

Sequential decomposition of repeated games with asymmetric information and dependent states

Deepanshu Vasal

Abstract—We consider a finite horizon repeated game with N selfish players who observe their types privately and take actions, which are publicly observed. Their actions and types jointly determine their instantaneous rewards. In each period, players jointly observe actions of each other with delay 1, and private observations of the state of the system, and get an instantaneous reward which is a function of the state and everyone’s actions. The players’ types are static and are potentially correlated among players. An appropriate notion of equilibrium for such games is Perfect Bayesian Equilibrium (PBE) which consists of a strategy and a belief profile of the players which is coupled across time and as a result, the complexity of finding such equilibria grows double-exponentially in time. We present a sequential decomposition methodology to compute structured perfect Bayesian equilibria (SPBE) of this game, introduced in [1], where equilibrium policy of a player is a function of a common and a private belief state. This methodology computes SPBE in linear time. In general, the SPBE of the game problem exhibit signaling behavior, i.e. players’ actions reveal part of their private information that is payoff relevant to other players.

I. INTRODUCTION

Information asymmetry among strategic agents is an important topic, which has seen some very influential works such as [2] and [3]. Akerlof in [2] and Spence in [3] modeled a market of cars and a job market, respectively, as instances of information asymmetry in a game, and show interesting behavior of strategic agents derived from these models. Specifically, Akerlof in [2] showed that in a market of cars, where the quality of car is known only to the seller, lower prices can drive out good cars from the market. Spence in [3] showed that in equilibrium in a job market, a candidate can ‘signal’ her higher productivity to a potential employer by opting for higher education credentials. While these works showed very interesting and relevant phenomena for static information asymmetry, in the real world however, there exists many such, and even more complicated decision making scenarios which involves strategic decision makers with *dynamically evolving* information asymmetry. Some instances of such systems include: (a) in cyber-physical systems, many cyber and physical devices are connected to each other which have different information and they make a decision to optimize their performance objectives; (b) in a wind energy market a wind energy producer observes its own wind production privately and publicly observes the output of other producers which also determine the prices, and its objective is to generate output that maximizes its revenue; (c) in a social network, people have private opinions about a topic and also publicly observe actions of others, based on which they make a decision to maximize their utility. All such scenarios can be modeled as a dynamic game

Deepanshu Vasal is with the Department of Electrical and Computer Engineering, and with Simons Center for Mathematics of Networks, University of Texas, Austin, TX, USA. Email: dvasal at utexas.edu.

This research was funded by Department of Defense #W911NF1510225 to The University of Texas at Austin.

This paper has originally appeared on arxiv.org on January 10, 2018 as working paper 1801.03459.

of asymmetric information¹ [4]. Such problems are gaining more interest with applications such as alpha-go by Deepmind [5] for the symmetric information game Go, and for asymmetric/imperfect information games such as Texas Hold’em in [6].

Dynamical systems with strategic players are modeled as dynamic stochastic games, introduced by Shapley in [7]. Discrete-time dynamic games with Markovian structure have been studied extensively to model many practical applications, in engineering as well as economics literature [8], [9]. In dynamic games with perfect and symmetric information, subgame perfect equilibrium (SPE) is an appropriate equilibrium concept and there exists a backward recursive methodology to find all the SPEs of these games (see [10]–[12] for a more elaborate discussion). Maskin and Tirole in [13] introduced the concept of Markov perfect equilibrium (MPE) for dynamic games with symmetric information, where equilibrium strategies are dependent on some payoff relevant Markovian state of the system, rather than on the entire history. In dynamic games of asymmetric information, there are strategic players who are affected by an underlying process that is dynamically evolving, and the players make asymmetric observations about that process. In such games and more generally in any dynamic multi-agent decision problem with asymmetric information, a player’s action not just either *explore* or *exploit* the system², as it happens in a single agent problem, but also *signal* i.e. reveal part of its private information to the other players that is payoff relevant to them.³ Some appropriate notions of equilibrium for such games is Perfect Bayesian Equilibrium (PBE) or Sequential Equilibrium (SE) [4], [10] which involve an equilibrium strategy profile and an equilibrium belief profile of all the players, among other refinements. In these equilibria, the equilibrium strategies and beliefs are coupled together through a joint fixed-point equation in the space of strategies and beliefs for all players and for all histories of the game.⁴ Since the history of such games grows exponentially, the complexity of finding equilibria of such game grows double exponentially in time, rendering such problems intractable. We refer the reader to the Introduction section of [14] for a thorough introduction and a brief literature survey on dynamic games with asymmetric information.

Recently, there have been a number of results on finding an information state for different classes of such games that decomposes these games across time (in an analogous way a dynamic program decomposes a dynamic optimization problem), and thus reduces the complexity of finding these equilibria from double-exponential to linear in time. Authors in [14], [15], and independently, authors in [16], [17], presented such a sequential decomposition for games to find structured perfect Bayesian equilibrium (SPBE) and common information based perfect Bayesian equilibrium (CIB-

¹Sometimes also referred to as dynamic games of incomplete/imperfect information.

²Exploitation refers to making a decision based on whereas exploration refers to taking action that improves the current estimate of the state of the system even at some cost in the present, but that improves future reward.

³Equivalently, signaling occurs in such decision problems if players’ beliefs on a payoff relevant state are strategy dependent i.e. they depend on the strategies of one or more players.

⁴In comparison, Nash equilibrium for a static game is a fixed-point equation in the space of probability measures on possible actions of the players [10]

(PBE), respectively, where each player has a type or a state that evolves (conditionally) independently of other players' types in a Markovian way. Authors in [18] extended those results to LQG games (i.e. with linear state update, quadratic instantaneous costs and Gaussian random variables), and to games with conditionally independent hidden Markovian types in [19], where instead of perfectly observing its own type, each player makes independent noisy observations about it, respectively. Authors in [20] considered such dynamic game with a system state and delayed information sharing pattern where each players learns every agents private observations and actions with delay of d -steps. Authors in [17, ch. 5] generalized that dynamic game with a Markovian state where players make a common and private observations of the state of the system, where these observations are conditionally independent among players, conditioned on the current state and previous action. In this paper, we consider a model that adds to this literature, where there exists players' types that are static and *arbitrary correlated* among the players.

In this paper, similar to [14], we present a backward-forward methodology to compute SPBE of the game, where each player's strategy is a function of the same common information belief state and private state. These equilibria of the game are analogous to Markov Perfect equilibria (MPE) [13] of symmetric information games.

The paper is structured as follows. In Section II, we present the general model. In Section III, we define information belief states, which act as the motivation for the structure of the equilibrium policies. In Section IV, present the backward-forward methodology to compute SPBE of the game in linear time. We discuss some remarks in Section V. We conclude in section VI.

A. Notation

We use uppercase letters for random variables and lowercase for their realizations. For any variable, subscripts represent time indices and superscripts represent player identities. We use notation $-i$ to represent all players other than player i i.e. $-i = \{1, 2, \dots, i-1, i+1, \dots, N\}$. We use notation $a_{t:t'}$ to represent vector $(a_t, a_{t+1}, \dots, a_{t'})$ when $t' \geq t$ or an empty vector if $t' < t$. We use a_t^{-i} to mean $(a_t^1, a_t^2, \dots, a_t^{i-1}, a_t^{i+1}, \dots, a_t^N)$. We remove superscripts or subscripts if we want to represent the whole vector, for example a_t represents (a_t^1, \dots, a_t^N) . In a similar vein, for any collection of finite sets $(\mathcal{X}^i)_{i \in \mathcal{N}}$, we denote $\times_{i=1}^N \mathcal{X}^i$ by \mathcal{X} . We denote the indicator function of any set A by $I_A(\cdot)$. For any finite set \mathcal{S} , $\Delta(\mathcal{S})$ represents space of probability measures on \mathcal{S} and $|\mathcal{S}|$ represents its cardinality. We denote by P^g (or E^g) the probability measure generated by (or expectation with respect to) strategy profile g . We denote the set of real numbers by \mathbb{R} . For a probabilistic strategy profile of players $(\beta_t^i)_{i \in \mathcal{N}}$ where probability of action a_t^i conditioned on $(a_{1:t-1}, x^i)$ is given by $\beta_t^i(a_t^i | a_{1:t-1}, x^i)$, we use the short hand notation $\beta_t^{-i}(a_t^{-i} | a_{1:t-1}, x^{-i})$ to represent $\prod_{j \neq i} \beta_t^j(a_t^j | a_{1:t-1}, x^j)$. All equalities and inequalities involving random variables are to be interpreted in *a.s.* sense.

II. GENERAL MODEL

We consider a discrete-time dynamical system with N selfish players in the set $\mathcal{N} \triangleq \{1, 2, \dots, N\}$. We consider finite horizon $\mathcal{T} \triangleq \{1, 2, \dots, T\}$ with perfect recall. The system state is $x = (x^1, \dots, x^N)$, where $x \in \mathcal{X}$ is distributed as

$$P(x) = Q^x(x) \quad (1)$$

where $a_t = (a_t^1, \dots, a_t^N)$, and a_t^i is the action taken by player i at time t . At start of period t , players jointly observe a_{t-1} , and make private observations x , where $x^i \in \mathcal{X}^i$ is the private observation of player $i, \forall t \in 1, \dots, T$,

Player i takes action $a_t^i \in \mathcal{A}^i$ at time t upon observing $a_{1:t-1}$ which is common information among players, and x^i which is player i 's private information. The sets $\mathcal{A}^i, \mathcal{X}^i$ are assumed to be

finite. Let $g^i = (g_t^i)_t$ be a probabilistic strategy of player i where $g_t^i : \mathcal{A}^{t-1} \times (\mathcal{X}^i) \rightarrow \Delta(\mathcal{A}^i)$ such that player i plays action a_t^i according to $A_t^i \sim g_t^i(\cdot | a_{1:t-1}, x^i)$. Let $g := (g^i)_{i \in \mathcal{N}}$ be a strategy profile of all players. At the end of interval t , player i gets an instantaneous reward $R_t^i(X, A_t)$. The objective of player i is to maximize its total expected reward in a *perfect* sense

$$J^{i,g} := \mathbb{E}^g \left[\sum_{t=1}^T R_t^i(X, A_t) \right]. \quad (2)$$

III. STRUCTURAL RESULTS

The problem described in previous section is a repeated game with asymmetric information where a player's strategy is of the form $a_t^i \sim g_t^i(\cdot | a_{1:t-1}, x^i)$. Equivalent of such decentralized problems when players are cooperative and have the same objective can not be solved using classical tools from the theory of Markov Decision Processes (MDP) [21]. However, there exists some key ideas in the literature such as agent-by-agent approach [22] and common information approach [23] that present structural results of the optimum policies for a class of such systems with non-classical information structure and cooperative players. Such techniques have been used in dynamic team problems such as [24]–[26]. For the problem described in previous section, we use an approach inspired by [23].

For every agent $i \in \mathcal{N}$ and any policy profile g , let $\pi_t \in \Delta(\mathcal{X})$ be a common belief on x conditioned on the common information $a_{1:t-1}$, defined as follows.

$$\pi_t(x) := P^g(x = x | a_{1:t-1}) \quad (3)$$

A. Common agent approach

An alternative way to view the problem is as follows. As is done in common information approach [23], at time t , a fictitious common agent observes the common information $a_{1:t-1}$ and generates prescription functions $\gamma_t = (\gamma_t^i)_{i \in \mathcal{N}} = \psi_t[a_{1:t-1}]$. Player i uses these prescription functions γ_t^i to operate on its private information x^i to produce its action a_t^i , i.e. $\gamma_t^i : x^i \rightarrow \Delta(\mathcal{A}^i)$ and $a_t^i \sim \gamma_t^i(\cdot | x^i) = \psi[a_{1:t-1}](\cdot | x^i)$. It is easy to see that for any g policy profile of the players there exists an equivalent ψ profile of the common agent (and vice versa) that generates the same control actions for every realization of the information of the players.

We define a special type of common agent's policy as follows. We call a common agent's policy be of type θ if the common agent observes the common belief π_t derived from the common observation $a_{1:t-1}$, and generates prescription functions $\gamma_t = (\gamma_t^i)_{i \in \mathcal{N}} = \theta_t[\pi_t]$. Player i uses these prescription function γ_t^i to operate on its private type x^i to produce its action a_t^i , i.e. $\gamma_t^i : \Delta(\mathcal{X}) \rightarrow \Delta(\mathcal{A}^i)$ and $a_t^i \sim \gamma_t^i(\cdot | x^i) = \theta[\pi_t](\cdot | x^i)$. Equivalently, we call a common agent's policy be of type θ if for all $i \in \mathcal{N}$ and for all time t , player i 's action a_t^i depends on its information $(a_{1:t-1}, x^i)$ through the belief states $\pi_t(x) = P^\theta(x = x | a_{1:t-1})$ and x^i .

In the next lemma we show that for any given θ policy, the belief states π_t can be updated recursively as follows. Let $\pi_1(x) := Q^x(x)$.

Lemma 1: For any given policy of type θ , there exists update functions F_t , independent of θ , such that

$$\pi_{t+1} = F_t(\pi_t, \gamma_t, a_t). \quad (4)$$

Proof: Please see Appendix I. ■

IV. REPEATED GAMES WITH ASYMMETRIC INFORMATION

A. Solution concept: PBE

We introduce perfect Bayesian equilibrium (PBE) as an appropriate equilibrium concept for the game considered. Any history of this

game at which players take action is of the form $h_t = (x, a_{1:t-1})$. Let \mathcal{H}_t be the set of such histories at time t . At any time t player i observes $h_t^i = (x^i, a_{1:t-1})$ and all players together observe $h_t^c = (a_{1:t-1})$ as common history. Let \mathcal{H}_t^i be the set of observed histories of player i at time t and \mathcal{H}_t^c be the set of common histories at time t . An appropriate concept of equilibrium for such games is PBE [11] which consists of a pair (β^*, μ^*) of strategy profile $\beta^* = (\beta_t^{*,i})_{t \in \mathcal{T}, i \in \mathcal{N}}$ where $\beta_t^{*,i} : \mathcal{H}_t^i \rightarrow \Delta(\mathcal{A}^i)$ and a belief profile $\mu^* = (\mu_t^{*,i})_{t \in \mathcal{T}, i \in \mathcal{N}}$ where $\mu_t^{*,i} : \mathcal{H}_t^i \rightarrow \Delta(\mathcal{X}^i)$ that satisfy sequential rationality so that $\forall i \in \mathcal{N}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i, \beta_t^i$

$$\begin{aligned} & \mathbb{E}^{(\beta^{*,i}, \beta^{*,i}, \mu^*)} \left\{ \sum_{n=t}^T R_n^i(X, A_n) | h_t^i \right\} \\ & \geq \mathbb{E}^{(\beta^i, \beta^{*,i}, \mu^*)} \left\{ \sum_{n=t}^T R_n^i(X, A_n) | h_t^i \right\}, \end{aligned} \quad (5)$$

and the beliefs are updated using Bayes' rule, whenever possible.

In this paper, we first define a common belief π_t^* as a belief on x . Then player i derives its equilibrium belief $\mu_t^{*,i}$ on (x^{-i}) by conditioning π_t^* on its private type x^i i.e. $\mu_t^{*,i}(x^{-i}) = \pi_t^*(x^{-i} | x^i)$. We define structured perfect Bayesian equilibrium as follows.

Definition 1 (SPBE): A structured perfect Bayesian equilibrium (SPBE) is a PBE of the considered dynamic game where at any time t , for any agent i , its equilibrium strategy $\beta_t^{*,i}$ depends on player i 's information $(a_{1:t-1}, x^i)$ through the common belief and its private type x^i .

We conjecture that as in [14], such *structured* strategies form a rich class where any expected reward profile of the players that can be generated from any general strategy profile can also be generated using such structured strategy profile.

In the following we present a backward-forward methodology to compute SPBE of this game.

1) Backward Recursion: In this section, we define an equilibrium generating function $\theta = (\theta_t^i)_{i \in \mathcal{N}, t \in \mathcal{T}}$ and a sequence of functions $(V_t^i)_{i \in \mathcal{N}, t \in \{1, 2, \dots, T+1\}}$, where $V_t^i : \Delta(\mathcal{X}) \times \mathcal{X}^i \rightarrow \mathbb{R}$, in a backward recursive way, as follows.

1. Initialize $\forall \pi_{T+1} \in \Delta(\mathcal{X}), x^i \in \mathcal{X}^i$,

$$V_{T+1}^i(\pi_{T+1}, x^i) := 0. \quad (6)$$

2. For $t = T, T-1, \dots, 1$, $\forall \pi_t \in \Delta(\mathcal{X})$, let $\theta_t[\pi_t]$ be generated as follows. Set $\tilde{\gamma}_t = \theta_t[\pi_t]$, where $\tilde{\gamma}_t$ is the solution, if it exists⁵, of the following fixed-point equation, $\forall i \in \mathcal{N}, x^i \in \mathcal{X}^i$,

$$\begin{aligned} & \tilde{\gamma}_t^i(\cdot | x^i) \in \arg \max_{\gamma_t^i(\cdot | x^i)} \mathbb{E}^{\tilde{\gamma}_t^i(\cdot | x^i), \pi_t} \left\{ R_t^i(X, A_t) + \right. \\ & \left. V_{t+1}^i(F_t(\pi_t, \tilde{\gamma}_t, A_t), x^i) | x^i \right\}, \end{aligned} \quad (7)$$

where expectation in (7) is with respect to random variables (X, A_t) through the measure $\sum_{x^{-i}, x} \pi_t(x^{-i} | x^i) \gamma_t^i(a_t^i | x^i) \tilde{\gamma}_t^{-i}(a_t^{-i} | x^{-i})$, F_t is defined in Lemma 1. Furthermore, set

$$\begin{aligned} & V_t^i(\pi_t, x^i) := \mathbb{E}^{\tilde{\gamma}_t^i(\cdot | x^i), \pi_t} \left\{ R_t^i(x, A_t) + \right. \\ & \left. V_{t+1}^i(F_t(\pi_t, \tilde{\gamma}_t, A_t), x^i) | x^i \right\}. \end{aligned} \quad (8)$$

It should be noted that (7) is a fixed-point equation where the solution of the above fixed-point equation $\tilde{\gamma}_t^i$ appears in both, the left-hand-side and the right-hand-side of the equation. However, it is not the outcome of the maximization operation as in a best response equation similar to that of a Bayesian Nash equilibrium.

⁵Existence of general solution of this per stage fixed-point equation is discussed in Remark 4.

2) Forward Recursion: Based on θ defined above in (6)–(8), we now construct a set of strategies β^* and beliefs μ^* for the game \mathcal{D} in a forward recursive way, as follows, where $\beta_t^{i,*} : \mathcal{A}^{t-1} \times \mathcal{X}^i \rightarrow \Delta(\mathcal{A}^i)$ and $\mu_t^{i,*} : \mathcal{A}^{t-1} \times \mathcal{X}^i \rightarrow \Delta(\mathcal{X}^i)$. We first define the update function of equilibrium common belief $\pi_t^* : \mathcal{A}^{t-1} \rightarrow \Delta(\mathcal{X})$, where for every private history $(a_{1:t-1}, x^i)$ of player i , its equilibrium belief $\mu_t^{i,*}$ is constructed by conditioning common belief π_t^* on its private belief x^i .

1. Initialize at time $t = 0, \forall i \in \mathcal{N}$,

$$\pi_1^*[\phi](x) := Q^x(x) \quad (9)$$

2. For $t = 1, 2, \dots, T, i \in \mathcal{N}, \forall a_{1:t-1}, x^i$

$$\beta_t^{*,i}(a_t^i | a_{1:t-1}, x^i) := \theta_t^i[\pi_t^*[a_{1:t-1}]](a_t^i | x^i) \quad (10a)$$

$$\mu_t^{*,i}(x^{-i} | a_{1:t-1}, x^i) := \pi_t^*[a_{1:t-1}](x^{-i} | x^i) \quad (10b)$$

$$\pi_{t+1}^*[a_{1:t}] := F_t(\pi_t^*[a_{1:t-1}], \theta_t[\pi_t^*[a_{1:t-1}]], A_t) \quad (10c)$$

where F_t is defined in Lemma 1.

In the following theorem, we show that the equilibrium strategy and belief profile (β^*, μ^*) defined above constitute a PBE of the game considered.

Theorem 1: A strategy and belief profile (β^*, μ^*) , constructed through backward/forward recursive methodology is a PBE of the game, i.e. $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, x^i), \beta_t^i, \mu_t^i$,

$$\begin{aligned} & \mathbb{E}^{\beta_t^{*,i}, \beta_t^{*,i}, \mu_t^{*,i}} \left\{ \sum_{n=t}^T R_n^i(X, A_n) | a_{1:t-1}, x^i \right\} \geq \\ & \mathbb{E}^{\beta_t^i, \beta_t^i, \mu_t^i} \left\{ \sum_{n=t}^T R_n^i(X, A_n) | a_{1:t-1}, x^i \right\}. \end{aligned} \quad (11)$$

Proof: The proof is provided in Appendix II. ■

V. REMARKS

A few remarks are in order.

Remark 1: We note that in Step-2 in Backward Recursion, a fixed-point equation is solved in $(\tilde{\gamma}_t^i, V_t^i)_{i \in \mathcal{N}}$ for each $\pi_t \in \Delta(\mathcal{X})$ and for each $t \in T, T-1, \dots, 1$. Since Backward recursion dominates the computational complexity, the complexity of our methodology is linear in time. However, we do not make any claim about the complexity of solving the fixed-point in each instant, which, together with its existence as discussed below, is an important open question for future research.

Remark 2: We emphasize that even though the backward-forward methodology presented above finds a class of equilibrium strategies that are structured i.e. depend on the common belief and private state, the unilateral deviations of players in (11) are considered in the space of general strategies, i.e., the methodology does not make any bounded rationality assumptions.

Remark 3: Intuition of the proof: In such games, one could use the one-shot deviation principle [27] to argue that sequential rationality reduces to showing that no player wants to deviate unilaterally in β_t^i at any time t , keeping the rest of the strategy $\beta_{t+1:T}^{i,*}$ as the equilibrium strategy. We argue that this is equivalent to (7) i.e. for a given (π_t^*, x^i) , a player's unilateral deviation in its strategy $\beta_t^{i,*}$ is the same as unilateral deviation in its measure $\gamma_t^i(\cdot | x^i)$ on its action a_t^i (and not on the whole function $\gamma_t^i(\cdot | \cdot)$). This is because under such unilateral deviations, a player uses the same future information states (π_{t+1}^*, x^i) as it would have done under an equilibrium strategy, whose update depends on equilibrium $\tilde{\gamma}_t$. This is so because player uses the same equilibrium π_{t+1}^* to predict other player's actions, and update of its equilibrium private belief x^i does not depend on γ_t^i .

Remark 4: While it is known that for any finite dynamic game with asymmetric information and perfect recall, there always exists a PBE [10, Prop. 249.1], existence of SPBE is not always guaranteed. It is clear from our methodology that existence of SPBE boils down to existence of a solution to the fixed-point equation (7) at

every stage. Specifically, at each time t given the functions V_{t+1}^i for all $i \in \mathcal{N}$ from the previous round (in the backwards recursion) equation (7) must have a solution $\tilde{\gamma}_t^i$ for all $i \in \mathcal{N}$. Generally, existence of equilibria is shown through Kakutani's fixed point theorem, as is done by proving existence of a mixed strategy Nash equilibrium of a finite game [10], [28]. This is done by showing existence of fixed point of the best-response correspondences of the game. Among other conditions, it requires the *closed graph property* of the correspondences, which is usually implied by the continuity property of the utility functions involved. For (7) establishing existence is not straightforward due to: (a) potential discontinuity of the π_t update function F_t when the denominator in the Bayesian update is 0 and (b) potential discontinuity of the value functions, V_{t+1}^i . It is noted in [17, ch. 5] through [29] that for dynamic zero-sum games with asymmetric information, the value function is a continuous function, on the basis of which and the above mentioned arguments, the existence of equilibria in sequential decomposition for such games is established. Furthermore, authors in [18] consider an LQG dynamic game with asymmetric information (with linear state update, quadratic costs and Gaussian noise), where they present sufficient algorithmic conditions for such an equilibrium to exist. Authors in [30] study a public goods game for which signaling equilibria were found numerically using such methodology where agents have independent types which are perfectly observed by them and agents observe each others' actions in each time-period. Having said this, existence of the fixed-point equation for a more general class of problems remain an open question.

Remark 5: This model also allows to incorporate many bounded rationality models. Some examples include using a discount factor δ or by restricting the search for optimum $\gamma_t^i(\cdot|x^i)$ in (7) in the space of functions that are linear in private information variables.

Remark 6: In this paper, we considered a model with static types of the players. If the types of the players were also dynamic, say in a Markovian or a controlled Markovian way, we believe such a decomposition is not possible. The reason is that there is no consistent set of sufficient belief statistics that summarizes the observed history of the players in such a way that does not lead to *infinite regress of beliefs*. We pose it as an open problem if there exists any special structure of dynamic evolution of private beliefs with correlated types that allows for sequential decomposition in dynamic games with asymmetric information.

VI. CONCLUSION

In this paper, we considered a model of repeated game where there is an underlying state of the system that is static and players jointly observe actions of other players with delay 1 and correlated private observations of the state of the system. Each player receives a reward that is a function of the state and actions of all the players. We define a common information belief state and private information belief states of the players. We then presented a backward-forward methodology similar to the one presented in [1] to compute its structured perfect Bayesian equilibria (SPBE). Future work includes proving such a methodology for discounted infinite-horizon case and specializing results to many practical settings such as games on graph, where players who are connected on the graph have correlated private information. Some practical applications of interest include security games for cyber-physical systems and Bayesian learning games in a social network with fully rational and potentially adversarial agents. An important future direction would be to investigate if this new result in the theory of such games can facilitate more efficient way of computing equilibria in [6] which can have significant implications in developing a software to solve real-world strategic problems.

VII. ACKNOWLEDGEMENT

The author would like to thank Achilleas Anastasopoulos for useful comments and pointing an error in previous draft, and Francois Baccelli and Sriram Vishwanath for the encouragement and support.

APPENDIX I

Lemma 2: There exists an update function F_t of π_t , independent of θ

$$\pi_{t+1} = F_t(\pi_t, \gamma_t, a_t) \quad (12)$$

Proof:

$$\pi_{t+1}(x) = P^\theta(x|a_{1:t}, \gamma_{1:t+1}) \quad (13a)$$

$$= P^\theta(x|a_{1:t}, \gamma_{1:t}) \quad (13b)$$

$$= \frac{P^\theta(x, a_t|a_{1:t-1}, \gamma_{1:t})}{\sum_{x, a_t} P^\theta(x, a_t|a_{1:t-1}, \gamma_{1:t})} \quad (13c)$$

$$= \frac{\pi_t(x) \left(\prod_{i=1}^N \gamma_t^i(a_t^i|x^i) \right)}{\sum_x \pi_t(x) \prod_{i=1}^N \gamma_t^i(a_t^i|x^i)} \quad (13d)$$

if the denominator is not 0 and

$$\pi_{t+1}(x) = \pi_t(x) \quad (13e)$$

otherwise. Thus we have,

$$\pi_{t+1} = F_t(\pi_t, \gamma_t, a_t) \quad (13f)$$

■

APPENDIX II

(PROOF OF THEOREM 1)

In the following theorem, we will assume that the equilibrium strategies and beliefs (β^*, μ^*) are generated using an equilibrium function θ . Moreover, we will also use π_t^* map corresponding to μ_t^* as defined in (10). With slight abuse of notation, we use both beliefs and belief functions as superscripts on expectations, where the reference is clear from the context. These functions when used as superscripts in expectation denote the belief functions, which when applied on the conditioned random variables, define beliefs on the random variables of interest inside the expectation.

Proof: We prove (11) using induction and from results in Lemma 3, 4 and 5 proved in Appendix III. For base case at $t = T$, $\forall i \in \mathcal{N}$, $(a_{1:t-1}, x^i) \in \mathcal{H}_T^i, \beta^i$

$$\mathbb{E}^{\beta_T^{*i}, \pi_T^{*i}} \left\{ R_t^i(X, A_t) | a_{1:t-1}, x^i \right\} = V_T^i(\pi_T^*[a_{1:t-1}], x^i) \quad (14a)$$

$$\geq \mathbb{E}^{\beta_T^i, \pi_T^*} \left\{ R_t^i(X, A_t) | a_{1:t-1}, x^i \right\} \quad (14b)$$

where (14a) follows from Lemma 5 and (14b) follows from Lemma 3 in Appendix III.

Let the induction hypothesis be that for $t + 1$, $\forall i \in \mathcal{N}$, $(a_{1:t}, x^i) \in \mathcal{H}_{t+1}^i, \beta^i$,

$$\mathbb{E}^{\beta_{t+1:T}^{*i}, \pi_{t+1}^{*i}} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) | a_{1:t-1}, x^i \right\} \geq \mathbb{E}^{\beta_{t+1:T}^i, \pi_{t+1}^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) | a_{1:t-1}, x^i \right\}. \quad (15a)$$

Then $\forall i \in \mathcal{N}$, $(a_{1:t-1}, x^i) \in \mathcal{H}_t^i, \beta^i$, we have

$$\mathbb{E}^{\beta_{t:T}^{*i}, \pi_t^{*i}} \left\{ \sum_{n=t}^T R_n^i(X, A_n) | a_{1:t-1}, x^i \right\} = V_t^i(\pi_t^*[a_{1:t-1}], x^i) \quad (16a)$$

$$\geq \mathbb{E}^{\beta_t^i, \pi_t^*} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[a_{1:t-1}, A_t], x^i) | a_{1:t-1}, x^i \right\} \quad (16b)$$

$$= \mathbb{E}^{\beta_t^i, \pi_t^*} \left\{ R_t^i(X, A_t) + \mathbb{E}^{\beta_{t+1:T}^{*i}, \pi_{t+1}^{*i}} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) | a_{1:t-1}, A_t, x^i \right\} | a_{1:t-1}, x^i \right\} \quad (16c)$$

$$\geq \mathbb{E}^{\beta_t^i \beta_t^{*-i}, \pi_t^*} \left\{ R_t^i(X, A_t) + \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t-1}, A_t, x^i \right\} \middle| a_{1:t-1}, x^i \right\} \quad (16d)$$

$$= \mathbb{E}^{\beta_t^i \beta_t^{*-i}, \pi_t^*} \left\{ R_t^i(X, A_t) + \mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t-1}, A_t, x^i \right\} \middle| a_{1:t-1}, x^i \right\} \quad (16e)$$

$$= \mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ \sum_{n=t}^T R_n^i(X, A_n) \middle| a_{1:t-1}, x^i \right\}, \quad (16f)$$

where (16a) follows from Lemma 5, (16b) follows from Lemma 3, (16c) follows from Lemma 5, (16d) follows from induction hypothesis in (15a) and (16e) follows from Lemma 4. ■

APPENDIX III

As we did in the previous theorem, in the following lemmas, we would assume that the equilibrium strategies and beliefs (β^*, μ^*) are generated using an equilibrium function θ . We also use π_t^* corresponding to μ_t^* as defined in (10).

Lemma 3: This lemma states that the reward that player i would get on playing equilibrium strategy will be greater or equal to the reward it would get if it deviates only at time t , keeping the rest of its strategy as equilibrium strategy. $\forall t \in \mathcal{T}, i \in \mathcal{N}, (a_{1:t-1}, x^i) \in \mathcal{H}_t^i, \beta_t^i$

$$V_t^i(\pi_t^*[a_{1:t-1}], x^i) \geq \mathbb{E}^{\beta_t^i \beta_t^{*-i}, \pi_t^*} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[a_{1:t-1}, A_t]) \middle| a_{1:t-1}, x^i \right\} \quad (17)$$

Proof: We prove this lemma by contradiction.

Suppose the claim is not true for t . This implies $\exists i, \hat{\beta}_t^i, \hat{a}_{1:t-1}, x^i$ such that

$$\mathbb{E}^{\hat{\beta}_t^i \beta_t^{*-i}, \pi_t^*} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[\hat{a}_{1:t-1}, A_t]) \middle| \hat{a}_{1:t-1}, x^i \right\} > V_t^i(\pi_t^*[\hat{a}_{1:t-1}]). \quad (18)$$

We will show that this contradicts the definition of V_t^i in (8).

Construct $\hat{\gamma}_t^i(a_t^i | x^i) = \hat{\beta}_t^i(a_t^i | \hat{a}_{1:t-1}, x^i)$

Then for $\hat{a}_{1:t-1}, x^i$, we have

$$V_t^i(\pi_t^*[\hat{a}_{1:t-1}]) < \mathbb{E}^{\hat{\beta}_t^i, \beta_t^{*-i}, \pi_t^*, x^i} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[\hat{a}_{1:t-1}, A_t]) \middle| \hat{a}_{1:t-1}, x^i \right\} \quad (19a)$$

$$= \sum_{\substack{x, a_t, a_t, \\ x^i, x^{-i}}} \left[R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[\hat{a}_{1:t-1}, a_t]) \right] \pi_t^*[\hat{a}_{1:t-1}](x^{-i} | x^i) \quad (19b)$$

$$= \sum_{x, a_t, x^i, x^{-i}} \left[R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[\hat{a}_{1:t-1}, a_t]) \right] \pi_t^*[\hat{a}_{1:t-1}](x^{-i} | x^i) \hat{\gamma}_t^i(a_t^i | x^i) \beta_t^{*-i}(a_t^{-i} | \pi_t^*[\hat{a}_{1:t-1}], x^{-i}) \quad (19c)$$

$$= \mathbb{E}^{\hat{\gamma}_t^i(\cdot | x^i) \beta_t^{*-i}, \pi_t^*[\hat{a}_{1:t-1}]} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[\hat{a}_{1:t-1}, A_t]) \middle| x^i \right\} \quad (19d)$$

$$\leq \max_{\hat{\gamma}_t^i(\cdot | x^i)} \mathbb{E}^{\hat{\gamma}_t^i(\cdot | x^i) \beta_t^{*-i}, \pi_t^*[\hat{a}_{1:t-1}]} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[\hat{a}_{1:t-1}, A_t]) \middle| x^i \right\} \quad (19e)$$

$$= V_t^i(\pi_t^*[\hat{a}_{1:t-1}]) \quad (19f)$$

where (19a) follows from (18), (19c) follows from the definition of $\hat{\gamma}_t^i$, and (19f) follows from the definition of V_t^i in (8). ■

Lemma 4: $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, x^i) \in \mathcal{H}_{t+1}^i$ and β_t^i

$$\mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t}, x^i \right\} = \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t}, x^i \right\}. \quad (20)$$

Proof: Since the above expectations involve random variables $X^{-i}, A_{t+1:T}$, we consider $P^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*}(x, a_{t+1:T} | a_{1:t}, x^i)$.

$$P^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*}(x, a_{t+1:T} | a_{1:t}, x^i) = P^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*}(x | a_{1:t}, x^i) P^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_t^*}(a_{t+1:T} | x, a_{1:t}) \quad (21a)$$

We first note that

$$P^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*}(a_{t+1:T} | x, a_{1:t}) = \beta_{t+1}^i(a_{t+1}^i | a_{1:t-1}, x^i) \left(\beta_{t+1}^{*-i}(a_{t+1}^{-i} | a_{1:t-1}, x^{-i}) \right) P^{\beta_{t+2:T}^i \beta_{t+2:T}^{*-i}, \pi_t^*}(a_{t+2:T} | a_{1:t}, x) \quad (21b)$$

$$= P^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*}(a_{t+1:T} | a_{1:t}, x) \quad (21c)$$

Thus using the above equation, (21a) is given by

$$P^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*}(x, a_{t+1:T} | a_{1:t-1}, x^i) \quad (21d)$$

$$= P^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*}(x, a_{t+1:T} | a_{1:t-1}, x^i). \quad (21e)$$

Lemma 5: $\forall i \in \mathcal{N}, t \in \mathcal{T}, a_{1:t-1} \in \mathcal{H}_t^i, x^i \in (\mathcal{X}^i)$ ■

$$V_t^i(\pi_t^*[a_{1:t-1}], x^i)$$

$$= \mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ \sum_{n=t}^T R_n^i(X, A_n) \middle| a_{1:t-1}, x^i \right\}. \quad (22)$$

Proof:

We prove the lemma by induction. For $t = T$,

$$\mathbb{E}^{\beta_T^i \beta_T^{*-i}, \pi_T^*} \left\{ R_T^i(X, A_T) \middle| a_{1:t-1}, x^i \right\} = \sum_{x^{-i}, a_T, x^{-i}} R_T^i(X, A_T) \pi_T^*[a_{1:t-1}](x^{-i} | x^i) \beta_T^{*-i}(a_T^{-i} | a_{1:t-1}, x^{-i}) \quad (23a)$$

$$= V_T^i(\pi_T^*[a_{1:t-1}], x^i), \quad (23b)$$

where (23b) follows from the definition of V_t^i in (8) and the definition of β_T^* in the forward recursion in (10). Suppose the claim is true for $t + 1$, i.e., $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, x^i) \in \mathcal{H}_{t+1}^i$

$$V_{t+1}^i(\pi_{t+1}^*[a_{1:t}], x^i) = \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t}, x^i \right\}. \quad (24)$$

Then $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, x^i) \in \mathcal{H}_t^i$, we have

$$\mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ \sum_{n=t}^T R_n^i(X, A_n) \middle| a_{1:t-1}, x^i \right\} = \mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ R_t^i(X, A_t) + \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t-1}, A_t, x^i \right\} \middle| a_{1:t-1}, x^i \right\} \quad (25a)$$

$$= \mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*-i}, \pi_t^*} \left\{ R_t^i(X, A_t) + \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*-i}, \pi_{t+1}^*} \left\{ \sum_{n=t+1}^T R_n^i(X, A_n) \middle| a_{1:t-1}, A_t, x^i \right\} \middle| a_{1:t-1}, x^i \right\} \quad (25b)$$

$$= \mathbb{E}^{\beta_{t:T}^{*,i}, \beta_{t:T}^{*, -i}, \pi_t^*} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[a_{1:t-1}, A_t], x^i) \mid a_{1:t-1}, x^i \right\} \quad (25c)$$

$$= \mathbb{E}^{\beta_t^{*,i}, \beta_t^{*, -i}, \pi_t^*} \left\{ R_t^i(X, A_t) + V_{t+1}^i(\pi_{t+1}^*[a_{1:t-1}, A_t], x^i) \mid a_{1:t-1}, x^i \right\} \quad (25d)$$

$$= V_t^i(\pi_t^*[a_{1:t-1}], x^i) \quad (25e)$$

where (25b) follows from Lemma 4 in Appendix III, (25c) follows from the induction hypothesis in (24), (25d) follows because the random variables involved in expectation, X, A_t do not depend on $\beta_{t+1:T}^{*,i}$ and (25e) follows from the definition of V_t^i in (8). ■

REFERENCES

- [1] D. Vasal and A. Anastasopoulos, "A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information," Tech. Rep., Aug. 2015. [Online]. Available: <http://arxiv.org/abs/1508.06269>
- [2] G. A. Akerlof, "The market for lemons": Quality uncertainty and the market mechanism," *The quarterly journal of economics*, pp. 488–500, 1970.
- [3] M. Spence, "Job market signaling," *The quarterly journal of Economics*, pp. 355–374, 1973.
- [4] D. Fudenberg and J. Tirole, "Perfect bayesian equilibrium and sequential equilibrium," *journal of Economic Theory*, vol. 53, no. 2, pp. 236–260, 1991.
- [5] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, p. 354, 2017.
- [6] N. Brown and T. Sandholm, "Safe and nested subgame solving for imperfect-information games," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 689–699. [Online]. Available: <http://papers.nips.cc/paper/6671-safe-and-nested-subgame-solving-for-imperfect-information-games.pdf>
- [7] L. S. Shapley, "Stochastic games," *Proceedings of the national academy of sciences*, vol. 39, no. 10, pp. 1095–1100, 1953.
- [8] T. Başar and G. Olsder, *Dynamic Noncooperative Game Theory, 2nd Edition*. Society for Industrial and Applied Mathematics, 1998.
- [9] J. Filar and K. Vrieze, *Competitive Markov decision processes*. Springer Science & Business Media, 2012.
- [10] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, ser. MIT Press Books. The MIT Press, 1994, vol. 1.
- [11] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.
- [12] G. J. Mailath and L. Samuelson, *Repeated games and reputations: long-run relationships*. Oxford university press, 2006.
- [13] E. Maskin and J. Tirole, "Markov perfect equilibrium: I. observable actions," *Journal of Economic Theory*, vol. 100, no. 2, pp. 191–219, 2001.
- [14] 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*, 2018.
- [15] D. Vasal and A. Anastasopoulos, "A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information," in *American Control Conference*, Boston, US, 2016, available on arXiv.
- [16] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, "Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 222–237, 2017.
- [17] H. T. Jahormi, "On design and analysis of cyber-physical systems with strategic agents," Ph.D. dissertation, University of Michigan, Ann Arbor, 2017.
- [18] D. Vasal and A. Anastasopoulos, "Signaling equilibria of dynamic LQG games with asymmetric information," in *Conference on Decision and Control*, 2016.
- [19] —, "Decentralized Bayesian learning in dynamic games," in *Allerton Conference on Communication, Control, and Computing*, 2016. [Online]. Available: <https://arxiv.org/abs/1607.06847>
- [20] H. Tavafoghi, Y. Ouyang, and D. Teneketzis, "On stochastic dynamic games with delayed sharing information structure," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 7002–7009.
- [21] P. Kumar and P. Varaiya, "Stochastic systems," 1986.
- [22] Y.-C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.
- [23] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *Automatic Control, IEEE Transactions on*, vol. 58, no. 7, pp. 1644–1658, 2013.
- [24] A. Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," *Automatic Control, IEEE Transactions on*, vol. 58, no. 9, pp. 2377–2382, 2013.
- [25] A. Mahajan and D. Teneketzis, "On the design of globally optimal communication strategies for real-time communication systems with noisy feedback," *IEEE J. Select. Areas Commun.*, no. 4, pp. 580–595, May 2008.
- [26] D. Vasal and A. Anastasopoulos, "Stochastic control of relay channels with cooperative and strategic users," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3434–3446, Oct 2014.
- [27] E. Hendon, H. J. Jacobsen, and B. Sloth, "The one-shot-deviation principle for sequential rationality," *Games and Economic Behavior*, vol. 12, no. 2, pp. 274–282, 1996.
- [28] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.
- [29] P. R. Milgrom and R. J. Weber, "Distributional strategies for games with incomplete information," *Mathematics of operations research*, vol. 10, no. 4, pp. 619–632, 1985.
- [30] D. Vasal, A. Sinha, and A. Anastasopoulos, "A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information," Tech. Rep., Aug. 2015. [Online]. Available: <http://arxiv.org/abs/1508.06269v3>