Multi-Agent Sequential Hypothesis Testing

A Sequential Game of Synchronizing Coordination

Kwang-Ki K. Kim and Jeff S. Shamma

***Abstract*— This paper considers multi-agent sequential hypothesis testing and presents a framework for strategic learning in sequential games with explicit consideration of both temporal and spatial coordination. The associated Bayes risk functions explicitly incorporate costs of taking private/public measurements, costs of time-difference and disagreement in actions of agents, and costs of false declaration/choices in the sequential hypothesis testing. The corresponding sequential decision processes have well-defined value functions with respect to (a) the belief states for the case of conditional independent private noisy measurements that are also assumed to be independent identically distributed over time, and (b) the information states for the case of correlated private noisy measurements. A sequential investment game of strategic coordination and delay is also discussed as an application of the proposed strategic learning rules.**

I. Introduction

Many interactive decision processes of multi-agent systems are essentially modeled as games with incomplete information. Each player's payoff is a function of his action, the opponents' actions, and the state of nature that includes all uncertain parameters defining games. Assuming all participants are Bayesian rational players, which means that their strategies are contingent on their (Bayesian) beliefs about the state of the world,* a decision-maker must take the others' beliefs into account as well as his own belief. Harsanyi [1] proposed the concept of "types" that are determined by a prior chance-move and correspond to private information available to players. (Bayesian) Rational behaviors of players are determined by beliefs about the types and the opponents' type-contingent strategies. This means that the interactive decision processes require higher-order beliefs, i.e., players' beliefs about other players' beliefs, player's beliefs about other players' belief about other player's beliefs, ad infinitum. The requirement of higher-order beliefs makes the associated games intractable, even though there are some Bayesian games that allow tractable analysis, e.g., global games.

For games of incomplete information, learning the value of state of the world or nature by taking private/public noisy measurements/signals is indispensable for sequential decision-making. In particular, measurements can be sequentially taken with some costs and a learning rule is to determine an optimal stopping-time and an optimal choice that minimize the associated Bayes risks (or maximize Bayes rewards) [2]–[4]. In [5], [6], some results on single-agent sequential hypothesis tests based on dynamic programming are presented.

For an unknown state of nature $\theta \in \Theta$, a Bayesian statistician has a posterior distribution at time t denoted by $p_t^\theta \triangleq \mathbf{P}(h = h_\theta | \mathcal{F}^t)$ where h_θ denotes a hypothesis corresponding to $\theta \in \Theta$ and $\mathcal{F}^t$ refers to a filtration adapted from a sequence of available measurements $\{\zeta_0, \zeta_1, \cdots, \zeta_t\}$. This paper extends the results of single-agent sequential hypothesis testing to multi-agent settings for which temporal coordination between agents is also taken into account as well as coordination of actions.

For multi-agent sequential hypothesis testing, game theoretic frameworks of modeling strategic interactions and decision processes can be conveniently employed. In [7], it was shown that the so-called common q-belief can be used as an approximate common knowledge in the sense that equilibria of the game with common q-belief have continuity at $q = 1$ (or as $q \to 1$ from below), with respect to the true game of the actual θ. A natural question for learning in games with incomplete information (i.e., Bayesian games) is under which conditions common q-belief can be achieved. Partial answers for such a question are presented in [8]–when each agent has a finite signal space, the agents are able to commonly learn the value of the state of nature corresponding to incomplete information. However, their learning environments including payoffs are not realistic, since it is assumed that waiting is costless, i.e., obtaining measurements/signals and delay in actions do not cost. This also implies that achieving common q-beliefs is considered only in the asymptotic manner. In realistic sequential learning, taking signals/measurements and/or delaying choices should be considered to be costly in strategic learning, so that learning processes should stop in finite time with probability one.

For an interactive learning problem, one goal is to find an optimal policy for each agent to minimize the cost of errors (type I and type II), the cost of observations, and the cost of disagreement and time-difference in choices among agents. It is assumed that there is no communication or information exchange between agents. To our best knowledge, this paper is the first work on multi-agent sequential hypothesis testing that considers temporal synchronization of learning between agents as well as coordination in learning the value of unknown states, for which taking measurements and delaying choices are costly. This formulation presents a novel framework of multi-agent sequential hypothesis testing

K.-K. K. Kim is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, (kwangki.kim@ece.gatech.edu).

J. S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, (shamma@gatech.edu), and with King Abdullah University of Science and Technology (KAUST), (jeff.shamma@kaust.edu.sa).

*We use the term "state of the world" to denote the concatenation of the "state of nature" and "behaviors" of the opponents.

with consideration of synchronizing coordination, agreeing actions, and inverse (i.e., inference) problems. In this paper, we particularly focus on two-player sequential hypothesis testing with (a) conditionally independent noisy measurements that are independent identically distributed and (b) correlated noisy measurements. For the case of (a), the belief state corresponding to the posterior probability distribution over the state of nature is a sufficient statistic for the associated interactive sequential decision problems, whereas the case of (b) does not allow abstraction of information states into belief states for the well-defined statistical sequential decision processes. This is, roughly speaking, because the corresponding Bayes risks and optimal cost-to-go functions are not computable in terms of the belief states only.

It is expected that our work can be extended to strategic learning in dynamic games, e.g., dynamic global games [9] and dynamic coordination games with options to delay [10], about which such an extension for sequential games of strategic coordination and delay is partially discussed in Section IV. The existing literature on this subject focuses on analysis in terms of high-order beliefs [11] and global games [12], whereas our work presents methods of multi-agent sequential learning that neither require approximate common/higher-order beliefs nor restrict game environments.

II. Preliminaries

A. Single-agent sequential hypothesis testing

1) Bayes risk: We formulate statistical inference as a decision problem. Consider the cost functions $C : \Theta \times \mathcal{A} \to \mathbb{R}$ where $\Theta = \{\theta_1, \cdots, \theta_n\}$ denotes the set of states of nature and $\mathcal{A} = \{a_1, \cdots, a_\ell\}$ refers to the set of actions. Denote the set of information vectors by $\mathcal{I}$. Then the risk function associated with an action $a \in \mathcal{A}$ for given information vector $I \in \mathcal{I}$ is

$$R(a, I) = \mathbf{E}[C(\theta, a)|I] = \sum_{i=1}^{n} C(\theta_i, a) P(\theta_i|I)\,.$$

Consider an information-contingent decision rule $\mu : \mathcal{I} \to \mathcal{A}$ or a randomized counterpart $\mu : \mathcal{I} \to \Delta(\mathcal{A})$, where $\Delta(\mathcal{A})$ denotes the set of probability distributions over $\mathcal{A}$. Then the risk function of the decision rule $\mu : \mathcal{I} \to \Delta(\mathcal{A})$ is

$$\begin{aligned} R^{\mu}(I) &= \mathbf{E}[C(\theta, \mu(I))|I]\,, \\ &= \sum_{k=1}^{\ell} \sum_{i=1}^{n} C(\theta_i, a_k) \mu(a_k|I) P(\theta_i|I)\,, \end{aligned}$$

where $\mu(a_k|I)$ refers to the probability of taking action $a_k \in \mathcal{A}$ for given information vector $I \in \mathcal{I}$, by abuse of notation. For pure (i.e., deterministic) strategies, an optimal policy can be obtained by

$$\mu^*(I) := \arg\min_{a \in \mathcal{A}} R(a, I)$$

and since the risk function $R(a, I)$ can be fully characterized by the posterior $P(\cdot|I) \in \Delta(\Theta)$, there exists a belief-contingent policy $\bar{\mu} : \Delta(\Theta) \to \mathcal{A}$ such that

$$\bar{\mu}(P(\cdot|I)) \equiv \mu^*(I),$$

which indeed follows from the fact that the belief state is a sufficient statistic for this decision problem. Similar observations are extended to multi-agent sequential hypothesis testing for the case of conditionally independent private measurements and are presented in Section III-C in this paper.

2) Dynamic programming solution: This section summarizes the classical results on sequential hypothesis tests and we refer the interested readers to the research monographs [5], [6], [13] for details. For $\theta \in \Theta$, $p_t^\theta = \mathbf{P}(h = h_\theta|\zeta_0, \cdots, \zeta_t)$, where h_θ refers to the hypothesis corresponding to $\theta \in \Theta$. For notational simplicity and technical convenience, consider the set of binary states $\Theta = \{0, 1\}$ and define the associated belief state of hypothesis h_0 by

$$p_t \triangleq \mathbf{P}(h = h_0|\zeta_0, \cdots, \zeta_t)\,.$$

Recursive Bayesian updates of belief states corresponding to posterior distributions is given by

$$p_{t+1}(\zeta_{t+1}) = \frac{p_t f_0(\zeta_{t+1})}{p_t f_0(\zeta_{t+1}) + (1 - p_t) f_1(\zeta_{t+1})}\,,$$

which defines a Markov process, provided ζ_{t+1} is independent of $\{\zeta_s \in \mathcal{Z} : s \leq t\}$. $f_\theta(\cdot) = \mathbf{P}(\cdot|\theta)$ refers to the conditional probability corresponding to each hypothesis h_θ for $\theta \in \Theta$.

Consider a finite time-horizon sequential hypothesis testing with the cost of continuing observations $C_{\rm o} > 0$ and the costs of erroneous declaration $C_{\rm e} : \Theta \times \mathcal{D} \to \mathbb{R}_+$ where $\mathcal{D} = \{0, 1\} \equiv \Theta$. The optimal expected cost for the last period is

$$\bar{J}_N(p_N) = \min_{a \in \mathcal{D}} \{\underbrace{p_N C_{\rm e}(0, a) + (1 - p_N) C_{\rm e}(1, a)}_{\bar{R}(a, p_N)}\},$$

where $C_{\rm e}(0, 0) = C_{\rm e}(1, 1) = 0$.

At time $t \leq N - 1$, the optimal expected cost-to-go is

$$\bar{J}_t(p_t) = \min\left\{\min_{a \in \mathcal{D}} \bar{R}(a, p_t), C_{\rm o} + W_t(p_t)\right\}\,,$$

where $W_t(p_t) := \mathbf{E}\big[\bar{J}_{t+1}\left(p_{t+1}(\zeta_{t+1})\right)|p_t\big]$ and the expectation is taken over a random process ζ_{t+1} whose probability distribution is the marginal probability corresponding to the current belief state p_t, i.e., $p_t f_0(\zeta_{t+1}) + (1 - p_t) f_1(\zeta_{t+1})$.

Assume $C_{\rm o} + W_{N-1}\Big(\frac{C_{\rm e}(1,0)}{C_{\rm e}(1,0)+C_{\rm e}(0,1)}\Big) < \frac{C_{\rm e}(1,0)C_{\rm e}(0,1)}{C_{\rm e}(1,0)+C_{\rm e}(0,1)}$.[†] Then an optimal policy of sequential binary hypothesis testing with the belief state p_t at each time t is

$$\begin{cases} \text{declare } h_0 & \text{if } p_t > \overline{\gamma}_t\,, \\ \text{declare } h_1 & \text{if } p_t < \underline{\gamma}_t\,, \\ \text{continue} & \text{if } \underline{\gamma}_t \leq p_t \leq \overline{\gamma}_t\,, \end{cases}$$

where the thresholds $\underline{\gamma}_t$ and $\overline{\gamma}_t$ solve the equations

$$\begin{aligned} \underline{\gamma}_t C_{\rm e}(0, 1) &= C_{\rm o} + W_t(\underline{\gamma}_t)\,, \\ (1 - \overline{\gamma}_t) C_{\rm e}(1, 0) &= C_{\rm o} + W_t(\overline{\gamma}_t)\,, \end{aligned}$$

[†]This assumption is required to guarantee that the intervals $[\underline{\gamma}_t, \overline{\gamma}_t]$ corresponding to the action of continuation are non-empty for all $t \leq N-1$. If this assumption is not satisfied then one can reduce the horizon length since $\overline{J}_{N-1}$ is identical to $\overline{J}_N$.

that also satisfy the monotonic inequalities: $\overline{\gamma}_t \geq \overline{\gamma}_{t+1}$, $\overline{\gamma}_t \leq 1 - \frac{C_\mathrm{o}}{C_\mathrm{e}(1,0)}$, $\underline{\gamma}_t \leq \underline{\gamma}_{t+1}$, and $\underline{\gamma}_t \leq \frac{C_\mathrm{o}}{C_\mathrm{e}(0,1)}$ for all t.

Proposition 1. Suppose that prior distribution has non-trivial entries, i.e., $p_0 \in \mathring{\Delta}(\Theta)$[‡] and taking a noisy measurement associated with the state of nature is costly, i.e., $C_\mathrm{o} > 0$. Suppose that the cost functions $C_\mathrm{e} : \Theta \times \mathcal{D} \to \mathbb{R}$ satisfies $C_\mathrm{e}(\theta, \hat{\theta}) > C_\mathrm{o}$ for all $\theta \neq \hat{\theta}$, where $C_\mathrm{e}(\theta, \hat{\theta})$ refers to the cost of declaring $\hat{\theta}$ when the true state of nature is θ. Suppose $0 < \frac{f_\theta(\zeta)}{f_{\theta'}(\zeta)} < \infty$ for all $\theta, \theta' \in \Theta$ and all $\zeta \in \mathcal{Z}$, and the random processes $\{\zeta_t\}$ are independently and identically distributed (i.i.d.) with the conditional probabilities $f_\theta(\cdot)$ for $\theta \in \Theta$. Then there exists a constant $\epsilon(C_\mathrm{o}) > 0$ such that no Bayesian sequential hypothesis testing minimizing the associated Bayes risk can achieve individual q-belief with $q \geq 1 - \epsilon(C_\mathrm{o})$.

Remark 1. Proposition 1 shows that no learning rule minimizing Bayes risk can achieve common q-belief with arbitrary high probability q when taking measurements is costly. This observation is a motivation of our work and sequential learning rules presented in this paper can be of either finite or infinite time-horizon and explicitly take costs of taking measurements and delaying choices into account.

III. Strategic Multi-Agent Learning for Synchronizing Coordination

Uncertain fundamentals are summarized by the state of nature θ and each player observes different noisy measurements of the state over time. Suppose that the noise technology, i.e., characteristics of the noisy channels available to players, is common knowledge, received signals of each player can be used to form and update beliefs about the state of nature.

A. Information states, belief states, and predictive inference

1) Information and belief states: Consider a time-invariant state of nature $\theta \in \Theta$ and the associated noisy measurements at time $t+1$

$$y^i_{t+1} = g^i_t(y^i_t, u^i_t, v^i_t, \theta),\ i \in \mathcal{P}\,,$$

where $\mathcal{P}$ denotes the set of agents, u^i refers to a sequence of actions of agent i, and v^i refers to a random process corresponding to noisy observations of agent i. Denote the information state available to agent i at time t by I^i_t: $I^i_0 = y^i_0$ and

$$I^i_t = (y^i_0, y^i_1, \cdots, y^i_t; u^i_0, u^i_1, \cdots, u^i_{t-1})\,.$$

Denote the belief state of agent i at time t by β^i_t:

$$\beta^{\theta,i}_t = P(\theta | I^i_t),\ \theta \in \Theta,\ t \geq 0\,,$$

that is the posterior probability distribution over Θ for given information I^i_t. As this paper focuses on two player sequential hypothesis testing, by abuse of notation we use the following notations: $x := y^1$, $z := y^2$, $I := I^1$, $a := u^1$, $b := u^2$, $K := I^2$, $\alpha^\theta := \beta^{\theta,1}$, and $\beta^\theta := \beta^{\theta,2}$. Assume also that $x_t \in \mathcal{X}$ and $z_t \in \mathcal{Z}$ for all $t \in \mathbb{Z}_+$ where the signal spaces $\mathcal{X}$ and $\mathcal{Z}$ are finite.

For the case of a binary state of nature $\Theta = \{0, 1\}$, at each time t, two players compute the following information and belief states:

$$I_t = (x_0, x_1, \cdots, x_t; a_0, a_1, \cdots, a_{t-1}),\ \alpha_t = P(\theta = 0 | I_t);$$
$$K_t = (z_0, z_1, \cdots, z_t; b_0, b_1, \cdots, b_{t-1}),\ \beta_t = P(\theta = 0 | K_t);$$

where $a_t, b_t \in \mathcal{A} := \{0, \mathrm{w}, 1\}$ with

$$\begin{array}{rcl} 0 & \Leftrightarrow & \text{Declare } \theta = 0\,, \\ \mathrm{w} & \Leftrightarrow & \text{Wait for another signal}\,, \\ 1 & \Leftrightarrow & \text{Declare } \theta = 1\,. \end{array}$$

Note that in our framework of sequential learning, the opponent's actions are not observable. For player 1, if the action $a_t = \mathrm{w}$, i.e., continuing observations, then the corresponding information and belief dynamics are obtained by

$$\begin{aligned} &I_{t+1}(x_{t+1}) = (I_t, x_{t+1})\,; \\ &x_{t+1} \sim \alpha_t f_0(x_{t+1}) + (1 - \alpha_t) f_1(x_{t+1})\,, \end{aligned} \tag{1}$$

and

$$\alpha_{t+1}(x_{t+1}) = \frac{\alpha_t f_0(x_{t+1})}{\alpha_t f_0(x_{t+1}) + (1 - \alpha_t) f_1(x_{t+1})}\,. \tag{2}$$

The belief and information dynamics of player 2 can be represented in the same way.

B. Cost functions and Bayes risk

1) Separable costs: Consider the following cost functions:

- Cost of taking measurements or signals:

$$C^i_\mathrm{o}(m) = \text{Cost of } m \text{ observations}$$

  For constant observation costs, $C^i_\mathrm{o}(m) = m C^i_\mathrm{o}$ with some constants C^i_o, $i \in \{1, \cdots, n\}$.
- Cost of erroneous decisions:

$$C^i_\mathrm{e}(\theta, a^i) \quad \text{for } \theta \in \Theta,\ a^i \in \mathcal{D} \equiv \Theta$$

- Cost of disagreement in the final choices:

$$C^i_\mathrm{d}(a^{-i}, a^i) \quad \text{for } a^{-i} \in \mathcal{D}^{n-1},\ a^i \in \mathcal{D}$$

  One might consider the separable costs $C^i_\mathrm{d}(a^{-i}, a^i) = \sum_{j \neq i} C^i_\mathrm{d}(a^j, a^i)$.
- Cost of time-separation (time-difference in declaration):

$$C^i_\mathrm{s}(\tau^{-i}, \tau^i) = \sum_{j \neq i} \max\left\{0, \tau^i - \tau^j\right\} C_\mathrm{s}$$

  where C_s denote a constant cost of one-step delay in declaration when someone made a declaration and τ^i refers to the stopping time of player i.

For a two-agent game of sequential costly hypothesis testing, the corresponding payoffs are represented in Fig. 1.

C. Solutions: Strategic learning

1) Sufficient statistics for decision processes: This section provides a brief introduction to sufficient statistics for decision processes. See [6, Sec. 5.4] for more details of sufficient

[‡] $\mathring{S}$ denotes the interior of a set S.

P2

P1	h_0	h_1	w
h_0	$-C_{\rm d}(0,0)$	$-C_{\rm d}(1,0)$	0
h_1	$-C_{\rm d}(0,1)$	$-C_{\rm d}(1,1)$	0
w	$-C_{\rm s}$	$-C_{\rm s}$	0

Nature

P1 (or P2)	h_0	h_1
h_0	$-C_{\rm e}(0,0)$	$-C_{\rm e}(1,0)$
h_1	$-C_{\rm e}(0,1)$	$-C_{\rm e}(1,1)$
w	$-C_{\rm o}$	$-C_{\rm o}$

Fig. 1: Two-agent statistical game for which there are two separable payoff matrices for each agent–one for interaction with the opponent and the other for interaction with nature.

statistics in Markov decision processes and see also [14, Sec. 2.9] for sufficient statistics in data and information processing. Define an optimal policy associated with some cost functions C_t by

$$\mu_t^*(I_t) := \arg\min_{u_t\in\mathcal{U}_t} \mathbf{E}[C_t(\vartheta_t, w_t, y_{t+1})|I_t, u_t]$$

where ϑ_t denotes the concatenation of unknowns, w_t denotes the concatenation of noise processes, y_{t+1} refers to the measurement at the one-step forward stage. Suppose that we can find a function $S_t(I_t)$ of the information state such that

$$\begin{aligned}&\min_{u_t\in\mathcal{U}_t} \mathbf{E}[C_t(\vartheta_t, w_t, y_{t+1})|I_t, u_t]\\ &\equiv \min_{u_t\in\mathcal{U}_t} \mathbf{E}[C_t(\vartheta_t, w_t, y_{t+1})|S_t(I_t), u_t]\ .\end{aligned}$$

Then such a function S_t is called a sufficient statistic. This implies that one can find an appropriate policy $\overline{\mu}_t$ achieving

$$\overline{\mu}_t(S_t(I_t)) \equiv \mu_t^*(I_t)$$

for every information state I_t.

2) Bayes's risk and value functions: Consider a game of two-agent sequential hypothesis testing. For player 1, consider the stage-risk function

$$R_t(I_t, a) = \begin{cases} \mathbf{E}\Big[C_{\rm e}(\theta_t, a) + \tilde{C}_{\rm s}(B_t, a, t)|I_t\Big] & \text{for } a\in 0,1,\\ C_{\rm o} & \text{for } a = {\rm w}.\end{cases} \tag{3}$$

where $B_t = (b_0, \cdots, b_t)$ denotes the sequence of actions of the opponent (player 2), $\tilde{C}_{\rm s}(B_t, a, t) := C_{\rm d}(a, b_t) + (t - \tau_t)C_{\rm s}$, and τ_t refers to the stopping-time of the opponent (player 2) for which if $b_t = {\rm w}$ then τ_t is set to be t by abuse of notation.

Consider a finite-horizon sequential hypothesis testing with horizon N.[§] At the terminal stage N, the optimal cost-to-go function is

$$J_N(I_N) = \min_{a\in\{0,1\}} R_N(I_N, a)\,. \tag{4}$$

The optimal cost-to-go function at time $t \le N-1$ is given by

$$J_t(I_t) = \min\left\{\min_{a_t\in\{0,1\}} R_t(I_t, a_t),\, C_{\rm o} + \mathbf{E}[J_{t+1}(I_{t+1})|I_t]\right\} \tag{5}$$

where $I_{t+1} = (I_t, x_{t+1})$ and the expectation is taken with respect to the random process x_{t+1} whose probability distribution is $\alpha_t f_0(\cdot) + (1-\alpha_t) f_1(\cdot)$ for the current belief state $\alpha_t := p(\theta_t = 0|I_t)$.

To compute the expectations associated with the preceding risk and optimal cost-to-go functions, the two conditional probabilities $P(\theta_t|I_t)$ and $P(B_t|I_t)$ are required. Suppose that the opponent's policy $\varphi_t(K_t) \in \mathcal{A} = \{0, 1, {\rm w}\}$ is (commonly) known. Then from predictive inference and for time-invariant $\theta \in \Theta$ and $B_t \in \mathcal{A}^{t+1}$, the belief over behaviors of the opponent can be described as

$$\begin{aligned}P(B_t|I_t) &= \sum_{K_t:\, b_t=\varphi_t(K_t)} P(K_t|I_t)\,,\\ &= \sum_{K_t:\, b_t=\varphi_t(K_t)} \sum_{\theta\in\Theta} P(K_t|I_t,\theta)P(\theta|I_t)\,,\\ &= \sum_{K_t:\, b_t=\varphi_t(K_t)} \sum_{\theta\in\Theta} \frac{P(K_t, I_t|\theta)}{P(I_t|\theta)} P(\theta|I_t)\,,\\ &= \sum_{K_{\tau_t}:\, b_{\tau_t}=\varphi_{\tau_t}(K_{\tau_t})} \sum_{\theta\in\Theta} \frac{P(K_{\tau_t}, I_t|\theta)}{P(I_t|\theta)} P(\theta|I_t)\,,\end{aligned} \tag{6}$$

where τ_t refers to the declaration-time of an action-string B_t, e.g., $B_t = ({\rm w}, {\rm w}, {\rm w}, 0, \cdots, 0)$ has $\tau_t = 3$.[¶]

3) Conditionally independent noisy measurements: For conditionally independent noisy signals, i.e., $P(K_t, I_t|\theta) = P(K_t|\theta)P(I_t|\theta)$, the conditional probability $P(B_t|I_t)$ has a compact form

$$P(B_t|I_t) = \sum_{K_{\tau_t}:\, b_{\tau_t}=\varphi_{\tau_t}(K_{\tau_t})} \sum_{\theta\in\Theta} P(K_{\tau_t}|\theta)P(\theta|I_t)\,. \tag{7}$$

Proposition 2. Suppose that private measurements/signals for agents are conditionally independent, and independently and identically distributed (i.i.d.) over time. Then the belief states are sufficient statistics for the preceding multi-agent sequential hypothesis testing.

Proof. The conditional probability $P(B_t|I_t)$ given in (7) is affinity dependent of $P(\theta|I_t)$ and not explicitly dependent of I_t, and probability distribution of x_{t+1} in (1) is fully characterized by $P(\theta|I_t)$ under the i.i.d. assumption. This implies that the optimal cost-to-go functions (4) and (5) can be rewritten as $J_N(I_N) \equiv \overline{J}_N(P(\cdot|I_N))$ where

$$\overline{J}_N(P(\cdot|I_N)) := \min_{a\in\{0,1\}} R_N(P(\cdot|I_N), a)\,, \tag{8}$$

and $J_t(I_t) \equiv \overline{J}_t(P(\cdot|I_t))$ for $t \le N-1$ where

$$\begin{aligned}&\overline{J}_t(P(\cdot|I_t))\\ &:= \min\left\{\min_{a_t\in\mathcal{D}} R_t(P(\cdot|I_t), a_t),\, C_{\rm o} + \mathbf{E}[J_{t+1}(I_{t+1})|P(\cdot|I_t)]\right\},\\ &= \min\left\{\min_{a_t\in\mathcal{D}} R_t(P(\cdot|I_t), a_t),\, C_{\rm o} + \mathbf{E}\big[\overline{J}_{t+1}(P(\cdot|I_{t+1}))|P(\cdot|I_t)\big]\right\},\end{aligned} \tag{9}$$

respectively. It follows from the fact that the predicted posterior $P(\cdot|I_N)$ at N is a function of $P(\cdot|I_{N-1})$ and x_N whose distribution is independent of I_{N-1} under the

[§]This paper focuses on finite time-horizon cases and the results can be extended to the cases of infinite time-horizon in straightforward manners.

[¶]Note that due to the consistency constraint of sequential actions, the actual cardinality of the set of B_t's is indeed $2t+3$, not 3^{t+1}.

i.i.d. assumption, but dependent of $P(\cdot|I_{N-1})$ only. By mathematical induction, the last equality can be obtained for all $t \leq N-1$. Therefore, the belief state $P(\cdot|I_t)$ at each time t is a sufficient statistic to determine the associated optimal cost-to-go function. □

Lemma 1. The value functions $\overline{J}_t : \Delta(\Theta) \to \mathbb{R}$ are concave for all $t = 0, 1, \cdots, N$.

Corollary 1. For conditionally independent measurements that are i.i.d. over time given the state of nature, a threshold policy is the best response to any policy $\varphi_t : K_t \mapsto b_t$ of the opponent.

Remark 2. The results in Corollary 1 are somewhat surprising. It implies that the best response is independent of the opponent's beliefs about the state of nature, regardless of belief-based or information-based policy the opponent uses, provided the opponent's prior belief over the state of nature is known.

Remark 3. As with single agent sequential hypothesis testing [6], from the stationarity of state of nature and the monotonicity of dynamic programming, we have a monotonically non-decreasing sequence of value functions: $\overline{J}_t(p) \leq \overline{J}_{t+1}(p)$ for all $p \in \Delta(\Theta)$ and t.

For well-defined dynamic programming solutions with finite-horizon N, it is required that

- for conditionally independent measurement cases, there exists $p \in \Delta(\Theta)$ such that
$$\overline{W}_{N-1}(p) + C_{\mathrm{o}} < \min_{a \in \mathcal{D}} R_N(p, a)\,;$$
- for correlated measurement cases, there exists $I \in \mathcal{X}^{N+1}$ such that
$$W_{N-1}(I) + C_{\mathrm{o}} < \min_{a \in \mathcal{D}} R_N(I, a)\,,$$
where $W_t(I_t) := \mathbf{E}[J_{t+1}(I_{t+1})|I_t]$, $I_{t+1} = (I_t, x_{t+1})$, and the expectation is taken with respect to the random process x_{t+1} whose probability distribution is $\alpha_t f_0(\cdot) + (1-\alpha_t) f_1(\cdot)$ for the current belief state $\alpha_t = p(\theta_t = 0|I_t)$.

4) Correlated noisy measurements: In the presence of correlations between noisy measurements of players, computations of Bayes risk become more complicated since abstraction of the information states into the belief states in lower dimensional spaces results in non-measurable quantities.

Proposition 3. The belief states are not sufficient statistics for the associated decision problem of multi-agent sequential hypothesis testing.

Proof. As previously shown, to compute the expectations associated with the risk and optimal cost-to-go functions of two player sequential hypothesis testing, the two conditional probabilities $P(\theta|I_t)$ and $P(B_t|I_t)$ are required. From (6), $P(B_t|I_t) \neq P(B_t|P(\cdot|I_t))$ for cases of correlated measurements. This implies that the associated (stage) Bayes risk (3) cannot be computed with respect to the belief state $P(\theta|I_t)$, and consequently, the equalities (8) and (9) do not hold, even under the i.i.d. assumption for noisy measurements. The information states I_t are indispensable for the corresponding decision problem and information-processing to generate the corresponding belief state $P(\theta|I_t)$ results in insufficient statistics. □

This is, roughly speaking, because the opponent's stopping-time in sequential hypothesis testing is not measurable with respect to the belief states, but are measurable only in terms of the information states. This is even for the case when the opponent's cost functions are independent of the player's actions.

IV. Application to Sequential Games of Strategic Coordination and Delay

Consider the following example taken from [15] with some modifications, for which payoffs are given by Fig. 2. Then

		P2: Invest	P2: Not Invest
P1	Invest	θ, θ	$\theta - 1$, 0
	Not Invest	0, $\theta - 1$	0, 0

Fig. 2: Consider the state of nature $\theta \in \Theta = \{-1, 2\}$.

the optimal strategies are

- for $\theta = 2$, each player has a dominant strategy to invest;
- for $\theta = -1$, each player has a dominant strategy not to invest.

Consider an option to delay agents' choices of investment and the corresponding (symmetric) payoff matrices in Fig. 3. Denote the corresponding payoff functions by $U_1, U_2 : \mathcal{A} \times \mathcal{B} \times \Theta \to \mathbb{R}$, where $\mathcal{A} = \mathcal{B} = \{\mathrm{I}, \mathrm{NI}, \mathrm{w}\}$ is the set of actions including an option to delay choices for each player.

Assumption 1. Assume that

- private signals are conditionally independent and each random signal process is i.i.d;[‖]
- information-contingent strategies $\mu_t : \mathcal{X}^{t+1} \to \mathcal{A}$ and $\nu_t : \mathcal{Z}^{t+1} \to \mathcal{B}$ are common knowledge; and
- characteristics of noisy channels $p(x_t|\theta)$ and $p(z_t|\theta)$ for $x_t \in \mathcal{X}$, $z_t \in \mathcal{Z}$ and $\theta \in \Theta$ are common knowledge.

Lemma 2. Under Assumption 1, the belief states $p(\theta|I_t)$ corresponding to $I_t = (x_0, \cdots, x_t)$ and $p(\theta|K_t)$ corresponding to $K_t = (z_0, \cdots, z_t)$ are sufficient statistics for the sequential game with the stage payoffs in Fig. 3.

5) A sequential leader-follower game of investment: Suppose that player 2 is a leader whose payoffs are independent of choices of player 1. Consider the stage payoffs in Fig. 4 for player 2. Consider the stage payoffs in Fig. 3 for player 1. The preceding leader-follower game of strategic coordination and delay in investment has the following properties:

- The leader's optimal policy is a threshold policy of single-agent sequential hypothesis testing that is presented in Section II-A. Fig. 5 shows the optimal *reward*-to-go function for a finite horizon sequential decisions

[‖] We assume finite sets of signals $\mathcal{X}$ and $\mathcal{Z}$ and such assumption can be trivially relaxed to be continuous and continuous signals do not change the results in this section, whereas existence of an equilibrium can be dependable on the characteristics of signals.

P1 \ P2	I	NI	w
I	2	1	$(\lambda_0+1)\delta$
NI	0	0	0
w	$-C_o$	$-C_o$	$-C_o$

for $\theta = 2$

P1 \ P2	I	NI	w
I	-1	-2	$(\lambda_1-2)\delta$
NI	0	0	0
w	$-C_o$	$-C_o$	$-C_o$

for $\theta = -1$

Fig. 3: A symmetric game of investment with an optional delay denoted by w: $\lambda_0, \lambda_1 \in]0,1[$ can be interpreted as probabilities that the opponent (P2) invests at the next stage with probability λ_0 when $\theta = 2$ and with probability λ_1 when $\theta = -1$. $\delta \in [0,1]$ denotes a discounting factor and C_o refers to the cost of delaying choices and taking signals.

P2 \ Nature	$\theta = 2$	$\theta = -1$
I	$\rho_0 + 1$	$\rho_1 - 2$
NI	$-\rho_0 - 1$	$-\rho_1 + 2$
w	$-C_o$	$-C_o$

Fig. 4: Payoffs of a leader (P2): $\rho_0, \rho_1 \in [0,1]$ are given constants and can be interpreted as personal (i.e., subjective) probabilities that someone made an investment before leader's participation when $\theta = 2$ and $\theta = -1$, respectively.

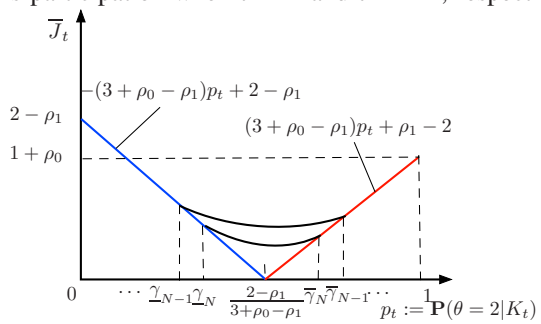

Fig. 5: Determining an optimal threshold policy for the leader with the stage payoffs in Fig. 4 and a finite time-horizon N. Note that this corresponds to sequential decisions maximizing Bayes rewards, so the value functions are convex in the belief state p_t for all t.

(of maximization) corresponding to the stage payoffs in Fig. 4. An optimal threshold policy is to choose (i) Invest (I) if $p_t > \overline{\gamma}_t$, (ii) Not Invest (NI) if $p_t < \underline{\gamma}_t$, and (iii) Wait (w) if $p_t \in [\underline{\gamma}_t, \overline{\gamma}_t]$ at time t.

- If the follower's private signals are conditionally independent of the leader's private signals, the follower's optimal policy that is the best response to the leader's threshold policy is a threshold policy, which is similar to the policy presented in Section III-C.3.

V. More Than Two Players

This paper so far has mostly focused on two-player cases. The associated results can be extended to many player cases with minuscule changes. In particular, the separable cost functions presented in Section III-B.1 with conditionally independent private noisy i.i.d. signals make such extensions trivial, since the associated predictive inference about the other players' behaviors are independently made. In contrast, for the case of correlated private noisy signals, the separable nature of cost functions is not amenable to computations– this is because the associated predictive inference about the other players' behaviors cannot be independently made and the curse of dimension is inevitable. Our future work would be to consider mean-field methods of approximation for cases of a large number of players.

VI. Conclusions

This paper presented a framework of multi-agent sequential hypothesis testing for which both temporal and spatial coordination are explicitly considered by introducing costs of taking private measurements, time-difference in stopping-time and disagreement between agents, and false declaration about the state of nature. It was shown that for conditionally independent noisy measurements, the belief states are sufficient statistics, whereas such abstraction of the information states into the belief states is not satisfactory for the case of correlated noisy measurements. We also discussed an application of the proposed sequential learning to sequential games of strategic coordination and delay for which synchronizing coordination in choices is of particular interest in investment games.

Acknowledgement

The authors acknowledge support for this work under grants AFOSR MURI projects #FA9550-09-1-0538 and #FA9550-10-1-0573, and ARO MURI project #W911NF-12-1-0509.

References

[1] J. C. Harsanyi, "Games with incomplete information played by Bayesian players, Part I. The basic model," *Management science*, vol. 14, no. 3, pp. 159–182, 1967.

[2] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[3] D. Blackwell and M. A. Girshick, *Theory of games and statistical decisions*. New York: John Wiley and Sons, 1954.

[4] M. H. DeGroot, *Optimal Statistical Decisions*. Hoboken, NJ: John Wiley & Sons, 2004.

[5] P. Whittle, *Optimization over time*. John Wiley & Sons, Inc., 1982.

[6] D. P. Bertsekas, *Dynamic programming and optimal control 3rd edition, volume I*. Belmont, MA: Athena Scientific, 2005.

[7] D. Monderer and D. Samet, "Approximating common knowledge with common beliefs," *Games and Economic Behavior*, vol. 1, no. 2, pp. 170–190, 1989.

[8] M. W. Cripps, J. C. Ely, G. J. Mailath, and L. Samuelson, "Common learning," *Econometrica*, vol. 76, no. 4, pp. 909–933, 2008.

[9] S. Morris and H. S. Shin, "Global games: theory and applications," *Econometric Society Monographs*, vol. 35, pp. 56–114, 2003.

[10] D. Gale, "Dynamic coordination games," *Economic theory*, vol. 5, no. 1, pp. 1–18, 1995.

[11] A. Dasgupta, J. Steiner, and C. Stewart, "Dynamic coordination with individual learning," *Games and Economic Behavior*, vol. 74, no. 1, pp. 83–101, 2012.

[12] A. Dasgupta, "Coordination and delay in global games," *Journal of Economic Theory*, vol. 134, no. 1, pp. 195–225, 2007.

[13] D. P. Bertsekas, *Dynamic programming and optimal control 4th edition, volume II*. Belmont, MA: Athena Scientific, 2012.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[15] H. Carlsson and E. v. Damme, "Global games and equilibrium selection," *Econometrica*, vol. 61, no. 5, pp. 989–1018, 1993.