (specifically, conditionally independent on a hidden state of the world) between agents, in conjunctions with strategies with time-invariant domains, and so sufficient statistics (beliefs) are defined. As a result, we are forced to deal with private beliefs and the aforementioned issue of the infinite sequence of beliefs on beliefs has to be resolved. This is what makes the considered model interesting and more challenging compared to the previous works.

The remaining part of the paper is structured as follows. In Section II the general model is described. Section III is a review of the solution concept that we have considered in this paper. We develop our main results in Section IV. In section V, we describe the specific case of the model that is an LQG game, followed by the development of a concrete example in Section VI together with numerical results. We conclude in Section VII. Most of the proofs of theorems and lemmas are relegated to the Appendices at the end of the paper.

B. Notation

We use upper case letters for scalar and vector random variables and lower case letters for their realizations. We use the notation \( \mathbb{P}(a|b) \) to denote the probability \( \mathbb{P}(A = a|B = b) \) for discrete random variables and to denote the measure \( \mathbb{P}(da|B = b) \) for continuous random variables. Bold upper case letters are used to denote matrices. Subscripts denote time indices and superscripts represent player identities. The notation \( -i \) denotes the set of all players except \( i \). All vectors are column vectors. The transpose of a matrix \( A \) (or vector) is denoted by \( A' \). We use semicolons ";" for vertical concatenation of matrices (or vectors). For any vector (or matrix) with time and player indices, \( a_t^i \) (or \( A_t^i \)), \( a_t^{-i} \) denotes the vertical concatenation of vectors (or matrices) \( a_1^i, a_2^i, \ldots, a_t^i, a_{t+1}^i, \ldots \). Further, \( a_{1:t}^i \) means \( (a_1^i, a_2^i, \ldots, a_t^i) \). In general, for any vector with time and player indices, \( a_t^i \), we remove the superscript to show the vertical concatenation of the whole vectors and we remove the subscript to show the set of all vectors for all times. The matrix of all zeros with appropriate dimensions is denoted by \( 0 \) and the identity matrix of appropriate dimensions is denoted by \( I \). For two matrices \( A \) and \( B \), \( \mathcal{S}(A,B) \) represents the block diagonal concatenation of these matrices, i.e.,
\[
\begin{bmatrix}
A & 0 \\
0 & B
\end{bmatrix}
\]
(it applies for any number of matrices).

By \( \mathcal{S}(A^{-i}) \), we mean the block diagonal concatenation of matrices \( A_j \) for \( j \neq i \). Further, \( \mathcal{S}(A;B) \) represents \( B'AB \).

For the equation \( \left[ \begin{array}{c} a; \\
b; \\
c \end{array} \right] = A \left[ \begin{array}{c} a; \\
b; \\
c \end{array} \right] \), the notation \( (A)_{a,b} \) denotes the submatrix of \( A \) corresponding to rows \( a \) and columns \( b \). We use "::" for either of the row or column subscripts to indicate the whole rows or columns, e.g., \( (A)_{::b} \) denotes the submatrix of \( A \) corresponding to columns \( b \). The trace of matrix \( A \) is denoted by \( \text{tr}(A) \). We use \( \delta(\cdot) \) for the Dirac delta function. For any Euclidean set \( S \), \( \Delta(S) \) represents the space of all probability measures on \( S \).

II. MODEL

We consider a discrete time dynamic system with \( \mathcal{N} = \{1, 2, ..., N\} \) strategic players over a finite time horizon \( T = \{1, 2, ..., T\} \). There is a static unknown state of the world \( V \sim Q_V(\cdot) \). Each player has a private noisy observation \( X_t^i \) of \( V \) at every time step \( t \in T \). At time \( t \), player \( i \) takes action \( a_t^i \in A^i \) which is observed publicly by all players. The private observations are generated according to the kernel \( X_t^i \sim Q_X^i(\cdot|V, A_{t-1}) \) and they are independent across agents given \( V \) and \( A_{t-1} \), i.e.,
\[
\mathbb{P}(X_t^i|V, A_{1:t-1}, X_{1:t-1}) = \prod_{i \in \mathcal{N}} Q_X^i(X_t^i|V, A_{t-1}) 
\]

We assume that players have perfect recall and we can construct the history of the system at time \( t \) as \( h_t = (v, x_{1:t}, a_{1:t-1}) \in \mathcal{H}_t \) and the information set of player \( i \) at time \( t \) as \( h_t^i = (x_{1:t}, a_{1:t-1}) \in \mathcal{H}_t^i \). At the end of time step \( t \), each player \( i \) receives the reward \( r_t^i(v, a_t) \). We assume that the rewards are not observed by the players until the end of the time horizon.

Let \( g^i = (g_t^i)_{t \in T} \) be a probabilistic strategy of player \( i \), where \( g_t^i : \mathcal{H}_t^i \to \Delta(A^i) \), meaning that player \( i \)'s action at time \( t \) is generated according to the distribution \( A_t^i \sim g^i_t(\cdot|h_t^i) \). The strategy profile of all players is denoted by \( g \). For the strategy profile \( g \), player \( i \)'s total expected reward is
\[
J^{i,g} := \mathbb{E}^g \left\{ \sum_{t=1}^{T} r_t^i(V, A_t) \right\},
\]
and her objective is to maximize her total expected reward.

III. SOLUTION CONCEPT

We can model this system as a dynamic game with asymmetric information and an appropriate solution concept for such games is Perfect Bayesian Equilibrium (PBE). A PBE consists of a pair \((\beta, \mu)\) (an assessment) of strategy profile \( \beta = (\beta_t^i)_{t \in T, i \in \mathcal{N}} \) and belief system \( \mu = (\mu_t^i)_{t \in T, i \in \mathcal{N}} \) where \( \mu_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{H}_t^i) \) satisfies Bayesian updating and sequential rationality holds. Bayesian updating includes both on and off equilibrium histories. This condition requires the beliefs to be Bayesian updated, if possible, given any history, whether that history is on equilibrium or off equilibrium [4]. To be more specific, given history \( h_t^i \), which could be on or off-equilibrium, and the realizations at time \( t \), i.e., \( a_t, x_{t+1}^i \), the beliefs should be updated according to Bayes rule if \( \mathbb{P}^g(a_t, x_{t+1}^i|h_t^i) > 0 \). Otherwise, the beliefs could be updated arbitrarily. For any \( i \in \mathcal{N}, \ t \in T, h_t^i \in \mathcal{H}_t^i, \beta^i, \) sequential rationality imposes the following condition for the strategy profile \( \beta \):
\[
\mathbb{E}^{\beta^i} \left\{ \sum_{n=t}^{T} r_n^i(V, A_n)|h_t^i \right\} \geq \mathbb{E}^{\tilde{\beta}^i} \left\{ \sum_{n=t}^{T} r_n^i(V, A_n)|h_t^i \right\}
\]

IV. STRUCTURED PBE

The domain of the strategies \( g_t^i(\cdot|h_t^i) \) is expanding in time. Finding such strategies is complicated with the complexity growing exponentially with the time horizon. For this reason, we consider summaries for \( h_t^i \in \mathcal{H}_t^i, \) i.e., \( S(h_t^i) \), that are time invariant. We are interested in PBEs with strategies, \( g_t^i(\cdot|h_t^i) = \)
ψ t (|S(h t ))), that are functions of h t only through the summaries S(h t ). These PBEs are called structured PBEs [7]. In contrast to H t , the set of summaries does not grow in time and therefore, finding such structured PBEs is less complicated than a general PBE. According to [7], we can show that players can guarantee the same rewards by playing structured strategies compared to the general non-structured ones. In dynamic games with asymmetric information, summaries are usually the belief of players over the unknown variables of the game.

Define the private beliefs over the unknown state of the world V as
\[ \xi_i^t(v) = \mathbb{P}^g(v|h_t^i) = \mathbb{P}^g(v|x_{1:t}, a_{1:t-1}) \]

We further define the conditional public belief over the private beliefs as follows
\[ \pi_t(\xi_i^t|v) = \mathbb{P}^g(\xi_i^t|v, h_t^i) = \mathbb{P}^g(\xi_i^t|v, a_{1:t-1}). \]

Lemma 1 (Conditional Independence of Private Beliefs): We have the following equation for the conditional public belief
\[ \pi_t(\xi_i^t|v) = \prod_{i \in N} \pi_i^t(\xi_i^t|v), \]
where \( \pi_i^t(\xi_i^t|v) = \mathbb{P}(\xi_i^t|v, a_{1:t-1}) \). Similarly, we have
\[ \mathbb{P}^g(x_{1:t}|v, a_{1:t-1}) = \prod_{i \in N} \mathbb{P}^g(x_{1:t}^i|v, a_{1:t-1}). \]

Proof: See Appendix A.

Note that this conditional independence holds regardless of the strategy profiles \( g \). Using this result, and with a slight abuse of notation, we can summarize the conditional public belief into the vector \( \pi_t = [\pi_1^t, \ldots, \pi_N^t] \).

We are interested in strategies of the form \( \xi_i^t \sim \psi_t(\xi_i^t, \pi_t) = \gamma_t(\xi_i^t) \), where \( \gamma_t = \theta(\pi_t) \) and we will prove that such structured strategies form a PBE of the game. Note that with the above decomposition of the strategy \( \psi \) into partial strategies \( \gamma \) and the strategy \( \theta \), designing strategies \( \psi \) is equivalent to designing \( \theta \).

A. Belief Update

In this subsection, we present two lemmas regarding the beliefs and their update rules.

Lemma 2: The private beliefs can be updated as \( \xi_{i+1}^t = F_t^i(\xi_i^t, \pi_i^t, a_i^t, x_{i+1}^t) \), where \( F_t^i \) is defined through
\[ \xi_{i+1}^t(v) = \frac{\int_{v_{i-1}} \xi_i^t(v) \prod_{j \in i} \pi_i^t(\xi_j^t|v) \gamma_j^t(a_j^t|\xi_j^t) Q_X(x_{i+1}^t|v, a_i^t) \mathbb{d}v}{\int_{v_{i-1}} \xi_i^t(v) \prod_{j \in i} \pi_i^t(\xi_j^t|v) \gamma_j^t(a_j^t|\xi_j^t) Q_X(x_{i+1}^t|v, a_i^t) \mathbb{d}v} \]

Proof: See Appendix B.

Note that this update depends on the strategy profile \( g \) only through the partial function \( \gamma_t \), i.e., it is independent of the strategy \( \theta \). We will also use the notation \( \xi_{i+1} = F_t^i(\xi_i^t, \pi_i^t, \gamma_t^i, a_i^t, x_{i+1}^t) \) for the update function of the vector of private beliefs.

Lemma 3: The conditional public beliefs can be updated as \( \pi_{i+1}^t = F_t^i(\pi_t, \gamma_t^i, a_t) \), where \( F_t^i \) is defined through
\[ \pi_{i+1}^t(\xi_i^t|v) = \frac{\int_{v_{i-1}} \pi_i^t(\xi_i^t|v, \gamma_t^i(a_t^i|\xi_i^t)) Q_X(x_{i+1}^t|v, a_t^i) \mathbb{d}v}{\int_{v_{i-1}} \pi_i^t(\xi_i^t|v) \gamma_t^i(a_t^i|\xi_i^t) Q_X(x_{i+1}^t|v, a_t^i) \mathbb{d}v} \]

Proof: See Appendix C.

B. Equilibrium Strategies

In this subsection, we will show that structured strategies of the form \( \gamma_t(\xi_i^t) \), where \( \gamma_t = \theta(\pi_t) \) form sPBE of the game. The following theorem formalizes this result and presents the fixed point equation characterizing the equilibrium strategies.

Theorem 1: The strategy profile \( \gamma_t^i = \theta^{\pi_i^t}(\pi_t) \) characterized by the following fixed point equation, forms a sPBE of the game. For all \( i \in N \),
\[ \gamma_t^{a_t^i}(\xi_i^t) = \arg \max_{\gamma_t(\xi_i^t)} \mathbb{E}[r_t^i(\pi_t, \xi_i^t, a_t^i) + J_{t+1}^i(F_t^i(\pi_t, \gamma_t^i, a_t^i), F_t^i(\xi_i^t, \pi_t, \gamma_t^{a_t^i-i}, a_t, X_{t+1}^i))|\pi_t, \xi_i^t], \]

where \( r_t^i(\pi_t, \xi_i^t, a_t^i) = \mathbb{E}[r_t^i(V, A_t)|\pi_t, \xi_i^t, a_t^i] \).

Proof: See Appendix D.

We remark here that in equation (10) the update rule of the public belief \( \pi_t \) is using the equilibrium strategies \( \gamma_t^i \) and therefore, for each time instance \( t \), the collection of equations of the form (10a) for all \( i \in N \) constitutes a fixed point equation over the strategy profile \( \gamma_t^i \). The reason for this is that in characterizing a PBE, one needs to fix the belief structure and then finds the equilibrium strategies corresponding to those beliefs. On the other hand, the beliefs have to be consistent with the equilibrium strategies. This creates a fixed point equation over \( \gamma_t^{a_t^i} \). Furthermore, the above equation has to be solved simultaneously for all \( i \in N \), thus creating the fixed point equation over the strategy \( \gamma_t^i \).

C. Discussion

In this section, we characterized the sufficient statistics of the histories of the considered dynamic game. As we...
mentioned in the Introduction, these summaries include private beliefs, \( \xi_i^t \). One may wonder how we resolved the issue with the chain of private beliefs that was discussed in the Introduction. In other words, how did we resolve the issue of possibly requiring an infinite hierarchy of beliefs on beliefs. In the previous development, we actually proved that this chain stops at the second step. To see this, consider the introduction of private beliefs over others’ private beliefs, i.e., \( \mathbb{P}(\xi_{t-1}^i|h_{t}^i) \). The results of Lemma 1 show that

\[
\mathbb{P}(\xi_{t-1}^i|h_{t}^i) = \int \mathbb{P}(\xi_{t-1}^i, v|h_{t}^i) = \int \mathbb{P}(v|h_{t}^i) \mathbb{P}(\xi_{t-1}^i|v, h_{t}^i) = \int \mathbb{P}(\xi_{t-1}^i|v, h_{t}^i) \mathbb{P}(\xi_{t-1}^i|v, h_{t}^i) \mathbb{P}(v|h_{t}^i) = \int \mathbb{P}(\xi_{t-1}^i|v, a_{t-1}) \mathbb{P}(\xi_{t-1}^i|v, a_{t-1}) \mathbb{P}(v|h_{t}^i) = \int \mathbb{P}(\xi_{t-1}^i|v) \mathbb{P}(h_{t}^i),
\]

where (a) is due to the definition of the private beliefs and (7). The above implies that these beliefs can be evaluated by the public information, \( \pi_i \), and the first order private beliefs \( \xi_i^t \). This is the exact reason why \( \pi_i(\xi_i^t|v) \) was defined.

V. LQG Model

In this section, we study a specific instance of the model discussed so far which is the case when the unknown state of the world, \( V \), is a Gaussian random variable, the private observation kernels are linear and Gaussian and the instantaneous reward is quadratic. Therefore we have an LQG model. The motivation for studying this model stems from the general development in the previous section. In particular we required that equilibrium strategies are generated based on private beliefs and public beliefs on beliefs. In the LQG setting these beliefs can be greatly simplified, thus enabling us to more succinctly characterize the equilibrium strategies discussed in the previous section.

In this model, we consider an unknown state of the world \( V \sim N(0, \Sigma) \) with size \( N_v \). Each player has a private noisy observation \( X^i_t \) of \( V \) at every time step \( t \in T \)

\[
x^i_t = v + w^i_t,
\]

where \( W^i_t \sim N(0, Q^i) \) and all of the noise random vectors \( W^i_t \) are independent across \( i \) and \( t \) and also independent of \( V \). The values of \( \Sigma \) and \( Q^i \), \( \forall i \in N \) are common knowledge between players. Note that in order to maintain the linearity of private observations, we have considered uncontrolled private observations unlike the general model in first part of the paper. More discussion on this matter can be found in section V-C.

We have \( a^i_t \in \mathcal{A}^i = \mathbb{R}^{N_v} \). The instantaneous reward\(^2\) is given by

\[
r^i_t(v, a^i_t) = [v' a^i_t] R^i_t [v a^i_t] = \mathcal{Q}d(R^i_t; [v a^i_t]),
\]

where \( R^i_t \) is a symmetric matrix of appropriate dimensions.

A. Equilibrium Beliefs

In this setting, we will show that the private beliefs \( \xi_i^t \) are Gaussian and since any Gaussian belief can be expressed in terms of its mean and covariance matrix, we define the summaries such that they include the mean and covariance matrices of the beliefs of the players over \( V \). The mean of each player’s belief, i.e., her estimate of \( V \), will be her private information. The covariance matrix, however, can be calculated publicly. We define the private estimate of players over \( V \) as follows. For all \( i \in N_v, t \in T \),

\[
\tilde{v}_i^t = \mathbb{E}[V|h_{t}^i] = \mathbb{E}[V|x_{1:t}^i, a_{1:t-1}],
\]

Since the private beliefs can be expressed in terms of their means and covariance matrices and since the covariance matrices are publicly calculated, the conditional public belief \( \pi_i(\xi_i^t|v) \) is equivalent to a belief over the private estimates. Intuitively, each player, in addition to her own estimate of \( V \), needs to interpret actions of others and predict their future actions. Hence, each player needs to have a belief over the estimates of other players on \( V \). We will show that this latter belief is also Gaussian and therefore, one needs to keep track of only its mean and covariance. We define the following quantity for all \( i \in N_v, t \in T \),

\[
\check{v}_i^{i,j} = \mathbb{E}[\tilde{V}_i^j|h_{t}^i] = \mathbb{E}[\tilde{V}_i^j|x_{1:t}^i, a_{1:t-1}],
\]

The quantity \( \check{v}_i^{i,j} \) is player \( i \)'s best estimate of \( V \) given her observations up to time \( t \). As mentioned before, this quantity is a private estimation for player \( i \) and is not measurable with respect to the sigma algebra generated by the observations of any other player \( j \). Hence, player \( i \) should form an estimate over the private estimates of other players and this is the reason \( \check{v}_i^{i,j} \) is defined. This in turn implies that players’ strategies should also be a function of their estimates over others’ estimates of \( V \). Hence, the same argument as the one in the first part of the paper about private beliefs holds and we need to define an estimate over estimates of players over other players’ estimates of \( V \). This argument continues as long as these estimates are private. Therefore, once again, we are faced with the problem of having to define a chain of private beliefs which are expressed as private estimates in this model. This chain stops whenever one of the estimates of players is public (or a public function of previous-step private estimates) and therefore, there is no need to form an estimate over it.

Indeed, we will show that \( \check{v}_i^{i,j} \) is a linear public function of \( \check{v}_i^{i,j} \), hence, there is no need to include \( \check{v}_i^{i,j} \) in the private part of the summary \( S(h_{t}^i) \) and therefore, no other player needs to form an estimate over it. The summary we use for \( h_{t}^i \) is defined as \( S(h_{t}^i) = (\check{v}_i^{i,j}, P(h_{t}^i)) \), where \( P(h_{t}^i) \) is the public

\(^2\)Unlike more standard LQG setting we consider “rewards” instead of “costs” to maintain consistency with the general problem discussed earlier.
Theorem 2: Assuming pure linear strategies of the form
\[ \hat{a}_i^t = \mathbf{L}_i^t \hat{v}_i^t - m_i^t, \]
in the summary, the private belief update function can be evaluated publicly. Consequently, the public belief \( \pi_i(\xi_t^i | v) \) can be reduced to a belief \( \pi_i^t(\hat{v}_i^t | v) \). Furthermore, \( \pi_i^t(\hat{v}_i^t | v) \) is Gaussian with mean \( E_i^t v + f_i^t \), where matrices \( E_i^t \), \( f_i^t \) can be evaluated publicly.

Proof: See Appendix E.

In the following we summarize the parameters needed to update each of the quantities introduced in the proof of Theorem 2 and we introduce update functions for each one.

\[
\begin{align*}
\hat{v}_i^{t+1} &= F_t(\hat{v}_i^t, \Sigma_i^{t+1}, E_i^t - f_i^t, L_i^{t-i}, m_i^{t-i}, a_t, x_{t+1}^i) \\
\Sigma_i^{t+1} &= F_{\Sigma_i}(\Sigma_i^{t+1}, L_t^{t-i}) \\
\Sigma_i^{t+2} &= F_{\Sigma_i}(\Sigma_i^{t+1}, E_t, L_t) \\
E_t &= F_E(E_t, \Sigma_i^{t+1}, \Sigma_i^{t+1}, L_t) \\
f_t &= F_f(f_t, \Sigma_i^{t+1}, \Sigma_i^{t+1}, E_t, L_t, m_t, a_t)
\end{align*}
\]

Equations (16a) and (16b) correspond to the private belief update and are similar in structure to the update function \( F_t \) of \( \xi_t^i \) in Lemma 2 for the general case. The remaining update functions correspond to the public belief update \( F_p \) in Lemma 3 for the general case.

Note that according to the above equations, the quantities \( \Sigma_i^{t+1}, \Sigma_i^{t+1}, E_t \) are updated recursively using the strategy matrices \( \mathbf{L}_i \). Hence, if one knows the strategies, one can calculate these quantities offline for the entire time horizon of the game. However, the quantity \( f_t \) is updated using the strategy matrices \( \mathbf{L}_i \) and vectors \( m_k \) as well as the realized actions \( a_t \) and therefore, they cannot be evaluated offline.

We reiterate at this point that Theorem 2 implies that the estimate of player \( i \) over private estimates of players \(-i\), i.e., \( \hat{v}_i^{t-i} \), is a linear function of \( \hat{v}_t \),

\[
\hat{v}_i^{t-i} = E(\hat{V}_t^{t-i} | h_t^i) = E(\hat{V}_t^{t-i} | V, A_{1:t-1}) | h_t^i = E(\hat{V}_t^{t-i} V + f_{t-i} | h_t^i) = E_t^{t-i} \hat{v}_i + f_{t-i},
\]

with matrices \( E_t^{t-i} \) and \( f_{t-i} \) being public information. As a result, assuming linear strategies of the form \( a_i^t = \mathbf{L}_i^t \hat{v}_i^t + m_i^t \) at equilibrium, one can form the summary \( S(h_t^i) = (\hat{v}_t, P(h_t^i)) \) and base the selection of the matrices \( \mathbf{L}_i^t \) and \( m_i^t \) on the public part of this summary, \( P(h_t^i) \). In the next section we show that indeed linear strategies can form an equilibrium and provide a methodology to find the quantities \( \mathbf{L}_i^t \) and \( m_i^t \).

B. Linear Structured PBE

Theorem 2 implies that \( S_t^i = \{ V, \hat{V}_t^{-1} \} \) is a jointly Gaussian random vector conditioned on player \( i \)'s observation till time \( t \), \( \forall i \in \mathcal{N}, t \in \mathcal{T} \). This implies that the beliefs over \( V \) are jointly Gaussian and so players need only keep track of their belief's mean (estimation) and covariance matrices. Furthermore, this theorem implies that a player's belief over other players' beliefs is also Gaussian and hence, players need to keep track of their estimation on other players' estimations, i.e., \( \hat{v}_i \). The important point of Theorem 2 is the statement that the estimation of players on others' estimations is a linear function of their own estimation and hence, in order to keep track of the estimation over other players' estimations, a player only needs to keep track of her own estimation over \( V \). Therefore, \( \hat{v}_i^t \) is a sufficient statistic for player \( i \)'s private observations till time \( t \).

In terms of the public summary, we see four public quantities, \( \Sigma_{t+1|i}, \Sigma_{t+1}, E_t \) and \( f_t \), in (16). With some abuse of notation, we define \( \Sigma = [\Sigma_{t+1|i}, \Sigma_{t+1}] \). We will show that the tuple \( (\Sigma_t, E_t, f_t) \) is the public summary of \( h_t^i \), i.e., \( P(h_t^i) \). Note that \( E_t \) and \( f_t \) are involved in the expression for the mean of the conditional public belief over \( \hat{v}_i \), hence, they correspond to the conditional public belief \( \pi_t \) in the first part of the paper. The covariance matrices \( \Sigma_{t+1|i}, \Sigma_{t+1} \) represent the covariance matrices of the private and conditional public beliefs. This implies that by having the tuple \( (\hat{v}_t^i, \Sigma_t, E_t, f_t) \), we have full characterization of the private and public belief and therefore, we have the summaries for the LQG game.

Therefore, we consider strategies of the form \( \gamma_i^t(\hat{v}_i^t, \Sigma_t, E_t, f_t) = \gamma_i^t(\hat{v}_i^t) \). In particular, we will now show that the strategy profile \( \gamma_i^t(\hat{v}_i^t, \Sigma_t, E_t, f_t) = \gamma_i^t(\hat{v}_i^t) \) \( \forall i \in \mathcal{N} \) where \( \gamma_i^t(\hat{v}_i^t) = \delta(a_i^t - \mathbf{L}_i^t \hat{v}_i - m_i^t) \), together with the corresponding Gaussian beliefs derived in Theorem 2, form a structured PBE of the game.

Theorem 3: The strategy profile \( \psi_i^t(\hat{v}_i^t, \Sigma_t, E_t, f_t, \hat{v}_i^t, \Sigma_t, E_t, f_t) = \gamma_i^t(\hat{v}_i^t) \) \( \forall i \in \mathcal{N} \), where \( \pi_i^t(\hat{v}_i^t, \Sigma_t, E_t, f_t) \) and \( f_t \) are PBE of the game.

Proof: See Appendix E.

One important result from the proof of Theorem 3 is that the reward to go, \( J_t^i(\hat{v}_i^t, \Sigma_t, E_t, f_t) \) is quadratic with respect to \( \hat{v}_t^i \) and \( f_t \), which are the only quantities in the summary that cannot be evaluated offline, i.e., we have

\[
J_t^i(\hat{v}_i^t, \Sigma_t, E_t, f_t) = q\mathbb{d}(Z_t^i; \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix}) + z_t^i \begin{bmatrix} \hat{v}_t^i \\ f_t \end{bmatrix} + o_t^i. \tag{18}
\]

Therefore, if we have the quantities \( Z_t^i, z_t^i, \) and \( o_t^i \) we can evaluate the reward to go for every value of \( \hat{v}_t^i \) and \( f_t \).

In the following, we propose a backward algorithm that evaluates the quantities \( Z_t^i, z_t^i, \) and \( o_t^i \) as well as the strategy matrices \( \mathbf{L}_i, \mathbf{M}_i \) and vectors \( m_i \) (we have \( m_i^t = \mathbf{M}_i^t f_t + \hat{m}_i^t \), according to the proof of Theorem 3) as functions of
Before stating the algorithm, we define the following functions.

\[ L_t = g_L(t, \Sigma_t, E_t) \quad (19a) \]
\[ M_t = g_M(t, \Sigma_t, E_t) \quad (19b) \]
\[ \bar{m}_t = g_{\bar{m}}(\Sigma_t, E_t) \quad (19c) \]
\[ Z_t = \psi_{Z,t}(\Sigma_t, E_t) \quad (19d) \]
\[ z_t = \psi_{z,t}(\Sigma_t, E_t) \quad (19e) \]
\[ o_t = \psi_{o,t}(\Sigma_t, E_t) \quad (19f) \]

**Backward Algorithm (Offline)**

1) Set \( t = T \). Set \( Z_{T+1} = \psi_{Z,T+1}(\Sigma_{T+1}, E_{T+1}) = 0 \), \( z_{T+1} = \psi_{z,T+1}(\Sigma_{T+1}, E_{T+1}) = 0 \) and \( o_{T+1} = \psi_{o,T+1}(\Sigma_{T+1}, E_{T+1}) \) for every \( \Sigma_{T+1}, E_{T+1} \).
2) Calculate \( L_T = g_L(T, \Sigma_T, E_T) \), \( M_T = g_M(T, \Sigma_T, E_T) \), \( \bar{m}_T = g_{\bar{m}}(\Sigma_T, E_T) \), and \( Z_T = \psi_{Z,T}(\Sigma_T, E_T) \) for every \( \Sigma_T, E_T \) and the corresponding \( \psi_{Z,T+1}(\cdot, \cdot) \) according to equation (68) and (70).
3) Set \( t = t - 1 \).
4) If \( t \geq 1 \), Go to step 3. Else stop.

Using the functions defined above, one can run the following forward algorithm to find the strategy matrices \( L_t \), \( M_t \) and vectors \( \bar{m}_t \) and the quantities \( Z_t, Z^*, o_t \).

**Forward Algorithm (Offline)**

1) Set \( t = 1 \).
2) Initialize the value of \( \Sigma_1 \) and \( E_1 \) according to equations (46) and (54).
3) Using \( \Sigma_t \) and \( E_t \), find \( L_t, M_t, \bar{m}_t \) and the quantities \( Z_t, Z^*, o_t \) according to equation (19).
4) Using \( \Sigma_t, E_t \) and \( L_t \), calculate \( \Sigma_{t+1} \) and \( E_{t+1} \) according to equations (16).
5) Set \( t = t + 1 \).
6) If \( t \leq T \), Go to step 3. Else stop.

**C. Model Extensions**

In this section, we investigate alternative models that can be studied with the methodology introduced in this paper and we explain how the results can be extended to such models.

As it is clear in equation (12), in the LQG model considered in this paper, the private observations are not controlled by the actions, unlike the general model of the first part of the paper. If we were to add control actions to equation (12), in order to maintain linearity, we would have added a term such as \( B^2_1 a_t \) and therefore, equation (12) would have looked like \( x^*_t = v + w^*_1 + B^2_1 a_t \). Since the actions are publicly observed, the amount of information that player i extracts from \( V \) remains the same with or without the term \( B^2_1 a_t \). Hence, because the private observations serve only as measurements of \( V \), adding control to equation (12) does not make any difference in the results.

Controlled private observations could make a difference in the LQG model if the private observations could affect the instantaneous rewards. That is, if the reward was \( r^*_i(v, a_t, x^*_t) = \sum(\mathcal{R}_t; \begin{bmatrix} v \\ a_t \\ x^*_t \end{bmatrix}) \). Note that the amount of information that \( x^*_t \) conveys about \( V \) is still the same as in the uncontrolled case. We can show that results similar to all of the ones in this paper will hold for this model with controlled private observations and this type of instantaneous reward. Note that in this case, the strategies would be linear in both the private estimation and the latest private observation.

We can also extend our results of the first part of the paper (the general model) to a model with the instantaneous reward being of the form of \( r^*_i(v, a_t, x^*_t) \). In this case, \( x^*_t \) should be added to the summaries and the results will hold.

**VI. Example**

In this section, we describe some numerical examples to show the equilibrium strategies discussed in this paper. In these examples, we derive the equilibrium strategies by solving a fixed point equation for the entire time horizon using the following algorithm. Note that the superscript \( (k) \) in \( A^{(k)} \) denotes the number of iterations performed. We define the convergence error as \( \epsilon^{(k)} = \max(|L_i^{(k)} - L_i^{(k+1)}|, |M_i^{(k)} - M_i^{(k+1)}|, |\bar{m}_i^{(k+1)} - \bar{m}_i^{(k)}|) \).

**Numerical Algorithm (Offline)**

1) Set \( k = 1 \).
2) Initialize \( L_1^{(1)} \), \( M_1^{(1)} \), and \( \bar{m}_1^{(1)} \) arbitrarily.
3) Using \( L_i^{(k)} \), evaluate \( \Sigma_1^{(k+1)} \), \( E_1^{(k+1)} \) according to equations (16) in a forward manner (using initial conditions \( \Sigma_1 \) and \( E_1 \) according to equations (46) and (54)).
4) Using \( L_i^{(k)} \), \( M_1^{(k)} = \Sigma_i^{(k)} \), \( \bar{m}_1^{(k)} \), \( \Sigma_i^{(k+1)} \), \( E_1^{(k+1)} \), evaluate \( L_i^{(k+1)} \), \( M_i^{(k+1)} \), and \( \bar{m}_i^{(k+1)} \) according to the backward algorithm.
5) Evaluate \( \epsilon^{(k)} \). If it is below the desired threshold, stop. Otherwise, go to step 4.

Note that in each step of the backward algorithm, one needs to solve a fixed point equation with respect to the strategy matrices and vectors to derive the functions defined in eq. (19) (see eq. (68) in Appendix F). However, in the numerical algorithm described above, we use the last iteration quantities for the right hand side of the equations and consequently, we do not need to solve any fixed point equations.

As a concrete example, we consider a setting where there is a project with an unknown attribute denoted by \( v \). There are two agents working on this project exerting a costly effort \( a_1^* \). The agents are rewarded based on the alignment of their effort with the project attribute, \( v \), as well as based on their cooperation. At each time slot, the agents have private observations, \( x_i^* \), of the project attribute. We consider two instances of the game where \( v \) is scalar in one and two dimensional vector in the other, while the efforts are scalars in both.

**A. Scalar State and Action**

We model the considered scenario for scalar \( v \) and scalar actions \( a_1^* \) with the instantaneous rewards being \( R_1^*(v, a_1) = a_1^* v + \frac{1}{2} a_1^2 a_1^2 - (a_1^*)^2 \) and \( R_2^*(v, a_1) = a_1^* v + \frac{1}{2} a_1^2 a_1^2 - (a_1^*)^2 \). That
is, we set \( R_1^1 = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \) and \( R_1^2 = \begin{bmatrix} 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & -1 \end{bmatrix} \).

Note that the term \( a^1_i v \) in the instantaneous rewards accounts for the alignment of \( a^1_i \) with \( v \), and the term \( a^2_i a^2_i \) denotes the cooperation between the agents.

Case 1: If we assume that agents perfectly observe \( V \), i.e., if we set \( Q^1 = 0 \) and \( Q^2 = 0 \), the following linear equilibrium strategy matrices and vectors are derived from the numerical analysis of this game for \( T = 2 \) and \( \Sigma = 1 \)

\[
L^1_1 = \frac{2}{3} \quad L^1_2 = \frac{2}{3} \\
L^2_1 = \frac{2}{3} \quad L^2_2 = \frac{2}{3}.
\] (20a)

Furthermore, we have \( \bar{m}^i_t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \). Note that since in this case, \( f_i = 0 \) for \( t = 1, 2 \), the strategy matrices \( M^i_t \) will not play any roles and are not presented here. These results imply that each agent will exert effort exactly equal to \( \frac{2}{3} V \). As it turns out, these strategies are myopic, i.e., we also observe these strategies in the case \( T = 1 \). The reason for having myopic strategies is that the observations are perfect and hence, the actions have no effect in shaping the future beliefs.

Case 2: Consider agents with equally imperfect observations, \( Q^1 = Q^2 = 1 \). The following strategy matrices are derived

\[
L^1_1 = 0.6722 \quad L^1_2 = 0.6722 \\
L^2_1 = 0.5333 \quad L^2_2 = 0.5333
\] (21a)

\[
M^1_1 = [0.0561 \ 0.2620] \quad M^1_2 = [0.2620 \ 0.0561] \\
M^2_1 = [0.0356 \ 0.1422] \quad M^2_2 = [0.1422 \ 0.0356],
\] (21b)

and \( \bar{m}^i_t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \). One can explain these results by paying attention to the interactions between the agents. At \( t = 1 \), agent one has a better estimation of \( V \) compared to agent two and therefore, she has higher \( L^1_1 \). At \( t = 2 \), agent two has learned the estimation of agent one through her action at \( t = 1 \) and therefore, the two agents have almost equal estimations. But this time, agent two exerts slightly higher effort to compensate agent one’s efforts at \( t = 1 \).

Case 3: If one agent has better observations than the other, i.e., \( Q^1 = 1 \), \( Q^2 = 2 \), the strategy matrices are changed as follows.

\[
L^1_1 = 0.6700 \quad L^1_2 = 0.6619 \\
L^2_1 = 0.5224 \quad L^2_2 = 0.5373
\] (22a)

\[
M^1_1 = [0.0520 \ 0.2701] \quad M^1_2 = [0.2738 \ 0.0605] \\
M^2_1 = [0.0348 \ 0.1433] \quad M^2_2 = [0.1393 \ 0.0358],
\] (22b)

and \( \bar{m}^i_t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \). One can explain these results by paying attention to the interactions between the agents. At \( t = 1 \), agent one has a better estimation of \( V \) compared to agent two and therefore, she has higher \( L^1_1 \). At \( t = 2 \), agent two has learned the estimation of agent one through her action at \( t = 1 \) and therefore, the two agents have almost equal estimations. But this time, agent two exerts slightly higher effort to compensate agent one’s efforts at \( t = 1 \).

Case 4: The interaction between agents can also be seen in a scenario where one agent has perfect observations and the other one has partial observations, i.e., \( Q^1 = 0 \), \( Q^2 = 2 \). The strategy matrices are given as follows.

\[
L^1_1 = 0.7125 \quad L^1_2 = 0.6781 \\
L^2_1 = 0.5000 \quad L^2_2 = 0.6250
\] (23a)

\[
M^1_1 = [0.0142 \ 0.1808] \quad M^1_2 = [0.1817 \ 0.0452] \\
M^2_1 = [0.0333 \ 0.1667] \quad M^2_2 = [0.1333 \ 0.0417],
\] (23b)

and \( \bar{m}^i_t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \).

Case 5: Finally, consider a case where both agents have very noisy observations, that is \( Q^1 \), \( Q^2 \) are large numbers. In this case, \( \bar{v}^i_t = 0 \) and \( f_i = 0 \). Therefore, the strategy matrices \( L^i_t \) and \( M^i_t \) do not play any roles and the actions will only follow \( \bar{m}^i_t \). For this game we obtain \( \bar{m}^i_t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \).

Case 6: We have also derived the strategy matrices of the game for larger values of \( T \). In Figure 1, we can see the plot of the strategy matrices \( L^i_t \) with respect to time for the symmetric case of \( Q^1 = Q^2 = 1 \) and for \( T = 10 \). As before, we observe a trend where as time goes by, the values of the strategy matrices decrease. The intuition behind why such behavior is observed is that more public information is observed as time goes by. Therefore, the players estimation over others’ estimations is mainly characterized by the public part of the state, \( f_i \), rather than the private estimates. This indicates that the matrix \( E_t \) decreases with time and as it is observed in our numerical results in Figure 1, it converges to zero. One can also see that the strategies decrease as \( E_t \) decreases. Therefore, the strategy matrices \( L^i_t \) decrease as time passes and they converge to 0.5, which is the equilibrium of the game when \( E_t = 0 \).

![Strategy matrices](image)

**Fig. 1:** Strategy matrices \( L^i_t \) and quantities \( E^i_t \) for \( T = 10 \).

**B. Game vs Centralized LQG**

In this subsection, we have compared the total rewards per time obtained through the game by players for \( Q^1 = Q^2 = 1 \) with a scenario in which both actions are taken by a single decision maker and the sum of the two rewards are collected by her. We have done this comparison for different time horizons \( T \) and Figure 2 depicts the plot of the total rewards per time obtained, \( J^T \), in the two considered scenarios. We notice that players are doing worse compared to the centralized decision
maker, specifically as the time horizon increases. The reason is that in the game scenario, the uncertainty in predicting the average reward-to-go increases drastically as time horizon increases. The centralized decision maker, however, benefits from time horizon increasing and her total reward per time converges to one. The reason is that as time goes by, the estimation over \( V \) becomes better and better and the reward converges to the one in the complete information case.

C. Two Dimensional State and Scalar Action

In this part, we consider a two dimensional attribute vector for the project, i.e., \( V \) is a two dimensional vector. Each agent tries to be aligned with one element of the attribute vector while maintaining cooperation with the other agent. We can model this alignment and cooperation of agents with \( R_1^t(v, a_t) = a_1^t v(1) + a_2^t v(2) - (a_1^t)^2 \) and \( R_2^t(v, a_t) = a_1^t v(2) + a_1^t a_2^t - (a_2^t)^2 \). That is, we set \( R_1^t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & -1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \) and
\[
R_2^t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & -1 \end{bmatrix}.
\]

We also set \( \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \).

Case 1: The following linear equilibrium strategy matrices are derived for the full information case.

\[
\begin{align*}
L_1^1 &= \begin{bmatrix} 0.7224 & 0.2402 \\ 0.4858 & 0.0842 \end{bmatrix}, & L_1^2 &= \begin{bmatrix} 0.2402 & 0.7224 \\ 0.0842 & 0.4858 \end{bmatrix}, \\
L_2^1 &= \begin{bmatrix} 0.2874 & 0.0780 \\ 0.1619 & 0.0281 \end{bmatrix}, & L_2^2 &= \begin{bmatrix} 0.0780 & 0.2874 \\ 0.0281 & 0.1619 \end{bmatrix},
\end{align*}
\]

and \( \bar{m}_1^t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \). Also, similar to the scalar case, \( M_i^t \) strategy matrices do not play any roles here since \( f_i = 0 \). We see that if \( V \) is perfectly observed, each agent will align her effort with a weighted average of \( V(1) \) and \( V(2) \) with the element corresponding to that agent having twice the weight. Also, similar to the scalar case, myopic strategies are played.

Case 2: Consider the partial information scenario with \( Q^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \) and \( Q^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \). The following linear equilibrium strategy matrices are derived.

\[
\begin{align*}
L_1^1 &= \begin{bmatrix} 0.7198 & 0.4232 \\ 0.5071 & 0.2055 \end{bmatrix}, & L_1^2 &= \begin{bmatrix} 0.4232 & 0.7198 \\ 0.2055 & 0.5071 \end{bmatrix}, \\
L_2^1 &= \begin{bmatrix} 0.3196 & 0.1506 \\ 0.1690 & 0.0685 \end{bmatrix}, & L_2^2 &= \begin{bmatrix} 0.1506 & 0.3235 \\ 0.0685 & 0.1370 \end{bmatrix},
\end{align*}
\]

and \( \bar{m}_i^t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \). Similar to the scalar scenario, we observe that the value of the strategy matrices decrease with time and again, \( \bar{m}_i^t = 0 \) for all of the cases.

Case 3: If each agent fully observes her corresponding element of the state and partially observes the other one, i.e., \( Q^1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \) and \( Q^2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \), we have the following linear equilibrium strategy matrices.

\[
\begin{align*}
L_1^1 &= \begin{bmatrix} 0.3239 & 0.0561 \\ 0.6293 & 0.1619 \end{bmatrix}, & L_1^2 &= \begin{bmatrix} 0.0561 & 0.3239 \\ 0.1619 & 0.6293 \end{bmatrix}, \\
L_2^1 &= \begin{bmatrix} 0.3380 & 0.1370 \\ 0.1690 & 0.0685 \end{bmatrix}, & L_2^2 &= \begin{bmatrix} 0.1370 & 0.3380 \\ 0.0685 & 0.1690 \end{bmatrix},
\end{align*}
\]

and \( \bar{m}_i^t = 0 \) for \( t = 1, 2 \) and \( i = 1, 2 \). An intuitive reason of why the second element and the first element of the strategy matrices \( L_1^1 \) and \( L_2^1 \), respectively, are larger than the previous case is that the second element and the first element of \( E_1^t \) and \( E_2^t \), respectively, have increased.

VII. Conclusion

In this paper, we studied a dynamic game with asymmetric information and dependent types and we characterized the structured perfect Bayesian equilibria of the game. We also studied a special case of our model that was Linear Quadratic Gaussian (LQG) non-zero-sum game and we characterized linear structured perfect Bayesian equilibria for the game. One of the important points that we made in this paper was that due to the conditional independence of the private signals, the private belief chain stops at two or any other given number of steps.
APPENDIX A
PROOF OF LEMMA 1

\[
\begin{align*}
\pi_t(\xi_t|v) &= \mathbb{P}(\xi_t|v, a_{t:t-1}) \\
&= \frac{\int_{x_{1:t}} \mathbb{P}(x_{1:t}, \xi_t, a_{t:t-1}|v)}{\int_{x_{1:t}} \mathbb{P}(x_{1:t}, a_{t:t-1}|v)} \tag{27a} \\
&= \int_{x_{1:t}} \prod_{i \in N} Q^i_X(x^i_s|v, a_{s-1}) \mathbb{P}(a_s^i|x^i_{1:s}, a_{1:s-1}) \\
&= \int_{x_{1:t}} \prod_{i \in N} Q^i_X(x^i_t|v, a_{t-1}) \mathbb{P}(\xi^i_t|x^i_{1:t}, a_{1:t-1}) \tag{27b} \\
&= \prod_{i \in N} Q^i_X(x^i_t|v, a_{t-1}) \mathbb{P}(\xi^i_t|x^i_{1:t}, a_{1:t-1}) \subseteq \prod_{i \in N} \mathbb{P}(\xi^i_t|v, a_{1:t-1}) \tag{27c} \\
&= \prod_{i \in N} \mathbb{P}(\xi^i_t|v, a_{1:t-1}) / \mathbb{P}(a_{1:t-1}|v) = \prod_{i \in N} \pi_t(\xi^i_t|v) \tag{27d} \\
&= \prod_{i \in N} \pi_t(\xi^i_t|v). \tag{27e}
\end{align*}
\]

The second part of the theorem is similarly proved as follows.

\[
\begin{align*}
\mathbb{P}(x_{1:t}|v, a_{1:t-1}) &= \frac{\mathbb{P}(x_{1:t}, a_{t:t-1}|v)}{\mathbb{P}(a_{1:t-1}|v)} \tag{28a} \\
&= \prod_{i \in N} \prod_{s=1}^{t-1} Q^i_X(x^i_s|v, a_{s-1}) \mathbb{P}(a_s^i|x^i_{1:s}, a_{1:s-1}) \\
&= \int_{x_{1:t}} \prod_{i \in N} \prod_{s=1}^{t-1} Q^i_X(x^i_s|v, a_{s-1}) \mathbb{P}(a_s^i|x^i_{1:s}, a_{1:s-1}) \mathbb{P}(\xi^i_t|x^i_{1:t}, a_{1:t-1}) \tag{28b} \\
&= \prod_{i \in N} \prod_{s=1}^{t-1} Q^i_X(x^i_s|v, a_{s-1}) \mathbb{P}(a_s^i|x^i_{1:s}, a_{1:s-1}) \\
&= \prod_{i \in N} Q^i_X(x^i_t|v, a_{t-1}) \mathbb{P}(\xi^i_t|x^i_{1:t}, a_{1:t-1}) \tag{28c} \\
&= \prod_{i \in N} \mathbb{P}(\xi^i_t|v, a_{1:t-1}) / \mathbb{P}(a_{1:t-1}|v) = \prod_{i \in N} \pi_t(\xi^i_t|v) \tag{28d} \\
&= \prod_{i \in N} \pi_t(\xi^i_t|v). \tag{28e}
\end{align*}
\]

APPENDIX B
PROOF OF LEMMA 2

Using Bayes rule we have

\[
\begin{align*}
\xi^i_{t+1}(v) &= \mathbb{P}(v|x_{t+1}^i, a_{1:t}) \\
&= \frac{\mathbb{P}(v, x_{t+1}^i, a_{1:t} | x_{1:t}^i, a_{1:t-1})}{\mathbb{P}(a_{1:t-1} | x_{1:t}^i, a_{1:t-1})} \tag{29a} \\
&= \frac{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(v, x_{t+1}^i, a_{1:t}, \xi^i_{t+1}, \tilde{v} | x_{1:t}^i, a_{1:t-1})}{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}^{\xi^i_{t+1}}(\tilde{v}, x_{t+1}^i, a_{1:t}, \xi^i_{t+1}, \tilde{v} | x_{1:t}^i, a_{1:t-1})} \tag{29b} \\
&= \frac{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(v | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1})}{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1}) \mathbb{P}(\tilde{v} | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1})} \tag{29c} \\
&= \frac{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(v | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1})}{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1}) \mathbb{P}(\tilde{v} | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1})} \tag{29d} \\
&= \frac{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(v | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1})}{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1}) \mathbb{P}(\tilde{v} | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1})} \tag{29e} \\
&= \frac{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(v | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1})}{\int_{\xi^i_{t+1}, \tilde{v}} \mathbb{P}(\xi^i_{t+1} | x_{1:t}^i | a_{1:t-1}) \mathbb{P}(\tilde{v} | x_{t+1}^i, a_{1:t-1}) \mathbb{P}(\xi^i_{t+1} | \tilde{v}, a_{1:t-1})} \tag{29f}.
\end{align*}
\]

APPENDIX C
PROOF OF LEMMA 3

Using Bayes rule we have

\[
\begin{align*}
\pi_t(\xi_{t+1}^i|v) &= \mathbb{P}(\xi_{t+1}^i|v, a_{1:t}) \tag{30a} \\
&= \frac{\mathbb{P}(\xi_{t+1}^i, x_{t+1}^i, a_{1:t} | v, a_{1:t-1})}{\mathbb{P}(a_{1:t-1} | v, a_{1:t-1})} \tag{30b} \\
&= \frac{\int_{\xi_{t+1}^i, x_{t+1}^i} \mathbb{P}(\xi_{t+1}^i, x_{t+1}^i | v, a_{1:t-1})}{\int_{\xi_{t+1}^i} \mathbb{P}(\xi_{t+1}^i | v, a_{1:t-1})}.
\end{align*}
\]
\[
\int_{\xi_t} \mathbb{P}(\xi_{t+1} | v, a_{1:t-1}) \mathbb{P}(a_t | \xi_t, a_{1:t-1}) \mathbb{P}(x_{t+1} | v, a_t) \times 1_{F^*(\xi_t, \pi_i, \gamma^{-i}_{t-i}, a_t, x_{t+1})}^{\xi_{t+1}} = \int_{\xi_t} \mathbb{P}(\xi_{t+1} | v, a_{1:t-1}) \mathbb{P}(a_t | \xi_t, a_{1:t-1}) \mathbb{P}(x_{t+1} | v, a_t) \times 1_{\text{randomized strategies}}^{\xi_{t+1}}.
\]

\[
\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t) = \frac{1}{\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}}}. \int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t)
\]

\[
\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t) = \frac{1}{\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}}}. \int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t)
\]

\[
\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t) = \frac{1}{\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}}}. \int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t)
\]

\[
\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t) = \frac{1}{\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}}}. \int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t)
\]

\[
\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t) = \frac{1}{\int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}}}. \int_{\xi_t} \prod_{j \in N} \pi_j^{\xi_j(v)} \gamma_j^{a_j^{\xi_j(v)}} Q_X(x_{t+1} | v, a_t)
\]

\[
\text{APPENDIX D}
\]

**Proof of Theorem 1**

To prove the theorem, we show that if every player \(-i\) plays according to strategy \(\gamma_{t-i}^i = \theta_t^{-i}(\pi_i)\), the best response of player \(i\) is of the form \(\gamma_t^{ia} = \theta_t(\pi_t)\) and it is derived from the given fixed point equation. We show that if we fix the update rule of \(\pi_t\) to \(\pi_{t+1} = F_t(\pi_t, \gamma_t^i, a_t) = F_t(\pi_t, \theta_t(\pi_t), a_t)\) and assume that player \(i\) is forced to use these beliefs as her true beliefs, then she faces an MDP with state \((\pi_t, x_t, \xi_t)\), action \(a_t\) and instantaneous reward \(r_t(\pi_t, x_t, \xi_t, a_t) = \mathbb{E}[r_t(V, A_t) | \pi_t, x_t, \xi_t, a_t]\).

We first need to prove that the state \((\pi_t, x_t, \xi_t)\) evolves according to a controlled Markov process. Indeed,

\[
\mathbb{P}(\pi_{t+1}, x_{t+1}, \xi_{t+1} | \pi_{1:t}, x_{1:t}, \xi_{1:t}, a_{1:t}) = \int_{\xi_{t-i}, a_{t-1}, x_{t-i}} \pi_t^{-1}(\xi_t^{-i} | v, a_t) \theta_t^{-i}(\pi_t)(a_t^{-i}) Q_X(x_{t+1} | v, a_t)
\]

\[
= \mathbb{P}(\pi_{t+1}, x_{t+1}, \xi_{t+1} | \pi_t, x_t, \xi_t, a_t).
\]

The average instantaneous reward can now be written as

\[
\mathbb{E}[r_t^i(V, A_t)] = \mathbb{E}[\mathbb{E}[r_t^i(V, A_t) | \pi_t, x_t, \xi_t, a_t]].
\]

\[
\mathbb{E}[r_t^i(V, A_t) | \pi_t, x_t, \xi_t, a_t] = \int_{\xi_{t-i}} \mathbb{E}[r_t^i(V, a_t) | \xi_t^{-i} | \pi_t, x_t, \xi_t, a_t]
\]

\[
= \int_{\xi_{t-i}} \mathbb{E}[r_t^i(V, a_t) | \xi_t^{-i} | \pi_t, x_t, \xi_t, a_t]
\]

\[
\text{APPENDIX E}
\]

**Proof of Theorem 2**

Throughout this proof, the submatrices that are not explicitly specified are all zero matrices with appropriate dimensions.

In order to prove the theorem we will define a dynamical system from the viewpoint of a specific user \(i\) and show inductively that it is a Gauss Markov model. Gaussianity of both private and conditional public beliefs follows from KF-type arguments.
For each player \( i \in \mathcal{N} \), we define an unobserved state vector as
\[
\tilde{s}^i_t = \begin{bmatrix} v \ ; \ \hat{v}_{i-1}^i \end{bmatrix},
\]
(36a)
and an observation vector
\[
y^i_t = \begin{bmatrix} a_{i-1}^i - m_{i-1}^i \ ; \ x_t^i \end{bmatrix}.
\]
(36b)

We will show that the random vector \( s^i_t \) evolves according to a linear Gaussian process,
\[
s^i_{t+1} = A^i_t s^i_t + \begin{bmatrix} 0 \\
D^i_t \end{bmatrix} a_{t-1}^i + \begin{bmatrix} 0 \\
H^i_t \end{bmatrix} w_{t-1}^i + \begin{bmatrix} 0 \\
d^i_t \end{bmatrix}
\]
(37a)
\[
y^i_t = C^i_t s^i_t + \begin{bmatrix} 0 \\
I \end{bmatrix} w_t^i,
\]
(37b)
where
\[
A^i_t = \begin{bmatrix} \mathbf{I} & 0 \\
G_t^{-i} \end{bmatrix},
\]
(37c)
\[
C^i_t = \begin{bmatrix} 0 & \Sigma_i^{-1} \end{bmatrix}.
\]
(37d)

Note that \( (y^i_{t+1}, a^i_{t-1}) \) is a shifted version of \( h^i_t \). We prove the validity of (37) and the claim of the theorem using induction. In particular, Lemma 4 below is the induction basis and the subsequent Lemma 5 is the induction step. This concludes the proof of the theorem.

**Lemma 4:** The following are true.

(a) \( \xi^i_t \) is Gaussian \( N(\hat{v}^i_t, \Sigma_t^i) \), with \( \hat{v}^i_t = \Sigma_t(\Sigma_t + Q^i)^{-1}x^i_t \) and \( \Sigma_t^i = \Sigma_t - \Sigma_t(\Sigma_t + Q^i)^{-1}\Sigma_t \). Consequently the public belief \( \pi_t^i(\xi^i_t | v) \) reduces to \( \pi_t^i(\hat{v}^i_t | v) \).

(b) (37) holds for \( t = 1 \).

(c) The public belief \( \pi_t^i(\hat{v}^i_t | v) \) is Gaussian with mean \( E[\hat{v}^i_t | v] = E^i_t v + f^i_t \), with \( E^i_t = \Sigma_t(\Sigma_t + Q^i)^{-1} \) and variance \( \Sigma_t = (\Sigma_t + Q^i)^{-1}Q^i(\Sigma_t + Q^i)^{-1} \).

**Proof:**

(a) We have \( x^i_t = v + w^i_t \) and \( \xi^i_t = \mathbb{P}(v | x^i_t) \), so due to joint Gaussianity of \( V \) and \( X^i_t \) we have that \( \xi^i_t = N(\hat{v}^i_t, \Sigma_t^i) \), with mean
\[
\hat{v}^i_t = E[V | x^i_t],
\]
(38a)
\[
= E[V] + E[V X^i_t]E[X^i_t X^i_t]^{-1}(x^i_t - E[X^i_t])
\]
(38b)
\[
= \Sigma_t(\Sigma_t + Q^i)^{-1}x^i_t;
\]
(38c)
and covariance matrix
\[
\Sigma_t^i = \Sigma_t - \Sigma_t(\Sigma_t + Q^i)^{-1} \Sigma_t.
\]
(39)

As a result the only private information of user \( i \) relevant to other users is \( \hat{v}^i_t \) and the public belief \( \pi_t^i(\xi^i_t | v) \) can be reduced to \( \pi_t^i(\hat{v}^i_t | v) \).

(b) We have \( s^i_t = \begin{bmatrix} v \ ; \ 0 \end{bmatrix} \) and \( s^i_2 = \begin{bmatrix} v \ ; \ \hat{v}^i_1 \end{bmatrix} \). The first row of (37a) is evidently true. For the second row, using the result (from part (a)) \( \hat{v}^i_t = \Sigma_t(\Sigma_t + Q^i)^{-1}(v + w^i_t) \), we can derive \( G_t^{-i}, H_t, D_t \) and \( d_t \) as
\[
G_t^{-i} = \begin{bmatrix} \Sigma_t(\Sigma_t + Q^i)^{-1} & 0 \end{bmatrix}
\]
(40a)
\[
H_t = \Sigma_t(\Sigma_t + Q^i)^{-1}
\]
(40b)
\[
D_t = 0
\]
(40c)
\[
d_t = 0
\]
(40d)
where \( \Sigma_t(\Sigma_t + Q^i)^{-1} \) is the vertical concatenation of the matrices \( \Sigma_t(\Sigma_t + Q^i)^{-1} \) for \( j \in \mathcal{N} \).

(c) Since \( \hat{v}^i_t = \Sigma_t(\Sigma_t + Q^i)^{-1}(v + w^i_t) \) we deduce that \( \pi_t^i(\hat{v}^i_t | v) \) is Gaussian with mean \( E[\hat{v}^i_t | v] = \Sigma_t(\Sigma_t + Q^i)^{-1}v \) and covariance matrix \( \Sigma_t^i = \Sigma_t(\Sigma_t + Q^i)^{-1}Q_t^i(\Sigma_t + Q^i)^{-1} \).

**Lemma 5:** Assuming pure linear strategies of the form \( \gamma^i_j(a^i_t | \tilde{v}^i_t) = \delta(a^i_t - L^i_j \tilde{v}^i_t - m^i_t) \) for all \( j \in \mathcal{N} \), and assuming that (37) holds for \( t \leq k \) and \( E[\hat{v}^k_t | v, a^i_{1:k-1}] = E^i_t v + f^i_k \), the following are true.

(a) \( \xi^i_{k+1} \) is \( N(\hat{v}^i_{k+1}, \Sigma^i_{k+1}) \) with
\[
\hat{v}^i_{k+1} = G^i_{k+1} \begin{bmatrix} \hat{v}^i_k \\
x^i_{k+1} \end{bmatrix} + d^i_{k+1},
\]
(41)
where \( G^i_{k+1}, d^i_{k+1} \) and \( \Sigma^i_{k+1} \) can be publicly evaluated. Consequently, the public belief \( \pi^i_{k+1}(\xi^i_{k+1} | v) \) can be reduced to a belief \( \pi^i_{k+1}(\hat{v}^i_{k+1} | v) \).

(b) (37) holds for \( t = k + 1 \).

(c) The conditional public belief, \( \pi^i_{k+1}(\hat{v}^i_{k+1} | v) \), are Gaussian with mean \( E[\hat{v}^i_{k+1} | V, a^i_{k+1}] = E^i_{k+1} V + f^i_{k+1} \) and covariance matrix \( \Sigma^i_{k+1} \), where matrices \( E^i_{k+1} \) and \( \Sigma^i_{k+1} \) and vector \( f^i_{k+1} \) can be publicly evaluated.

**Proof:**

(a) We first show one important result from the lemma assumptions. Notice that due to conditional independence of \( x^i_t \)'s given \( v \) across time and players, and since \( \hat{v}^i_t \) is a function of \( x^i_{1:k} \) and \( a^i_{1:k-1} \), we have
\[
\hat{v}^i_{k+1} = \begin{bmatrix} \hat{v}^i_k \\
x^i_{k+1} \end{bmatrix},
\]
(42a)
\[
= E[V | x^i_{1:k}, a^i_{1:k-1}]
\]
(42b)
\[
= E[V | x^i_{1:k}, a^i_{1:k-1}]
\]
(42c)
\[
= E[V | x^i_{1:k}, a^i_{1:k-1}]
\]
(42d)
\[
= E[V | x^i_{1:k}, a^i_{1:k-1}] + f^i_{k+1}
\]
(42e)
\[
= E^i_{k+1} v + f^i_{k+1}.
\]
(42f)

By using the assumption that (37) holds for \( t = k \), we form a linear Gaussian model with partial observations and use Kalman filter results [15, Ch.7]. Consider equation (37) for \( t = k \). By using standard Kalman filter results [15, Ch.7], we know that the belief over the system states given the observations is Gaussian and therefore, the private belief \( \xi^i_k \) is \( N(\hat{v}^i_k, \Sigma^i_k) \). We denote \( E[S^i_{k+1} | y^i_{1:k+1}, a^i_{1:k}] \) and \( E[S^i_{k+1} | y^i_{1:k}, a^i_{1:k-1}] \) by \( s^i_{k+1 | k+1} \) and \( s^i_{k+1 | k} \), respectively. We have
\[
s^i_{k+1 | k+1} = E[S^i_{k+1} | x^i_{1:k+1}, a^i_{1:k}]
\]
(43a)
\[
= E[S^i_{k+1} | x^i_{1:k+1}, a^i_{1:k}]
\]
(43b)
\[
= E[S^i_{k+1} | x^i_{1:k+1}, a^i_{1:k}]
\]
(43c)
}\[
\]
Therefore,

\[
\begin{align*}
\hat{v}_{k+1}^{i} &= \hat{v}_{k}^{i} + (J_{k+1}^{i})_{\cdot \cdot}^{v, i} (y_{k+1}^{i} - C_{k+1}^{i} \hat{s}_{k+1}^{i} | y_{k}^{i}) \\ 
&= \hat{v}_{k}^{i} + (J_{k+1}^{i})_{\cdot \cdot}^{v, i} \left[ a_{k} - m_{k} - \mathcal{D}(L_{k}^{-1}) \hat{v}_{k-1}^{i} \right] \\
&= \hat{v}_{k}^{i} + (J_{k+1}^{i})_{\cdot \cdot}^{v, i} \left[ -\mathcal{D}(L_{k-i}) E_{v, i}^{i} \hat{v}_{k-1}^{i} \right] \\
&\quad + (J_{k+1}^{i})_{\cdot \cdot}^{v, a_{i}} (a_{k} - m_{k} - \mathcal{D}(L_{k-i}) f_{k-1}^{i}) \\
&= G_{k+1}^{i} \left[ \hat{v}_{k}^{i} x_{k+1}^{i} \right] + \hat{d}_{k+1}^{i},
\end{align*}
\] (44a)

where

\[
\begin{align*}
(G_{k+1}^{i}, x^{i}) &= (J_{k+1}^{i})_{\cdot \cdot}^{v, x^{i}} \\
(G_{k+1}^{i}, \cdot \cdot) &= I - (J_{k+1}^{i})_{\cdot \cdot}^{v, a_{i}} \mathcal{D}(L_{k-i}) E_{v, i}^{i} - (J_{k+1}^{i})_{\cdot \cdot}^{v, x^{i}} \\
d_{k+1}^{i} &= (J_{k+1}^{i})_{\cdot \cdot}^{v, a_{i}} (a_{k} - m_{k} - \mathcal{D}(L_{k-i}) f_{k-1}^{i}).
\end{align*}
\] (45a)

The matrix \( J_{k+1}^{i} \) and the covariance matrix of \( S_{k+1}^{l} \) conditioned on \( y_{j}^{i}, y_{j+1}^{i} \) denoted by \( \Sigma_{k+1}^{i} | y_{k}, y_{k+1}^{i} \) and \( \Sigma_{k+1}^{i} | y_{k}, y_{k+1}^{i} \), respectively, can be derived from the standard Kalman filter equations as follows

\[
\begin{align*}
\Sigma_{k+1}^{i} | y_{k}, y_{k+1}^{i} &= A_{i}^{i} \Sigma_{k}^{i} | y_{k} A_{i}^{i} + \left[ 0 \ 0 \right] \mathcal{D}(Q_{k}^{-1}) \left[ 0 \ H_{k}^{i} \right] \right)^{t} \\
J_{k+1}^{i} &= \Sigma_{k+1}^{i} | y_{k}, y_{k+1}^{i} C_{k+1}^{i} \left[ \Sigma_{k+1}^{i} | y_{k}, y_{k+1}^{i} C_{k+1}^{i} \right]^{t} \\
&\quad + \left[ 0 \ I \right] Q_{k}^{i} \left[ 0 \ I \right]^{t})^{-1} \\
\Sigma_{j+1}^{i} | y_{k}, y_{k+1}^{i} &= (I - J_{k+1}^{i} C_{k+1}^{i}) \Sigma_{j+1}^{i} | y_{k}, y_{k+1}^{i} \\
\Sigma_{j+1}^{i} | y_{k}, y_{k+1}^{i} &= \mathbb{E}[S_{j+1}^{i} | y_{k}, y_{k+1}^{i}] - \mathbb{E}[S_{j}^{i} X_{j}^{i}]^{t} \mathbb{E}[X_{j}^{i} X_{j}^{i}]^{-1} \mathbb{E}[S_{j}^{i} X_{j}^{i}]^{t} \\
&= \left[ \Sigma \ 0 \right] - \left[ \Sigma \ 0 \right] (\Sigma + Q_{k}^{i})^{-1} \left[ \Sigma \ 0 \right] \\
&= \left[ \Sigma - \Sigma(\Sigma + Q_{k}^{i})^{-1} \Sigma \ 0 \right].
\end{align*}
\] (46a)

Note that, for notational simplicity, we remove the time subscripts from submatrix notation, so that (\( J_{k+1}^{i} \)) denotes (\( J_{k+1}^{i} \)) for all players \( i \).

Finally, we have \( \Sigma_{i}^{i} = (\Sigma_{i+1}^{i})_{\cdot \cdot}^{v, a_{i}} \). Unlike \( \hat{v}_{k}^{i} \), which is part of the private information of player \( i \), the matrix \( \Sigma_{i}^{i} \) is a public quantity due to the independence of equation (46c) to the private observations of player \( i \).

(b) Equation (37a) is obvious for the first part of the state, \( v \). In order to prove the other parts of equation (37a) for \( t = k + 1 \), we consider the dynamic system (37) for each of the players \( -i \) for \( t = k \) and we write (44) for players \( -i \). Since \( x_{k+1}^{i} \) is not part of \( y_{k+1}^{i} \), we can substitute it by \( v + w_{k+1}^{i} \) and derive \( G_{k+1}^{i} \) for all \( j \in -i \) as

\[
\begin{align*}
(G_{k+1}^{j}, v) &= (J_{k+1}^{j})_{\cdot \cdot}^{v, x^{i}} \\
(G_{k+1}^{j}, \cdot \cdot) &= I - (J_{k+1}^{j})_{\cdot \cdot}^{v, a_{i}} \mathcal{D}(L_{k-i}) E_{v, i}^{i} - (J_{k+1}^{j})_{\cdot \cdot}^{v, x^{i}} \\
(D_{k+1}^{j}, v) &= (J_{k+1}^{j})_{\cdot \cdot}^{v, a_{i}} \\
(d_{k+1}^{j})_{\cdot \cdot}^{v, a_{i}} &= (J_{k+1}^{j})_{\cdot \cdot}^{v, a_{i}} (a_{k} - m_{k} - \mathcal{D}(L_{k-i}) f_{k-1}^{i}) \\
&\quad + (J_{k+1}^{j})_{\cdot \cdot}^{v, a_{i}} (m_{k} - L_{k} f_{k}) \\
H_{i}^{k+1} &= \mathcal{D}(L_{j}^{-1}) \hat{v}_{k-1}^{i}. 
\end{align*}
\] (47a)

The notation \( -ij \) means all of the players except \( i \) and \( j \). We have derived the matrices \( A_{k}^{i} \), \( D_{k}^{i} \), \( H_{k}^{i} \), and vector \( d_{k+1}^{i} \) and so (37) holds for \( t = k + 1 \).

(c) In order to show that the conditional public belief \( \pi_{k+1}^{i} (\hat{v}_{k+1}^{i} | v) \) is Gaussian, we consider a conditional Gaussian Markov model. Note that the conditional public belief is publicly measurable conditioned on \( V \). We use this fact to form a conditional model, where the observations are the conditions in the conditional public belief and we derive conditional Kalman filters. Using (37) for \( t \leq k + 1 \), we can construct the following linear Gaussian model for \( t \leq k + 1 \),

State:

\[
\hat{s}_{t} = \left[ \begin{array}{c} \hat{s}_{t-1} \\
\end{array} \right].
\] (48a)

State Evolution:

\[
\hat{s}_{t+1} = \mathbf{A}_{t} \hat{s}_{t} + \mathbf{H}_{t} w_{t} + \mathbf{d}_{t},
\] (48b)

Observation:

\[
\hat{y}_{t} = \left[ \begin{array}{c} \hat{y}_{t-1} \\
\end{array} \right] = \mathbf{C}_{t} \hat{s}_{t},
\] (48c)

where

\[
\begin{align*}
\hat{A}_{t} &= \left[ \begin{array}{c} I \ 0 \\
\end{array} \right] \\
\hat{C}_{t} &= \left[ \begin{array}{c} I \ 0 \\
\end{array} \right] \mathcal{D}(L_{t-1}^{-1}) \\
\end{align*}
\] (49a)

Using this conditional Gaussian Markov model, we can conclude that the conditional public beliefs \( \pi_{k+1}^{i} (\hat{v}_{k+1}^{i} | v) \) are Gaussian and by using Kalman filter results for \( t = k + 1 \), we can write

\[
\begin{align*}
\hat{s}_{k+2}^{i+1} &= \mathbb{E}[\hat{s}_{k+2}^{i+1} | \hat{y}_{k+1}^{i} | y_{k+1}] \\
&= \mathbb{E}[\hat{s}_{k+2}^{i+1} | y_{k+1}, a_{k}] \\
&= \hat{A}_{k+1} \hat{s}_{k+1}^{i+1} + \hat{A}_{k+1} \hat{J}_{k+1}^{i} \left( \hat{y}_{k+1}^{i} - \hat{C}_{k+1} \hat{s}_{k+1}^{i+1} \right) + \hat{d}_{k+1}^{i+1}.
\end{align*}
\] (50a)
Therefore,
\[
\mathbb{E}[\tilde{V}_{k+1}|v,a_{1:k}] = (\mathcal{G}_{k+1})_{:,v}v + (\mathcal{G}_{k+1})_{:,\tilde{v}}\mathbb{E}[\tilde{V}_k|v,a_{1:k-1}]
\]
\[
- (\tilde{A}_{k+1}\tilde{J}_{k+1})_{:,v}\mathcal{D}(L_k)\mathbb{E}[\tilde{V}_k|v,a_{1:k-1}]
\]
\[
+ (\tilde{A}_{k+1}\tilde{J}_{k+1})_{:,\tilde{v}}(a_k - m_k) + (\tilde{d}_{k+1})_{\tilde{v}}
\]
(51a)

Using the assumption of \(\mathbb{E}[\tilde{V}_k|v,a_{1:k-1}] = \mathbb{E}v_k + f_k\), we have the following
\[
\mathbb{E}[\tilde{V}_{k+1}|v,a_{1:k}] = (\mathcal{G}_{k+1})_{:,v}v + \mathbb{E}v_k + f_k
\]
\[
- (\tilde{A}_{k+1}\tilde{J}_{k+1})_{:,v}\mathcal{D}(L_k)\mathbb{E}v_k + f_k
\]
\[
+ (\tilde{A}_{k+1}\tilde{J}_{k+1})_{:,\tilde{v}}(a_k - m_k) + (\tilde{d}_{k+1})_{\tilde{v}}
\]
(52a)

where
\[
\mathbb{E}v_k = (\mathcal{G}_{k+1})_{:,v}v + ((\mathcal{G}_{k+1})_{:,\tilde{v}}\mathbb{E}v_k + f_k)
\]
\[
- (\tilde{A}_{k+1}\tilde{J}_{k+1})_{:,v}\mathcal{D}(L_k)\mathbb{E}v_k + f_k
\]
\[
+ (\tilde{A}_{k+1}\tilde{J}_{k+1})_{:,\tilde{v}}(a_k - m_k) + (\tilde{d}_{k+1})_{\tilde{v}}
\]
(53a)

and similar to part (a) of the proof, the covariance matrix of \(\tilde{S}_{k+1}\) conditioned on \(\tilde{y}_{1:k+1}\) and \(\tilde{y}_{1:k}\), denoted by \(\Sigma_{k+1|k+1}\) and \(\Sigma_{k+1|k}\), respectively, and the matrix \(\tilde{J}_{k+1}\) are derived from the following Kalman filter equations.

\[
\Sigma_{k+1|k} = \tilde{A}_k\Sigma_{k|k}\tilde{A}_k^T + \tilde{H}_k\mathcal{D}(Q)\tilde{H}_k
\]
\[
\tilde{J}_{k+1} = \Sigma_{k+1|k}\tilde{C}_{k+1} (\Sigma_{k+1|k}\tilde{C}_{k+1})^{-1}
\]
\[
\Sigma_{k+1|k} = \mathbb{E}[\tilde{S}_k\tilde{S}_k^T] - \mathbb{E}[\tilde{S}_k V_k^T](\mathbb{E}[V_k V_k^T])^{-1}\mathbb{E}[V_k V_k^T]
\]
\[
= 0.
\]
(54a)

Note that if we know \(\Sigma_{k+1|k}, \Sigma_{k+1|k}, \mathbb{E}v_k and f_k\), we can publically evaluate all of the other quantities defined in this proof for \(k + 1\) for a given strategy matrices \(L_k\) and vectors \(m_k\) and therefore, we can find \(\Sigma_{k+2|k+1}, \Sigma_{k+2|k+1}, \mathbb{E}v_k\) and \(f_k\). We can also find \(G_{k+1}^i\) and \(d_{k+1}^i\), which are used to update \(\tilde{v}_k^i\) to \(\tilde{v}_{k+1}^i\).

**APPENDIX F**

**PROOF OF THEOREM 3**

We show that for any \(t \in T\), if all players \(-i\) play according to the strategy \(\gamma_t^{-i}(a_t^{-i}|\tilde{v}_t^{-i}) = \delta(a_t^{-i} - \mathcal{D}(L_t^{-i})\tilde{v}_t^{-i} - m_t^{-i})\), where \(m_t^{-i} = M_t^{-i}f_t + \tilde{m}_t^{-i}\), and the strategies of players are linear in \(\tilde{v}_k\) for \(k < t\), player \(i\) faces an MDP with state \((\tilde{v}_t^i, \Sigma_i, E_t, f_t)\) and her best response is of the form \(\gamma_t^i(a_t^i|\tilde{v}_t^i) = \delta(a_t^i - \mathcal{D}(L_t^i)\tilde{v}_t^i - m_t^i)\), where \(m_t^i = M_t^i f_t + \tilde{m}_t^i\).

By using the results from Theorem 2, given the strategy profile \(\gamma_t^i, (\tilde{v}_t^i, \Sigma_i, E_t, f_t)\) forms a Markov chain. Notice that \(\tilde{V}_{t+1}, \mathbb{E}_{t+1}, f_{t+1}\) are updated by \(\gamma_t\) which is linear and therefore, all results from Theorem 2 hold.

**Lemma 6**: One can write the expected value of the instantaneous reward \(\tilde{R}_t^i = \mathbb{E}[r_i(V_t, A_t)|a_t^i, \tilde{v}_t, \Sigma_t, E_t, f_t]\) as

\[
\tilde{R}_t^i = \text{qd}(\tilde{R}_t^i; \begin{bmatrix} \tilde{v}_t^i \\ a_t^i \\ \tilde{v}_t^{-i} \\ f_t \end{bmatrix}) + \tilde{b}_t^i \begin{bmatrix} \tilde{v}_t^i \\ a_t^i \\ \tilde{v}_t^{-i} \end{bmatrix} + \tilde{c}_t^i,
\]
(55)

where \(\tilde{R}_t^i, \tilde{b}_t^i\) and \(\tilde{c}_t^i\) are constructed in the proof.

**Proof**: Since we assume all players \(-i\) play according to \(\gamma_t^{-i}, \) we have \(a_t^{-i} = \mathcal{D}(L_t^{-i})\tilde{v}_t^{-i} + M_t^{-i}f_t + \tilde{m}_t^{-i}\) and so the instantaneous reward can be rewritten as follows.

\[
r_t^i(v, a_t) = \text{qd}(\tilde{R}_t^i; \begin{bmatrix} v \\ a_t \\ \tilde{v}_t^{-i} \end{bmatrix})
\]
(56a)

\[
= \text{qd}(\tilde{R}_t^i, \begin{bmatrix} v \\ a_t \\ \tilde{v}_t^{-i} \end{bmatrix}) + \tilde{b}_t^i \begin{bmatrix} v \\ a_t \\ \tilde{v}_t^{-i} \end{bmatrix} + \tilde{c}_t^i
\]
(56b)

where

\[
\tilde{R}_t^i = \begin{bmatrix} I_{N_t+N_a} & 0 & 0 \\ 0 & \mathcal{D}(L_t^{-i}) & M_t^{-i} \end{bmatrix}' \tilde{I}_{2,i+1} \tilde{R}_t^i \tilde{I}_{2,i+1}
\]
(57a)

\[
\tilde{I}_{2,i+1} = \begin{bmatrix} I_{N_t+N_a} & 0 & 0 \\ 0 & \mathcal{D}(L_t^{-i}) & M_t^{-i} \end{bmatrix}
\]
(57b)

where \(I_k\) is the identity matrix with size \(k \times k\).

\[
\tilde{b}_t^i = 2 \left[ \begin{bmatrix} 0 \\ \tilde{v}_t^{-i} \end{bmatrix} \right] \tilde{I}_{2,i+1} \tilde{R}_t^i \tilde{I}_{2,i+1} + \begin{bmatrix} 0 \\ \tilde{v}_t^{-i} \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{v}_t^{-i} \end{bmatrix}
\]
(57d)

\[
\tilde{c}_t^i = 2 \left[ \begin{bmatrix} 0 \\ \tilde{v}_t^{-i} \end{bmatrix} \right] \tilde{I}_{2,i+1} \tilde{R}_t^i \tilde{I}_{2,i+1} + \begin{bmatrix} 0 \\ \tilde{v}_t^{-i} \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{v}_t^{-i} \end{bmatrix}
\]
(57e)

We can now calculate the expected value of \(R_t^i\) as follows.

\[
\tilde{R}_t^i = \text{qd}(\tilde{R}_t^i, \begin{bmatrix} \tilde{v}_t^i \\ a_t^i \\ \tilde{v}_t^{-i} \end{bmatrix}) + \text{tr}(\tilde{R}_t^i \Sigma_t^i) + \tilde{b}_t^i \begin{bmatrix} \tilde{v}_t^i \\ a_t^i \\ \tilde{v}_t^{-i} \end{bmatrix} + \tilde{c}_t^i
\]
(58)

where

\[
\Sigma_t^i = \begin{bmatrix} \Sigma_t^i & 0 & \Sigma_t^i E_t^{-i} \\ 0 & 0 & 0 \\ E_t^{-i} \Sigma_t^i & 0 & 0 \end{bmatrix}
\]
(59)
By using \( \tilde{v}_t^{i,-1} = E_t^{-i} \tilde{v}_t^i + f_t^{-i} \), we can derive the equations for \( \hat{R}_t^i \), \( \hat{b}_t^i \) and \( \hat{c}_t^i \).

\[
\hat{R}_t^i = \begin{bmatrix} I_{N_v} & 0 & 0 \\ 0 & I_{N_v} & 0 \\ E_t^{-i} & 0 & \hat{I}_t^{-i} \\ 0 & 0 & I_{N_v,N} \end{bmatrix} \begin{bmatrix} I_{N_v} & 0 \\ 0 & I_{N_v} \\ E_t^{-i} & 0 \\ 0 & 0 & I_{N_v,N} \end{bmatrix}' \begin{bmatrix} I_{N_v} & 0 \\ 0 & I_{N_v} \\ E_t^{-i} & 0 \\ 0 & 0 & I_{N_v,N} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{v}_t^i \\ \hat{v}_t^i \\ \hat{V}_t^{-i} \\ X_{t+1}^{-i} \end{bmatrix}
\]

(60a)

\[
\hat{b}_t^i = \begin{bmatrix} I_{N_v} & 0 & 0 \\ 0 & I_{N_v} & 0 \\ E_t^{-i} & 0 & \hat{I}_t^{-i} \\ 0 & 0 & I_{N_v,N} \end{bmatrix} \begin{bmatrix} \tilde{v}_t^i \\ \hat{v}_t^i \\ \hat{V}_t^{-i} \\ X_{t+1}^{-i} \end{bmatrix}
\]

(60b)

\[
\hat{c}_t^i = \text{tr}(\hat{R}_t^i \hat{\Sigma}_t^i) + \hat{c}_t^i
\]

(60c)

In the next lemma, we show that the reward-to-go at time \( t \) is a quadratic function of \( \tilde{v}_t^i \) and \( f_t \). We do not care about its functionality with respect to \( \Sigma_t \) and \( E_t \) due to two reasons. First, they are part of the public part of the history and are not parameters of the partial strategies \( \gamma \). Second, they are not controlled by the actions. As we will see in the proof of this lemma, \( Z_t^i \), \( z_t^i \), and \( o_t \) are functions of \( \Sigma_t \) and \( E_t \).

Lemma 7: We have the following equation for the reward-to-go function, \( J_t^i(\tilde{v}_t^i; \Sigma_t, E_t, f_t) = \text{qd}(Z_t^i; \begin{bmatrix} \tilde{v}_t^i \\ f_t \end{bmatrix}) + z_t^i \begin{bmatrix} \tilde{v}_t^i \\ f_t \end{bmatrix} + o_t^i \). Note that the above equation only highlights the functionality of the reward-to-go with respect to \( \tilde{v}_t^i \) and \( f_t \). We do not care about its functionality with respect to \( \Sigma_t \) and \( E_t \).

Proof: We prove the lemma by backward induction. For \( T + 1 \), we have \( J_{T+1}^i(\tilde{v}_{T+1}^i; \Sigma_{T+1}, E_{T+1}, f_{T+1}) = 0 \) and by setting \( Z_{T+1}^i = 0, z_{T+1}^i = 0, o_{T+1}^i = 0 \), the equation holds.

Assume that the lemma holds for \( t + 1 \). We will show that it will also hold for \( t \).

\[
J_t^i(\tilde{v}_t^i; \Sigma_t, E_t, f_t) = \max \{|\tilde{v}_t^i|; A_t \}
\]

(61a)

\[
J_t^i(\tilde{v}_t^i; \Sigma_t, E_t, f_t) = \max \{\text{qd}(\hat{R}_t^i; \begin{bmatrix} \tilde{v}_t^i \\ a_t^i \end{bmatrix}), \hat{b}_t^i \begin{bmatrix} \tilde{v}_t^i \\ a_t^i \end{bmatrix}, \hat{c}_t^i \}
\]

(61b)

First consider the \( J_t^i \) part.

\[
\text{E}^{-i} \{\text{qd}(Z_{t+1}^i; \begin{bmatrix} \tilde{V}_t^{-i} \\ X_{t+1}^{-i} \end{bmatrix}) + z_{t+1}^i \begin{bmatrix} \tilde{V}_t^{-i} \\ X_{t+1}^{-i} \end{bmatrix} + o_{t+1}^i \begin{bmatrix} \tilde{V}_t^{-i} \\ X_{t+1}^{-i} \end{bmatrix} \}
\]

(61c)

and we have

\[
\hat{Z}_{t+1}^i = T_{t+1}^i \hat{G}_{t+1}^i \hat{Z}_{t+1}^i \hat{G}_{t+1}^i T_{t+1}^i
\]

(65a)

\[
T_{t+1}^i = \begin{bmatrix} I_{N_v} & 0 & 0 \\ 0 & I_{N_v} & 0 \\ 0 & 0 & I_{N_v,N} \end{bmatrix}^{-1} \begin{bmatrix} I_{N_v} & 0 & 0 \\ 0 & I_{N_v} & 0 \\ 0 & 0 & I_{N_v,N} \end{bmatrix} \begin{bmatrix} I_{N_v} & 0 & 0 \\ 0 & I_{N_v} & 0 \\ 0 & 0 & I_{N_v,N} \end{bmatrix}
\]

(65b)

\[
\hat{G}_{t+1}^i = \hat{G}_{t+1}^i + z_{t+1}^i \begin{bmatrix} \tilde{V}_t^{-i} \\ X_{t+1}^{-i} \end{bmatrix} + o_{t+1}^i \hat{G}_{t+1}^i T_{t+1}^i
\]

(65c)

where

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi}
\]

(64a)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64b)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi}
\]

(64c)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64d)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64e)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64f)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64g)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64h)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64i)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64j)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64k)

\[
(\hat{G}_{t+1}^i)_{\psi,\psi} = (\hat{G}_{t+1}^{-i})_{\psi,\psi} \hat{D}(L_t^{-i})
\]

(64l)
\[ (\Sigma_{t+1}^i)_{x}^{x'} \hat{v}_{x-x'} = \begin{bmatrix} (\Sigma_{t+1}^i)_{x'}^{x} \hat{v}_{x-x'} & E_{t+1}^{i} \Sigma_{i}^{\prime} \end{bmatrix}. \] (65f)

Therefore, one can write the expected reward-to-go as follows.

\[ J_{t}^{i}(\hat{v}_{t}^{i}, \Sigma_{t}^{i}, E_{t}, f_{t}) = \max \{ qd(\hat{R}_{t}^{i} - \hat{v}_{t}^{i}) \}_{a_{t}^{i}} + \hat{b}_{t}^{i} \left[ \begin{array}{c} \hat{v}_{t}^{i} \\ f_{t} \end{array} \right] + \hat{c}_{t}^{i} + qd(\hat{Z}_{t+1}^{i} - \hat{v}_{t}^{i}) \left[ \begin{array}{c} \hat{v}_{t}^{i} \\ f_{t} \end{array} \right] + \hat{z}_{t+1}^{i} \left[ \begin{array}{c} \hat{v}_{t}^{i} \\ f_{t} \end{array} \right] + \hat{o}_{t+1}^{i} \} \]

\[ \Rightarrow a_{t}^{i} = -(\hat{R}_{t}^{i} + \hat{Z}_{t+1}^{i})^{-1}((\hat{R}_{t}^{i} + \hat{Z}_{t+1}^{i})a_{t}^{i} \hat{v}_{t}^{i}f_{t} + 1/2 (\hat{b}_{t}^{i} + \hat{z}_{t+1}^{i})a_{t}^{i}) \] (66a)

\[ m_{t}^{i} = -(\hat{R}_{t}^{i} + \hat{Z}_{t+1}^{i})^{-1}(\hat{R}_{t}^{i} + \hat{Z}_{t+1}^{i})a_{t}^{i}f_{t} \] (66b)

Note that in order to derive the \( \gamma_{t+1}^{i} \) strategy matrix and vector, \( L_{t}^{i} \) and \( m_{t}^{i} \), we need to know \( L_{t-1}^{i} \) and \( m_{t-1}^{i} \). Clearly, the same is true for calculating \( L_{t-1}^{i} \) and \( m_{t-1}^{i} \). On the other hand, some of the quantities used in the proof, like \( G_{i+1}^{i} \), require \( L_{t}^{i} \) and \( m_{t}^{i} \) to be evaluated. Therefore, we have a fixed point equation over \( L_{t}^{i} \) and \( m_{t}^{i} \).

Note that we have such linear solution only if the matrix \( (\hat{R}_{t}^{i} + \hat{Z}_{t+1}^{i})a_{t}^{i}a_{t}^{i} \) is invertible and negative semi definite for all \( i \in \mathcal{N} \).

We conclude the proof of the theorem by noting that in Lemma 7, we proved that the reward to go is a quadratic function of \( \left[ \begin{array}{c} \hat{v}_{t}^{i} \\ f_{t} \end{array} \right] \) and as a result throughout the proof, we derived equation (68) for the best response strategy of player \( i \). Equation (68) indicates that the linear strategies in terms of \( \left[ \begin{array}{c} \hat{v}_{t}^{i} \\ f_{t} \end{array} \right] \) form equilibria of the game and the theorem is proved.

REFERENCES

[1] N. Heydaribeni and A. Anastasopoulos, “Linear equilibria for dynamic LQG games with asymmetric information and dependent types,” IEEE Conference on Decision and Control (CDC), 2019.

[2] M. J. Osborne and A. Rubinstein, *A course in game theory.* MIT press, 1994.

[3] D. Fudenberg and J. Tirole, *Game theory.* MIT press, 1991.

[4] J. Watson, “A general, practicable definition of perfect bayesian equilibrium,” unpublished draft, 2017.

[5] E. Maskin and J. Tirole, “Markov perfect equilibrium: I. observable actions,” Journal of Economic Theory, vol. 100, no. 2, pp. 191-219, 2001.

[6] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, “Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition,” IEEE Trans. Automatic Control, vol. 62, no. 1, pp. 222–237, Jan 2017.

[7] D. Vasal, A. Sinha, and A. Anastasopoulos, “A systematic process for evaluating structured perfect bayesian equilibria in dynamic games with asymmetric information,” IEEE Transactions on Automatic Control, vol. 64, no. 1, pp. 81–96, 2018.

[8] A. Mahajan and A. Nayyar, “Sufficient statistics for linear control strategies in decentralized systems with partial history sharing,” IEEE Transactions on Automatic Control, vol. 60, no. 8, pp. 2046–2056, 2015.

[9] S. Yüksel, “Stochastic nestedness and the belief sharing information pattern,” IEEE Transactions on Automatic Control, vol. 54, no. 12, pp. 2773–2786, 2009.

[10] H. Tavafoghi, Y. Ouyang, and D. Teneketzis, “A unified approach to dynamic decision problems with asymmetric information-part ii: Strategic agents,” arXiv preprint arXiv:1812.01132, 2018.

[11] D. Vasal and A. Anastasopoulos, “Signaling equilibria for dynamic LQG games with asymmetric information,” in Proc. IEEE Conf. on Decision and Control, Dec. 2016, pp. 6901–6908.

[12] D. Abreu, D. Pearce, and E. Stacchetti, “Toward a theory of discounted repeated games with imperfect monitoring,” Econometrica: Journal of the Econometric Society, pp. 1041–1063, 1990.

[13] S. Bhichchandani, D. Hirshleifer, and I. Welch, “A theory of fads, fashion, custom, and cultural change as informational cascades,” Journal of Political Economy, pp. 992–1026, 1992.

[14] Y-C. Ho et al., “Team decision theory and information structures in optimal control problems-part i,” IEEE Transactions on Automatic Control, vol. 17, no. 1, pp. 15–22, 1972.

[15] P. R. Kumar and P. Varaiya, Stochastic systems: estimation, identification, and adaptive control. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[16] H. S. Witsenhausen, “A counterexample in stochastic optimum control,” SIAM Journal on Control, vol. 6, no. 1, pp. 131–147, 1968.

[17] T. Başar, “Two-criteria LQG decision problems with one-step delay observation sharing pattern,” Information and Control, vol. 38, no. 1, pp. 21–50, 1978.

[18] E. Altman, V. Kambhely, and A. Silva, “Stochastic games with one step delay sharing information pattern with application to power control,” in 2009 International Conference on Game Theory for Networks. IEEE, 2009, pp. 124–129.
[19] I. Bistritz and A. Anastasopoulos, “Characterizing non-myopic information cascades in Bayesian learning,” in Proc. IEEE Conf. on Decision and Control, Miami Beach, FL, July 2018.

[20] I. Bistritz, N. Heydaribeni, and A. Anastasopoulos, “Do Informational Cascades Happen with Non-myopic Agents?” arXiv preprint arXiv:1905.01327, 2019.

[21] N. Heydaribeni, I. Bistritz, and A. Anastasopoulos, “Informational cascades can be avoided with non-myopic agents,” 57th Annual Allerton Conference, 2019.

[22] A. Gupta, A. Nayyar, C. Langbort, and T. Başar, “Common information based Markov perfect equilibria for linear-Gaussian games with asymmetric information,” SIAM Journal on Control and Optimization, vol. 52, no. 5, pp. 3228–3260, 2014.

[23] V. P. Crawford and J. Sobel, “Strategic information transmission.” Econometrica: Journal of the Econometric Society, pp. 1431–1451, 1982.

[24] F. Farokhi, A. M. Teixeira, and C. Langbort, “Gaussian cheap talk game with quadratic cost functions: When herding between strategic senders is a virtue,” in 2014 American Control Conference. IEEE, 2014, pp. 2267–2272.

[25] E. Kamenica and M. Gentzkow, “Bayesian persuasion.” American Economic Review, vol. 101, no. 6, pp. 2590–2615, 2011.

[26] M. O. Sayin and T. Başar, “Dynamic information disclosure for deception,” in 2018 IEEE Conference on Decision and Control (CDC). IEEE, 2018, pp. 1110–1117.

Nasimeh Heydaribeni received her B.S. and M.S. degrees in Electrical Engineering at Sharif University of Technology, Tehran, Iran, in 2015 and 2017, respectively. She is currently pursuing her Ph.D. degree in Electrical Engineering and Computer Science at University of Michigan, Ann Arbor. Her research interests are game theory and its applications in networked systems with emphasis on mechanism design, dynamic games with asymmetric information and information design.

Achilleas Anastasopoulos (S’97-M’99-SM’13) was born in Athens, Greece in 1971. He received the Diploma in Electrical Engineering from the National Technical University of Athens, Greece in 1993, and the M.S. and Ph.D. degrees in Electrical Engineering from University of Southern California in 1994 and 1999, respectively. He is currently an Associate Professor at the University of Michigan, Ann Arbor, Department of Electrical Engineering and Computer Science.

His research interests lie in the general area of communication and information theory, with emphasis in channel coding and multi-user channels; control theory with emphasis in decentralized stochastic control and its connections to communications and information theoretic problems; analysis of dynamic games and mechanism design for resource allocation on networked systems.

He is the co-author of the book Iterative Detection: Adaptivity, Complexity Reduction, and Applications, (Reading, MA: Kluwer Academic, 2001).

Dr. Anastasopoulos is the recipient of the “Myronis Fellowship” in 1996 from the Graduate School at the University of Southern California, the NSF CAREER Award in 2004, and was a co-author for the paper that received the best student paper award in ISIT 2009. He served as a technical program committee member for ICC 2003, 2015–2018; Globecom 2004, 2012; VTC 2007, 2014, 2015; ISIT 2015, SPAWC 2018, and is currently serving as the TPC co-Chair for the Communication Theory Symposium, ICC’21. He was an associate editor for the IEEE Transactions on Communications in 2003–2008.