Quickest Detection POMDPs With Social Learning: Interaction of Local and Global Decision Makers

Vikram Krishnamurthy, *Fellow, IEEE*

***Abstract*—We consider how local and global decision policies interact in stopping time problems such as quickest time change detection. Individual agents make myopic local decisions via social learning, that is, each agent records a private observation of a noisy underlying state process, selfishly optimizes its local utility and then broadcasts its local decision. Given these local decisions, how can a global decision maker achieve quickest time change detection when the underlying state changes according to a phase-type distribution? This paper presents four results. First, using Blackwell dominance of measures, it is shown that the optimal cost incurred in social-learning-based quickest detection is always larger than that of classical quickest detection. Second, it is shown that in general the optimal decision policy for social-learning-based quickest detection is characterized by multiple thresholds within the space of Bayesian distributions. Third, using lattice programming and stochastic dominance, sufficient conditions are given for the optimal decision policy to consist of a single linear hyperplane, or, more generally, a threshold curve. Estimation of the optimal linear approximation to this threshold curve is formulated as a simulation-based stochastic optimization problem. Finally, this paper shows that in multiagent sensor management with quickest detection, where each agent views the world according to its prior, the optimal policy has a similar structure to social learning.**

***Index Terms*—Adaptive sensing, Blackwell dominance, multiagent sensor scheduling, partially observed Markov decision process (POMDP), phase-type distribution, quickest time Bayesian change detection, social learning, stochastic dominance.**

I. Introduction

CLASSICAL Bayesian quickest time detection [43], [44] involves detecting a geometrically distributed change time by optimizing the tradeoff between false alarm frequency and delay penalty. The literature is vast, with applications in biomedical signal processing, machinery monitoring, and finance [39], [8], [34], [44]; see also [37] for team detection, and [52] and [53]. Classical quickest detection can be formulated as the following sequential protocol involving a countable number of agents: Suppose each agent acts once in a predetermined sequential order indexed by $k = 1, 2, \ldots$. Agent k receives an observation of the underlying state at time k and computes the posterior probability that the state has changed. It then reveals this posterior probability to subsequent agents. This process repeats until a stopping time at which the global decision maker announces a change. It is well known [43], [44] that the optimal policy to declare a change has a threshold (monotone) structure: if the posterior probability (belief state) exceeds a threshold, then a change is announced; otherwise, agents continue making observations.

A. Context

Motivated by understanding how local decisions affect global decision making in multiagent systems, this paper considers a generalization of the above classical quickest detection setup. Given local decisions from agents that are performing social learning, how can a global decision maker achieve quickest time change detection? In other words, how can a stochastic control problem (stopping time problem) be solved to make global decisions based on local decisions of agents? We consider phase-type distributed change times and interaction between local and global decision makers as outlined in the following two examples:

Example 1. Social-Learning-Based Quickest-Time Detection: Suppose that a multiagent system performs social learning [12][1] to estimate an underlying state as follows: Just as in the classical quickest detection protocol above, agents act sequentially in a predetermined order. However, instead of revealing its posterior distribution of change, each agent reveals its local decision to subsequent agents. The agent chooses its local decision by optimizing a local utility function (which depends on the public belief of the state and its local observation). Subsequent agents update their public belief based on these local decisions (in a Bayesian setting), and the sequential procedure continues. Given these local decisions, how can such a multiagent system detect a change in the underlying state and make a global decision to stop?

Example 2: Quickest-Time Detection With Adaptive Sensing: Consider a multisensor system where each adaptive sensor is equipped with a local sensor manager (controller). The multisensor system acts sequentially as follows: Based on the existing belief of the underlying state, the local sensor manager chooses (adapts) the sensor mode, e.g., low resolution or high resolution. The sensor then views the world based on this mode. Given the belief states and local sensor-manager decisions, how can such a multiagent system achieve quickest time change

Manuscript received July 02, 2010; revised August 26, 2011; accepted February 08, 2012. Date of publication May 25, 2012; date of current version July 10, 2012. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

The author is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vikramk@ece.ubc.ca).

Communicated by M. Lops, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2012.2201372

[1]Another way of viewing the social learning model is that there are finite number of agents that act repeatedly in some predefined order. If each agent picks its local decision using the current public belief, then the setup is identical to the social learning setup. We also refer reader to [1] and [2] for several recent results in social learning over several types of network adjacency matrices.

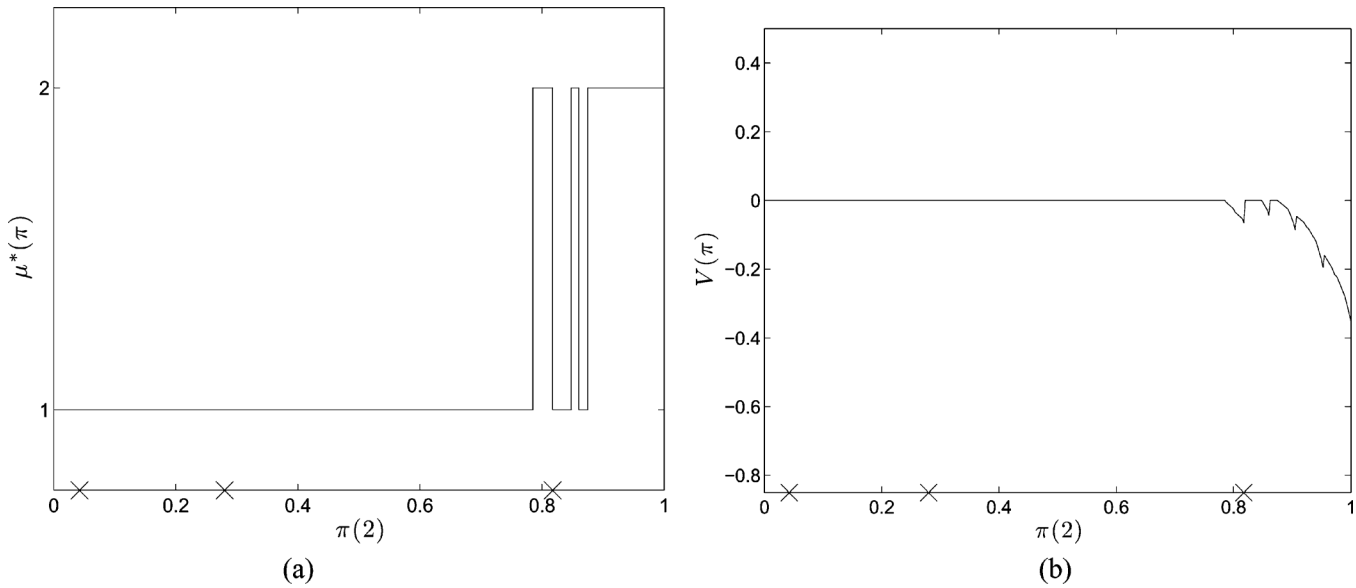

Fig. 1. Optimal decision policy for social-learning-based quickest time change detection for geometric distributed change time, see Example 1 of Section VII for details. (a) Optimal policy $\mu^*(\pi)$ is characterized by a triple threshold. (b) Value function $V(\pi)$ is nonconcave and discontinuous.

detection?[2] Quickest detection with such sensor management is of importance in automated tracking and surveillance systems [3], [5], [14]. In such cases, if individual agents or cluster heads are polled sequentially (e.g., round-robin fashion), then the resulting dynamics are very similar to the social learning setup.

Classical quickest detection is a trivial case of the above examples where agents reveal their local observation (instead of local decision) to subsequent agents. The above examples are nontrivial generalizations due to the interaction of the local and global decision makers.[3] In both examples, the local decision determines the belief state which determines the global decision (stop or continue) which determines the local decision at the next time instant and so on. This interaction of local and global decision making leads to discontinuous dynamics for the posterior probabilities (belief state) and unusual behavior as outlined below. We will show that the optimal decision policy has multiple thresholds and the stopping regions are nonconvex.

Fig. 1(a) gives a visual description of the optimal policy of social-learning-based quickest detection. It illustrates a triple threshold policy for geometric distributed change time. Complete details of this numerical example are given in Section VII. The horizontal axis $\pi(2)$ is the posterior probability of no change. The vertical axis denotes the optimal decision: $u = 1$ denotes stop and declare change, while $u = 2$ denotes continue. The multithreshold behavior of Fig. 1(a) is unusual: if it is optimal to declare a change for a particular posterior probability, it may not be optimal to declare a change when the posterior probability of a change is larger! Thus, the global decision (stop or continue) is a nonmonotone function of the posterior probability obtained from local decisions. Fig. 1(b) shows the associated value function obtained via stochastic dynamic programming. Unlike standard sequential detection problems where the value function is concave, the figure shows that the value function is nonconcave and discontinuous. To summarize, Fig. 1 shows that social-learning-based quickest detection results in fundamentally different decision policies compared to classical quickest time detection (which has a single threshold). Thus, making global decisions (stop or continue) based on local decisions (from social learning) is nontrivial.

B. Motivation and Related Works

Social Learning: In the last decade, social learning has been studied widely in economics to model the behavior of financial markets, crowds and social networks, see [1], [2], [12], [46], [31], and numerous references therein. Hellman's and Cover's seminal papers [15], [19] analyze learning with limited memory. [12, Chs. 3 and 4] gives an excellent exposition of social learning. An important result in social learning [6], [10] is that if the underlying state is a random variable and the observation and local decision spaces are finite, then agents eventually herd and end up making the same local decision irrespective of their observation. Such information cascades have been used in [12] to model sequences of financial trades, crashes and booms, and auctions. There is strong motivation to understand the interaction of local and global decision makers in social learning. Global decision making with social learning has recently been studied by several economists; for example, in [11], [12], [45], and [26], the authors describe how information externalities affect global and local decision making in social learning. This paper can be viewed as addressing a related problem: if individual agents make (simple) decisions by optimizing a local utility, how can the global system achieve

[2]The information flow patterns of Example 1 and 2 are similar. In Example 1, the sequence of events is $\text{prior} \rightarrow \text{observation} \rightarrow \text{local decision} \rightarrow \text{posterior}$. In Example 2, the sequence of events is $\text{prior} \rightarrow \text{local decision} \rightarrow \text{observation} \rightarrow \text{posterior}$.

[3]A signal-processing interpretation of social learning is as follows. Instead of using the posterior distribution to achieve quickest time detection, the decision maker (or individual agents) computes the maximum *a posteriori* (MAP) estimate of the underlying state at each time instant. Given these hard decision MAP state estimates (local decisions), how can the global decision maker achieve quickest change detection?

the (complex) task of detecting a change. In a non-Bayesian setting such problems of designing sophisticated global behavior given simple local behavior have also been studied in game-theoretic learning [18], [17], [27] involving correlated equilibria and global games.

PH-Distributed Change Time: This paper deals with quickest detection for PH-distributed change times. PH-distributions are used widely in queuing theory [36] and include geometric distributions as a special case. The optimal detection of a PH-distributed change point is useful since the family of all PH-distributions forms a dense subset for the set of all distributions, i.e., for any given distribution function F such that $F(0) = 0$, one can find a sequence of PH-distributions $\{F_n, n \geq 1\}$ to uniformly approximate F over $[0, \infty)$; see [36]. Therefore there is strong motivation to analyze quickest detection with PH-distributed change times and social learning. Quickest time change detection for PH-distributed change times is analyzed in [26]. The current paper generalizes these results to include social learning. A systematic investigation of the statistical properties of PH-distributions can be found in [36].

C. Main Results and Organization

This paper deals with characterizing the structure of the global quickest-time change detection policy in multiagent systems where individual agents make local myopic decisions when performing social learning. The main results and organization of the paper are as follows.

1) *Multiagent Protocol*: Section II presents the multiagent social learning protocol. The quickest time detection problem is formulated. We also point out in (21) the difference between the social learning model and the classical Kolmogorov–Shiryaev model for quickest change detection.
2) *Dynamic Programming Formulation and Dominance of Classical Detection:* In Section III, the optimal stopping policy is characterized in terms of stochastic dynamic programming. It is shown that the value function is in general nonconcave. Also Theorem 1 uses Blackwell ordering of measures to show that the optimal cost incurred in social-learning-based quickest detection is always larger than classical quickest detection. The result is intuitive since decision making using social learning is based on less information than classical quickest detection.
3) *Main assumptions and Multithreshold Policies:* Section IV starts with the main assumptions required to analyze the structure of the optimal quickest detection policy. These assumptions allow us to decompose the belief space into polytopes (see Theorem 2). On each of these polytopes, the conditional probability of a local decision given the underlying state and posterior distribution is a constant.
   The main result of Section IV is to characterize quickest time change detection policies when the probability of change, denoted ϵ, is small. When the probability of change equal to zero, Theorem 3 characterizes explicitly the multithreshold structure of the optimal decision policy and nonconcave behavior of the value function for sequential detection of a fixed state. Then, Corollary 1 shows that the optimal quickest-time detection policy for change probability ϵ, yields a cost that is within $O(\epsilon)$ of the optimal cost for zero change probability. An important ingredient in the proof of this result is characterization of fixed points of the social learning filter update (see Lemma 2) which also characterizes regions where the agents form information cascades in social learning.
4) *Phase-Type Distributed Change Times:* The next main result is to is to characterize the optimal policy of the global decision maker to achieve quickest time detection when the change time has a *phase-type* (PH) distribution and individual agents are performing social learning. As mentioned previously, PH-distributions can approximate arbitrary distributions and so are widely used in discrete-event systems.
   A PH-distributed change time can be modeled as a multistate Markov chain with an absorbing state, see [26] and also [36] for a systematic description. (For a two-state Markov chain, the PH-distribution specializes to the geometric distribution). So for quickest time detection with PH-distributed change time, the belief states (Bayesian posterior) lie in a multidimensional simplex of probability mass functions.
   Under what conditions will there exist a threshold stopping policy for quickest detection with PH-distributed change time and social learning? Under what conditions for the geometric change time case does the optimal policy coincide with the classical Kolmogorov–Shiryaev model?
   To answer these questions, the main results of Section V are as follows.
   a) Theorem 4 gives sufficient conditions under which the optimal decision policy for the global decision maker is myopic and characterized by a linear threshold hyperplane in the multidimensional simplex. For the geometric case, this results yields an identical threshold to the Kolmogorov–Shiryaev model.
   b) Theorem 5 gives sufficient conditions so that the optimal decision policy is characterized by a single switching curve in the multidimensional simplex. The result uses lattice programming [49] and structural results involving monotone likelihood ratio (MLR) stochastic orders [40], [28], and a novel modification of it. The result is useful because it implies that the global decision to stop can be implemented efficiently at each agent. Each agent simply needs to compare its belief state with respect to the threshold curve (in terms of an MLR partial order on the space of posterior distributions). Theorem 7 gives sufficient conditions on the optimal linear approximation to this curve that preserves the MLR increasing structure of the optimal decision policy. This linear approximation can be estimated via simulation based stochastic optimization.
5) *Multiagent Quickest Time Detection With Active Sensing:* Section VI considers multiagent quickest time detection outlined in Example 2 above. We show that the optimal policy is similar to that in social-learning-based quickest detection.

II. Social Learning Model and Protocol for Quickest Time Detection

In this section, the multiagent social learning model is presented in Section II-A. This constitutes the *local* decision-making framework for estimating an underlying state. Then, Section II-B formulates the costs incurred by the *global* decision maker in quickest time detection. Section II-C presents the global quickest time detection objective. Finally, Section II-D summarizes the entire social learning quickest detection model.

A. Multiagent Social Learning Model

Consider a countably infinite number of agents[4] performing social learning to estimate an underlying state process x. Each agent acts once in a predetermined sequential order indexed by $k = 1, 2, \ldots$. The index k can also be viewed as the discrete time instant when agent k acts.

Let $y_k \in \mathbb{Y} = \{1, 2, \ldots, Y\}$ denote the local (private) observation of agent k and $a_k \in \mathbb{A} = \{1, 2, \ldots, A\}$ denote the local decision agent k takes. Define the sigma algebras:

$$\begin{aligned} \mathcal{H}_k &\quad \sigma\text{-algebra generated by } (a_1, \ldots, a_{k-1}, y_k) \\ \mathcal{G}_k &\quad \sigma\text{-algebra generated by } (a_1, \ldots, a_{k-1}, a_k). \end{aligned} \tag{1}$$

The social learning model [10], [12] comprises the following ingredients.

1) *Absorbing-State Markov Chain and Phase-Type Distribution Change Times:* The state x_k represents the underlying process that changes at time τ^0. We model the change point τ^0 by a phase-type (PH) distribution. As mentioned in Section I, PH-distributions form a dense subset for the set of all distributions [36] and so can be used to approximate change times with arbitrary distribution. This is done by constructing a multistate Markov chain as follows: Assume the underlying state x_k evolves as a Markov chain on the finite state space $\mathbb{X} = \{1, \ldots, X\}$. Here state '1' is an absorbing state and denotes the state after the jump change. The states $2, \ldots, X$ can be viewed as a single composite state that x resides in before the jump.
The initial distribution is $\pi_0 = (\pi_0(i), i \in \mathbb{X})$, $\pi_0(i) = P(x_0 = i)$. We are only interested in the case where the change occurs after a least one measurement, so assume $\pi_0(1) = 0$. So the transition probability matrix P is of the form

$$P = \begin{bmatrix} 1 & 0 \\ \underline{P}_{(X-1)\times 1} & \bar{P}_{(X-1)\times(X-1)} \end{bmatrix} \tag{2}$$

Let the "change time" τ^0 denote the time at which x_k enters the absorbing state 1, i.e.,

$$\tau^0 = \inf\{k : x_k = 1\}. \tag{3}$$

The distribution of the change time τ^0 is equivalent to the distribution of the absorption time to state 1 and is given by

$$\nu_0 = \pi_0(1), \quad \nu_k = \bar{\pi}_0' \bar{P}^{k-1} \underline{P}, \quad k \geq 1 \tag{4}$$

where $\bar{\pi}_0 = [\pi_0(2), \ldots, \pi_0(X)]'$. So by appropriately choosing the pair (π_0, P) and state space dimension X, one can approximate any given discrete distribution on $[0, \infty)$ by the distribution $\{\nu_k, k \geq 0\}$; see [36, 240–243]. To ensure that τ^0 is finite, we assume states $2, 3, \ldots, X$ are transient. In the special case when x is a two-state Markov chain, the change time τ^0 is geometrically distributed.

2) *Local Observation*: Agent's k local (private) observation $y_k \in \mathbb{Y} = \{1, \ldots, Y\}$ is obtained from the observation likelihood distribution

$$B_{xy} = P(y_k = y | x_k = x). \tag{5}$$

The states $2, 3, \ldots, X$ are fictitious and are defined to generate the PH-distributed change time τ^0. So states $2, 3, \ldots, X$ are indistinguishable in terms of the observation y. That is, $P(y|2) = P(y|3) = \cdots = P(y|X)$ for all $y \in \mathbb{Y}$.

3) *Private Belief:* Using local observation y_k, agent k updates its private belief π_k^P defined as

$$\begin{aligned} \pi_k^P &= (\pi_k^P(i),\ i \in \mathbb{X}) \\ \pi_k^P(i) &= \mathbb{E}\{I(x_k = i)|\mathcal{H}_k\} = P(x_k = i | a_1, \ldots, a_{k-1}, y_k) \\ &\quad \text{initialized by } \pi_0. \end{aligned} \tag{6}$$

Thus, the private belief is the posterior distribution of the underlying state given the past local decisions and current observation. It is computed by agent k according to the following hidden Markov model (HMM) filter:

$$\begin{aligned} \pi_k^P &= T(\pi_{k-1}, y_k), \text{ where } T(\pi, y) = \frac{B_y P' \pi}{\sigma(\pi, y)} \\ \sigma(\pi, y) &= \mathbf{1}' B_y P' \pi \\ B_y &= \operatorname{diag}(B_{1y}, \ldots, B_{Xy}) \\ &\quad (X \times X \text{ diagonal matrix for each } y \in \mathbb{Y}) \end{aligned} \tag{7}$$

Also π_{k-1} denotes the public belief available at time $k-1$ (defined in Step 5 below).

4) *Agent's Local Decision:* Agent k then makes local decision $a_k \in \mathbb{A} = \{1, 2, \ldots, A\}$ to minimize myopically its expected cost. To formulate this, let $c(i, a)$ denote the nonnegative cost incurred if the agent picks local decision a when the underlying state is $x = i$. Denote the local decision X-dimensional cost vector

$$c_a = [\, c(1, a) \quad c(2, a) \quad \cdots \quad c(X, a).\,]. \tag{8}$$

Then, agent k chooses local decision a_k greedily to minimize its expected cost:

$$\begin{aligned} a_k &= a(\pi_{k-1}, y_k) = \arg\min_{a \in \mathbb{A}} \mathbb{E}\{c(x, a)|\mathcal{H}_k\} \\ &= \arg\min_{a \in \mathbb{A}} \{c_a' \pi_k^P\}. \end{aligned} \tag{9}$$

In quickest change detection, since states $2, 3, \ldots, X$ are indistinguishable in terms of observation y, we assume that $c(2, a) = c(3, a) = \cdots = c(X, a)$ for each $a \in \mathbb{A}$.

5) *Social learning Public Belief:* Finally, agent k broadcasts its local decision a_k. Subsequent agents $\bar{k} > k$ use deci-

[4] As mentioned earlier, the same setup holds if a finite number of agents are polled repeatedly in some predefined order, providing each agent picks its local decision based on the most recent public belief.

sion a_k to update their public belief of the underlying state x_k as follows: Define the public belief π_k as the posterior distribution of the state x given all local decisions taken up to time k

$$\begin{aligned}\pi_k &= \mathbb{E}\{x_k|\mathcal{G}_k\} = (\pi_k(i),\ i \in \mathbb{X}), \quad \pi_k(i) \\ &= P(x = i|a_1, \dots a_k), \quad \text{initialized by } \pi_0. \end{aligned} \tag{10}$$

Then, agents $\bar{k} > k$ update their public belief according to the following "social learning Bayesian filter":

$$\begin{aligned}\pi_k &= T^{\pi_{k-1}}(\pi_{k-1}, a_k), \text{where } T^{\pi}(\pi, a) = \frac{R_a^{\pi} P' \pi}{\sigma(\pi, a)},\ \sigma(\pi, a) \\ &= \mathbf{1}_X' R_a^{\pi} P' \pi. \end{aligned} \tag{11}$$

We use the notation $T^{\pi}(\cdot)$ to point out that the above Bayesian update map depends explicitly on the belief state π. [For notational simplicity, we have chosen not to use the superscript π for $\sigma(\pi, a)$]. This is a key difference compared to the HMM filter (7) where the Bayesian update map $T(\cdot)$ does not depend explicitly on belief state π. In (11), R_a^{π} denotes the diagonal matrix $R_a^{\pi} = \text{diag}(R_{i,a}^{\pi},\ i \in \mathbb{X})$ where

$$R_{i,a}^{\pi} = P(a_k = a|x_k = i, \pi_{k-1} = \pi) \tag{12}$$

denotes the conditional probability that agent k chose local decision a given state i. We call $R_{i,a}^{\pi}$ as the local decision likelihood probabilities in analogy to observation likelihood probabilities B_{iy} (5) in classical filtering.

Clearly observing the local decision a_k taken by agent k yields information about its local observation y_k. That is, a_k serves as a surrogate observation of the underlying state x_k. The following lemma summarizes how subsequent agents use a_k to compute the local decision likelihood probabilities R_{ia}^{π} in the social learning filter. The proof is straightforward and omitted.

Lemma 1: The local decision likelihood probability matrix R^{π} in the social learning Bayesian filter (11) is computed as

$$\begin{aligned} R^{\pi} &= BM^{\pi} \text{ where} \\ M_{y,a}^{\pi} &\stackrel{\triangle}{=} P(a|y, \pi) \\ &= \prod_{\tilde{a} \in \mathbb{A} - \{a\}} I(c_a' B_y P' \pi < c_{\tilde{a}}' B_y P' \pi). \end{aligned} \tag{13}$$

Here, R^{π} is a $\mathbb{Y} \times \mathbb{A}$ matrix; B, B_y are the private observation probabilities defined in (5) and (7); c_a, $c_{\tilde{a}}$ are the local cost vectors defined in (8); and $I(\cdot)$ denotes the indicator function. ■

The main implication of Lemma 1 is that the social learning Bayesian filter (11) is discontinuous in the belief state π, due to the presence of indicator functions in (13). The likelihood probabilities R^{π} in (12) are an explicit function of the belief state π—this is stark contrast to the standard quickest detection problems where the observation distribution is not an explicit function of the posterior distribution.

Summary: A key aspect of the information pattern in the above social learning protocol is that agent k does not have access to the private belief state π_{k-1}^P or private observations of previous agents. Instead each agent k only has access to the local decisions taken by previous agents together with its own current private observation y_k. The fact that the likelihood probabilities R^{π} is an explicit function of the public belief state π [see (13)] is an important aspect of social learning that is not present in classical sequential detection problems. It makes the Bayesian update of the public belief discontinuous with π and makes our proofs substantially harder than standard concavity arguments in classical quickest detection problems.

Belief State Space: Before proceeding with the quickest time detection formulation, we briefly describe the space in which the public belief π defined in (10) lives. The public belief belongs to the unit $X - 1$ dimensional simplex denoted as

$$\begin{aligned}&\Pi(X) \stackrel{\triangle}{=} \\ &\left\{\pi \in \mathbb{R}^X : \mathbf{1}_X' \pi = 1, \quad 0 \le \pi(i) \le 1 \text{ for all } i \in \mathbb{X}\right\}. \end{aligned} \tag{14}$$

So for geometric-distributed change times, the belief state space $\Pi(2)$ is the interval $[0, 1]$. For PH-distributed change times, the belief space $\Pi(X)$ is a multidimensional simplex. For example, $\Pi(3)$ is a 2-D unit simplex (equilateral triangle); $\Pi(4)$ is a tetrahedron, etc. The vertices of the unit simplex $\Pi(X)$ are the unit X-dimensional vectors $e_1, \dots, e_X$, where

$$\begin{aligned}&e_i \text{ denotes the unit vector with 1 in the } i\text{th position,} \\ &\qquad\qquad i \in \mathbb{X}. \end{aligned} \tag{15}$$

Of course the private belief π^P(6) also lives in $\Pi(X)$.

B. Quickest Time Detection: Costs Incurred by Global Decision Maker

With the above social-learning-based local decision framework, we now formulate the quickest time detection problem faced by the global decision maker. At each time k, given the public belief π_k, let u_k denote the global decision taken:

$$u_k = \mu(\pi_k) \in \{1\ (\text{announce change and stop}), 2\ (\text{continue})\}. \tag{16}$$

Thus, the global decision u_k is $\mathcal{G}_k$ measurable, where $\mathcal{G}_k$ is defined in (1). In (16), the policy μ belongs to the class of stationary decision policies denoted μ. Below we formulate the costs incurred when taking these global decisions u_k.

1) *Cost of announcing change and stopping:* If global decision $u_k = 1$ is chosen, then the social learning protocol of Section II-A terminates. If $u_k = 1$ is chosen before the change point τ^0, then a false alarm penalty is incurred. The false alarm event $\cup_{i \ge 2}\{x_k = i\} \cap \{u_k = 1\} = \{x_k \ne 1\} \cap \{u_k = 1\}$ represents the event that a change is announced before the change happens at time τ^0. To evaluate the false alarm penalty, let $f_i I(x_k = i, u_k = 1)$ denote the cost of a false alarm in state i, $i \in \mathbb{X}$, where $f_i \ge 0$. Of course, $f_1 = 0$ since a false alarm is only incurred if the stop action is picked in states $2, \dots, X$. The expected false alarm penalty is

$$\begin{aligned}\bar{C}(\pi_k, u_k = 1) &= \sum_{i \in \mathbb{X}} f_i \mathbb{E}\{I(x_k = i, u_k = 1)|\mathcal{G}_k\} = \mathbf{f}'\pi_k \\ &\text{where } \mathbf{f} = (f_1, \dots, f_X)',\ f_1 = 0. \end{aligned} \tag{17}$$

The false alarm vector $\mathbf{f}$ is chosen with increasing elements so that states further from state 1 incur larger penalties. (Obviously $f_i \ge 0$ since $f_1 = 0$).

2) *Delay cost of continuing:* If global decision $u_k = 2$ is taken then the social learning protocol of Section II-A continues to time $k+1$. A delay cost is incurred when the event $\{x_k = 1, u_k = 2\}$ occurs, i.e., no change is declared at time k, even though the state has changed at time k. The expected delay cost is

$$\bar{C}(\pi_k, u_k = 2) = d\,\mathbb{E}\{I(x_k = 1, u_k = 2)|\mathcal{G}_k\} = de_1'\pi_k \quad (18)$$

where $d > 0$ denotes the delay cost and e_1 is defined in (15).

Remarks:

1) Recall that the public belief state π depends on the local decisions a. Also the choice of global decision u determines when the local decision process terminates. This links the local and global decision makers.
2) The above costs (17), (18) should be viewed as an example only. The results of this paper also apply to more general stopping time problems with minor modifications if the global decisions u_k are $\mathcal{H}_k$ measurable (instead of $\mathcal{G}_k$ measurable), where $\mathcal{H}_k$ and $\mathcal{G}_k$ are defined in (1). More generally, $\bar{C}(\pi, u)$ can also include the local decision cost incurred in social learning, see remark at the end of Section V-B.

C. Quickest Time Detection Objective

Let $(\Omega, \mathcal{F})$ be the underlying measurable space where $\Omega = (\mathbb{X} \times \mathbb{U} \times \mathbb{Y})^\infty$ is the product space, which is endowed with the product topology and $\mathcal{F}$ is the corresponding product sigma-algebra. For any $\pi_0 \in \Pi(X)$, and policy $\mu \in \mu$, there exists a (unique) probability measure $\mathbb{P}^\mu_{\pi_0}$ on $(\Omega, \mathcal{F})$; see [20] for details. Let $\mathbb{E}^\mu_{\pi_0}$ denote the expectation with respect to the measure $\mathbb{P}^\mu_{\pi_0}$.

Let τ denote a stopping time adapted to the sequence of σ-algebras $\mathcal{G}_k$, $k \geq 1$, see (1). That is, with u_k determined by decision policy (16)

$$\tau = \{\inf k : u_k = 1\}. \quad (19)$$

For each initial distribution $\pi_0 \in \Pi(X)$, and policy μ, the following cost is associated:

$$J_\mu(\pi_0) = \mathbb{E}^\mu_{\pi_0}\{\sum_{k=1}^{\tau-1} \rho^{k-1}\bar{C}(\pi_k, u_k = 2) + \rho^{\tau-1}\bar{C}(\pi_\tau, u_\tau = 1)\}. \quad (20)$$

Here $\rho \in [0,1]$ denotes an economic discount factor. Since $\bar{C}(\pi, 1)$, $\bar{C}(\pi, 2)$ are nonnegative and bounded for all $\pi \in \Pi(X)$, stopping is guaranteed in finite time, i.e., τ is finite with probability 1 for any $\rho \in [0,1]$ (including $\rho = 1$).

Kolmogorov–Shiryaev criterion: Suppose $\mathbb{X} = \{1, 2\}$ implying that the change time τ^0 is geometrically distributed. Choose the false alarm vector $\mathbf{f} = f_2 e_2 = [0, f_2]'$ where f_2 is a positive constant, delay cost (18), and discount factor $\rho = 1$. Then the quickest time objective (20) assumes the classical Kolmogorov–Shiryaev criterion for detection of disorder [43]

$$J_\mu(\pi_0) = d\mathbb{E}^\mu_{\pi_0}\{(\tau - \tau^0)^+\} + f_2\,\mathbb{P}^\mu_{\pi_0}(\tau < \tau^0). \quad (21)$$

However, unlike classical quickest detection, the posterior (public belief) π has discontinuous dynamics given by the social learning Bayesian filter (11). (Recall from (11) and (13) that the dynamics of public belief π depend on the local decision costs c_a). ■

The goal of the global decision maker is to determine the change time τ^0 with minimal cost, that is, compute the optimal global decision policy $\mu^* \in \mu$ to minimize (20), where

$$J_{\mu^*}(\pi_0) = \inf_{\mu \in \mu} J_\mu(\pi_0).$$

The existence of an optimal stationary policy μ^* follows from [9, Prop. 1.3, Ch. 3].

D. Summary

In summary, the social-learning-based quickest detection problem with PH-distributed change time is specified by the model

$$(P, B, c, C, \rho, \mathbb{X}, \mathbb{Y}, \mathbb{A}, \mathbf{u}) \quad (22)$$

where P is the transition probability matrix (2), B is the private observation matrix (5), c are the local decision costs (8), C defined in (24) is the transformed global decision cost vector for quickest detection [in terms of false alarm $\mathbf{f}$(17) and delay penalty d(18)], and $\rho \in [0,1]$ is the discount factor (20). Also $\mathbb{X}$ is the state space, $\mathbb{Y}$ is the private observation space, $\mathbb{A}$ is the local decision space and $\mathbf{u} = \{1\ (\text{stop}),\ 2(\text{continue})\}$ is the global decision space.

III. Stochastic Dynamic Programming Formulation and Dominance of Classical Quickest Detection

Section III-A formulates the optimal decision policy for social-learning-based quickest detection as the solution of a stochastic dynamic programming problem. Section III-B describes why social-learning-based quickest detection is a nontrivial extension of the standard quickest detection problem. Finally, Section III-C presents our first structural result—it uses Blackwell dominance of measures to show that optimal cost incurred in quickest time detection with social learning is always larger than that with classical quickest detection.

A. Stochastic Dynamic Programming Formulation

Given the stopping time problem (20), it is well known [33] that the optimal policy $\mu^*(\pi)$ can be expressed as the solution of a stochastic dynamic programming problem in terms of the belief state π. Our characterization of the structure of the optimal policy $\mu^*(\pi)$ will be based on analyzing the structure of this dynamic programming problem.

The optimal stationary policy $\mu^* : \Pi(X) \to \{1, 2\}$ and associated value function $\bar{V}(\pi)$ of the stopping time problem (20) are the solution of "Bellman's dynamic programming equation"

$$\begin{aligned}
&\mu^*(\pi) \\
&= \arg\min\{\bar{C}(\pi, 1), \bar{C}(\pi, 2) + \rho \sum_{a \in \mathbb{A}} \bar{V}\left(T^\pi(\pi, a)\right)\sigma(\pi, a)\}, \\
&J_{\mu^*}(\pi_0) = \bar{V}(\pi_0) \\
&\bar{V}(\pi) \\
&= \min\{\bar{C}(\pi, 1),\ \bar{C}(\pi, 2) + \rho \sum_{a \in \mathbb{A}} \bar{V}\left(T^\pi(\pi, a)\right)\sigma(\pi, a)\}.
\end{aligned} \quad (23)$$

Here, the global decision maker's costs $\bar{C}(\pi,u)$ are defined in (17) and (18), T^π is the public belief Bayesian update (11), and the measure $\sigma(\pi,a)$ is defined in (11).

For our subsequent analysis, it is convenient to rewrite Bellman's equation as follows. Define the transformed value function and global decision costs $V(\pi)$, $C(\pi,1)$ and $C(\pi,2)$ as follows:

$$\begin{aligned} &V(\pi)=\bar{V}(\pi)-\mathbf{f}'\pi, \quad C(\pi,1)=0,\\ &C(\pi,2)=\bar{C}(\pi,2)-\mathbf{f}'\pi+\rho\mathbf{f}'P'\pi=C'\pi \\ &\text{where } C\stackrel{\triangle}{=}de_1-(I-\rho P)\mathbf{f} \text{ with elements denoted as } C_j,\\ &\quad j=1,\ldots,X. \end{aligned} \tag{24}$$

Then, clearly $V(\pi)$ satisfies Bellman's dynamic programming equation

$$\begin{aligned} \mu^*(\pi)&=\arg\min_{u\in\mathbb{U}} Q(\pi,u),\ J_{\mu^*}(\pi_0)=V(\pi_0)\\ V(\pi)&=\min_{u\in\{1,2\}} Q(\pi,u)\\ \text{where } Q(\pi,2)&=C(\pi,2)+\rho\sum_{a\in\mathbb{A}} V\left(T^\pi(\pi,a)\right)\sigma(\pi,a)\\ Q(\pi,1)&=C(\pi,1)=0. \end{aligned} \tag{25}$$

The above transformation[5] is convenient since the transformed stopping cost $C(\pi,1)=0$ and $C(\pi,2)=C'\pi$ in (24) captures all the costs involved in quickest detection. Of course, the optimal policy $\mu^*(\pi)$ and hence stopping set $\mathcal{S}$ remain unchanged with this coordinate transformation. The goal for the global decision maker is to determine the optimal stopping set denoted $\mathcal{S}$. That is, $\mathcal{S}$ is the set of public belief states π for which it is optimal to declare a change and stop:

$$\begin{aligned} \mathcal{S}&=\{\pi\in\Pi(X):\mu^*(\pi)=1\}\\ &=\{\pi\in\Pi(X):C(\pi,1)\le C(\pi,2)\\ &\quad+\rho\sum_{a\in\mathbb{A}} V\left(T^\pi(\pi,a)\right)\sigma(\pi,a)\}\\ &=\{\pi\in\Pi(X):\bar{C}(\pi,1)\le\bar{C}(\pi,2)+\rho\sum_{a\in\mathbb{A}}\bar{V}(T^\pi(\pi,a))\sigma(\pi,a)\}. \end{aligned} \tag{26}$$

Value Iteration Algorithm: Let $k=1,2,\ldots$, denote iteration number (the fact that we used k previously to denote time should not result in confusion). The value iteration algorithm is a fixed point iteration of Bellman's (25) and proceeds as follows: $V_0(\pi)=-\bar{C}(\pi,1)$ and

$$\begin{aligned} V_{k+1}(\pi)&=\min_{u\in\{1,2\}} Q_{k+1}(\pi,u)\\ \mu^*_{k+1}(\pi)&=\operatorname{argmin}_{u\in\{1,2\}} Q_{k+1}(\pi,u)\quad \pi\in\Pi(X)\\ \text{where } Q_{k+1}(\pi,2)&=C(\pi,2)+\rho\sum_{a\in\mathbb{A}} V_k\left(T(\pi,a)\right)\sigma(\pi,a)\\ Q_{k+1}(\pi,1)&=C(\pi,1)=0. \end{aligned} \tag{27}$$

Let $\mathcal{B}(X)$ denote the set of bounded real-valued functions on $\Pi(X)$. Since $C(\pi,1)$, $C(\pi,2)$, $\pi\in\Pi(X)$, are bounded, the value iteration algorithm (27) will generate a sequence of lower semicontinuous value functions $\{V_k\}\subset\mathcal{B}(X)$ that will converge pointwise as $k\to\infty$ to $V(\pi)\in\mathcal{B}(X)$, the solution of Bellman's equation, see [9, Prop. 1.3, Chap 3, Vol. 2]

Since the belief state space $\Pi(X)$ in (14) is a unit simplex, the value iteration algorithm (27) does not yield a practical solution methodology for computing stopping set $\mathcal{S}$ since $V_k(\pi)$ needs to be evaluated on the continuum $\pi\in\Pi(X)$. Although Bellman's equation and the value iteration algorithm is not useful from a computational point of view, in subsequent sections, we exploit its structure to characterize the stopping set $\mathcal{S}$ in (26). We then exploit this structure to devise stochastic gradient algorithms for approximating the optimal policy μ^* and thus determining the stopping set $\mathcal{S}$.

B. Why Social-Learning-Based Quickest Detection is Nontrivial

Let us illustrate why social-learning-based quickest detection results in a nontrivial behavior. We will show in Section IV that the belief space $\Pi(X)$ can be decomposed into $Y+1$ polytopes denoted $\mathcal{P}_1,\ldots,\mathcal{P}_{Y+1}$ such that on each of these polytopes $\mathcal{P}_l$, the belief state update $T^\pi(\pi,a)=T^l(\pi,a)$. Consider the value iteration algorithm (27) which is used as a basis for mathematical induction to prove properties associated with Bellman's (25). It can be expressed as[6]

$$\begin{aligned} &V_{k+1}(\pi)\\ &\quad=\min\{C'\pi+\rho\sum_a\sum_{l=1}^{Y+1} V_k(T^l(\pi,a))\sigma(\pi,a)I(\pi\in\mathcal{P}_l),0\}\\ &\quad=\min\{C'\pi+\rho\sum_a\sum_{l=1}^{Y+1} V_k(R^l_aP'\pi)I(\pi\in\mathcal{P}_l),0\}. \end{aligned} \tag{28}$$

It should be clear from (28) that if $V_k(\pi)$ is assumed to be concave on $\Pi(X)$, $V_{k+1}(\pi)$ is not necessarily concave on $\Pi(X)$. In fact, even if $V_k(\pi)$ is assumed to be concave in just one of the polytopes, say polytope $\mathcal{P}_l$, then $V_{k+1}(\pi)$ is not necessarily concave on $\mathcal{P}_l$, since $T^l(\pi,a)$ in (28) may map two distinct belief states in polytope $\mathcal{P}_l$ to two different polytopes. As will be shown in numerical examples, in general $V(\pi)$ will be discontinuous and nonconcave.

Classical quickest detection problems are special instances of partially observed Markov decision process (POMDP) stopping time problems [26]. In POMDPs, the belief state update T^π is not an explicit function of belief state π since the observation probabilities are not an explicit function of π. For such POMDP stopping time problems, the value iteration algorithm reads[7]

$$\underline{V}_{k+1}(\pi)=\min\{C'\pi+\rho\sum_y\sum_{l=1}^{Y+1}\underline{V}_k(B_yP'\pi),0\}$$

and is to be compared with (28). Since the composition of a concave function with a linear function preserves concavity, it is easily seen that if $\underline{V}_k(\pi)$ is piecewise linear and concave, then

[5]This transformation is used in [21, p. 389] to deal with stopping time problems. As a result of this transformation, the initial condition of the value iteration algorithm is modified [see (27)].

[6]Note that from (27), $V_k(\pi)$ is positively homogeneous, that is, for any $\alpha>0$, $V_k(\alpha\pi)=\alpha V_k(\pi)$. So choosing $\alpha=\sigma(\pi,a)$ which is the denominator term of T^π in (11) yields the expression in the second equality of (28).

[7]We use the notation $\underline{V}(\pi)$ to denote the value function of the classical stopping problem. This will be defined formally in Section III-C where we will show $\underline{V}(\pi)\le V(\pi)$, i.e., quickest detection with social learning always incurs a higher optimal cost than classical quickest detection.

so is $\underline{V}_{k+1}(\pi)$. So by mathematical induction on the value iteration algorithm, and since the sequence $\{\underline{V}_k(\pi)\}$ converges pointwise (actually uniformly for POMDPs) to $\underline{V}(\pi)$, the value function $\underline{V}(\pi)$ is concave and the stopping set $\mathcal{S}$ is a convex (and therefore connected) set [32]. The key difference in the above social learning quickest detection formulation is that the local decision likelihoods R^π (13) and, therefore, social learning filter T^π are explicit and discontinuous functions of π. This results in a possibly nonconcave value function $V(\pi)$ making determining $\mathcal{S}$ nontrivial.

C. Quickest Time Detection With Social Learning is More Expensive

This section presents our first main result. We prove that quickest detection with social learning is always more expensive than classical quickest detection. In social learning, agents have access to local decisions of previous agents instead of the actual observations. Thus, one would expect intuitively that this information loss results in less efficient quickest time change detection compared to classical quickest detection. Here, we confirm this intuition. The main idea is to use Blackwell dominance of observation measures.

1) Notation: First define the optimal policy and cost in classical quickest time detection. Similar to (25), the optimal policy $\underline{\mu}^*(\pi)$ and cost $\underline{V}(\pi)$ incurred in classical quickest detection, satisfies the following Bellman's equation:

$$\begin{aligned}
\underline{\mu}^*(\pi) &= \arg\min_{u\in\mathbb{U}} \underline{Q}(\pi,u),\ \underline{J}_{\mu^*}(\pi_0) = \underline{V}(\pi_0)\\
\underline{V}(\pi) &= \min_{u\in\{1,2\}} \underline{Q}(\pi,u)\\
\text{where } \underline{Q}(\pi,2) &= C(\pi,2) + \rho \sum_{y\in\mathbb{Y}} \underline{V}\left(T(\pi,y)\right)\sigma(\pi,y)\\
\underline{Q}(\pi,1) &= C(\pi,1) = 0
\end{aligned} \tag{29}$$

Recall $T(\pi,y)$ is the HMM Bayesian filter defined in (7). Thus, the only difference between the classical and social learning quickest detection problems is the update of the belief state, namely (7) in the classical setup versus (11) in the social learning formulation.

2) Main Result: The following theorem says that if the initial belief state is chosen from any of the polytopes $\mathcal{P}_{y^*},\ldots\mathcal{P}_{Y+1}$, the optimal detection policy with social learning incurs a higher cost than classical quickest detection.

Theorem 1: Consider the social learning quickest time detection problem (P,B,c,C,ρ) in (22) and associated value function $V(\pi)$ in (25). Consider also the classical quickest detection problem with value function $\underline{V}(\pi)$ in (29). Then, for any initial belief state $\pi\in\Pi(X)$, the optimal cost incurred by classical quickest detection is smaller than that of quickest detection with social learning. That is, $\underline{V}(\pi)\le V(\pi)$.

Since the theorem holds for the case $A=Y=2$ (equal number of local decision choices and observation symbols), a naive explanation that information is lost due to using fewer symbols in $\mathbb{A}$ compared to $\mathbb{Y}$ is not true.

The proof of Theorem 1 is given in Appendix B. Recall from (13) that $R^\pi = BM^\pi$ where B and M^π are stochastic matrices. Thus, observation y with conditional distribution specified by B is said to be more informative than (Blackwell dominates) observation a with conditional distribution R^π (see [40]). The main idea in the proof is that under the assumptions of Theorem 1, the value function $\underline{V}(\pi)$ is concave for $\pi\in\Pi(X)$. Then, the result is established using Jensen's inequality together with Blackwell dominance on the Bellman's equation. value iteration algorithm proves the result.

The first instance of a similar proof using Blackwell dominance for POMDPs was given in [50], see also [40], where it was used to show optimality of certain myopic policies. Our use of Blackwell dominance in Theorem 1 is somewhat different since we are using it to compare the value functions of two different dynamic programming problems. A useful consequence of Theorem 1 is that performance analysis of standard quickest detection problems [48] readily applies to form a lower bound for the cost incurred in social-learning-based quickest detection.

IV. Assumptions and Quickest Detection With Small Change Probabilities

This section comprises two parts.

1) Section IV-A lists the main assumptions (A1), (A2), (S) which result in a natural partition of belief space $\Pi(X)$ into $Y+1$ convex polytopes with decision likelihoods R^π [defined in (13)] being a constant (with respect to π) on each polytope (see Theorem 2). These polytopes play an important role in specifying the global quickest detection policy in the rest of the paper.
2) Section IV-B considers quickest time change detection with geometric distributed change time and gives explicit conditions for the optimal policy to have a double threshold. In particular, Theorem 3 and Corollary 1 show that the optimal quickest-time detection policy for change probability ϵ yields a cost that is within $O(\epsilon)$ of the optimal cost for sequential detection of a constant state.

A. Polytope Structure and Main Assumptions

Since the public belief state $\pi\in\Pi(X)$ is continuum (see (14)), as a first step in characterizing the optimal policy $\mu^*(\pi)$, we need to understand the structure of the decision likelihood probabilities R^π defined in (13). Even though the belief state $\pi\in\Pi(X)$ is continuum, it turns out that there are only 2^Y-1 possible local decision likelihood probability matrices R^π. Let $\mathcal{Q}_l, l=1,\ldots,2^Y-1$ denote the elements of the power set of $\mathbb{Y}$ (excluding, of course, the empty set). Define the following 2^Y-1 convex polytopes $\bar{\mathcal{P}}_l, l=1,2,\ldots,2^Y-1$:

$$\bar{\mathcal{P}}_l = \left\{\pi\in\Pi(X) : \begin{cases} (c_1-c_2)'B_yP'\pi < 0, & y\in\mathcal{Q}_l\\ (c_1-c_2)'B_yP'\pi \ge 0, & y\in\mathbb{Y}-\mathcal{Q}_l\end{cases}\right\}. \tag{30}$$

Recall the local cost vectors c_a are defined in (8). Then, from (13) it follows that M^π and hence R^π is a constant on each polytope $\mathcal{Q}_l$. Specifically, for rows $y\in\mathcal{Q}_l$, $M^\pi_{y1}=1$ and for rows $y\in\mathbb{Y}-\mathcal{Q}_l$, $M^\pi_{y2}=1$.

Although in general there are 2^Y-1 possible R^π matrices, we now show that by introducing assumptions (A1), (A2) and (S) below, there are only $Y+1$ distinct local decision likelihood matrices R^π. This forms an important preliminary step for characterizing the optimal global decision policy.

Recalling the notation in Section II-A, we list the following assumptions.

(A1) The observation distribution $B_{xy} = p(y|x)$ is TP2 (see Definition 6 in Appendix A), i.e., all second order minors of matrix B are nonnegative.

(A2) The transition probability matrix P is TP2. (All second order minors of P are nonnegative).

(A3) The elements of vector C in (24) are decreasing. A sufficient condition is that for $j \geq i$ and $i \geq 2$ the false alarm vector $\mathbf{f}$ and delay penalty d satisfy $f_i \geq \max\{1, \rho \mathbf{f}' P' e_i - d\}$ and $f_j - f_i \geq \rho \mathbf{f}' P' (e_j - e_i)$.

(S) The local decision cost vector c_a in (8) is submodular. That is, the elements $c(i,a)$ satisfy $c(1,2) > c(1,1)$ and $c(2,2) < c(2,1)$. (Recall from Section II-A that $c(2,a) = c(3,a) = \cdots = c(X,a)$ in quickest detection problems with PH-distributed change time).

Discussion of Assumptions:

Assumption (A1): The requirement that $P(y|x)$ is TP2 with respect to states $\{1,2\}$ and $y \in \mathbb{Y}$ holds for numerous examples, see Karlin's classic book [22] and also [23]. Examples include quantized Gaussians, quantized exponential distributions, Binomial, Poisson, etc. For example consider quantized Gaussians. Suppose $B_{iy} = P(y|x=i) = \frac{\bar{b}_{iy}}{\sum_{y=1}^{Y} \bar{b}_{iy}}$ where $\bar{b}_{iy} = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{1}{2}\frac{(y-g_i)^2}{\Sigma}\right)$, $\Sigma > 0$, and $g_1 < g_2$. Then, (A1) holds.

Assumption (A2) always holds trivially for $X = 2$. For $X > 2$, see [16] and [25] for numerous examples. Consider the tridiagonal transition probability matrix P with $p_{ij} = 0$ for $j \geq i+2$ and $j \leq i-2$. As shown in [16, pp. 99–100], a necessary and sufficient condition for tridiagonal P to be TP2 is that $p_{i,i}p_{i+1,i+1} \geq p_{i,i+1}p_{i+1,i}$. Such a diagonally dominant tridiagonal matrix satisfies Assumption (A2).

Assumption (A3) is a sufficient condition for $C(\pi, 2)$ to be decreasing in π with respect to the MLR order. We will use (A3) in Section V to obtain sufficient conditions for a threshold policy. Assumption (A3) always holds for the geometric distributed change times ($X = 2$). For PH-distributed change times ($X > 2$), Assumption (A3) can be viewed as design constraints the decision maker needs to take into account so that quickest detection with PH-distributed change times has a threshold policy [26]. Feasible values for the elements of $\mathbf{f}$ are straightforwardly obtained using a LP solver such as `linprog` in MATLAB.

Assumption (S) is only required for the problem to be nontrivial. If (S) does not hold and $c(i,1) < c(i,2)$ for $i = 1, 2$, then local decision $a = 1$ will always dominate decision $a = 2$ and the problem reduces to a standard quickest detection problem where the observed local decision $a = 1$ yields no information about the state. Assumption (S) implies $c(x,2) - c(x,1)$ is decreasing in $x \in \{1,2\}$, i.e., the local cost $c(x,a)$ is submodular which implies the zero crossing condition that is important in the proof of Theorem 2.

The following theorem is an abbreviated version of Theorem 2 presented in Appendix C. It will be used in the rest of the paper as a natural partition of the belief state space $\Pi(X)$. Recall that transition probability P, observation probability matrix B_y and local cost vector c_a are defined in (2), (7), and (8) respectively.

Theorem 2: Under (A1), (A2), and (S), the belief state space $\Pi(X)$ defined in (14) can be partitioned into at most $Y+1$ nonempty polytopes denoted $\mathcal{P}_1, \ldots, \mathcal{P}_{Y+1}$ where we have (31), shown at the bottom of the page. On each such polytope, the local decision likelihood matrix R^π defined in (13) is a constant with respect to belief state π. ■

As a consequence of Theorem 2 and (13), there are only $Y+1$ possible decision likelihood matrices R^π, one per polytope $\mathcal{P}_l$, $l = 1, \ldots, Y+1$. We will denote these decision likelihood matrices as

$$R^l = R^\pi = BM^l = BM^\pi, \pi \in \mathcal{P}_l, l = 1, \ldots, Y+1. \quad (32)$$

Example: To give some insight into the structure of decision likelihood matrix R^π, suppose $X = 2$ (state space), $Y = 3$ (observation space), $A = 2$ (local decision space). Then assuming (A1), (A2), (S), by Theorem 2 there are up to $Y+1 = 4$ convex polytopes. The matrices M^l defined in (13) and (32) are

$$M^4 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad M^3 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$M^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad M^1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}. \quad (33)$$

Then, from (32) the 4 possible decision likelihood matrices R^l are

$$R^1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad R^2 = \begin{bmatrix} B_{11} & B_{12}+B_{13} \\ B_{21} & B_{22}+B_{23} \end{bmatrix}$$
$$R^3 = \begin{bmatrix} B_{11}+B_{12} & B_{13} \\ B_{21}+B_{22} & B_{23} \end{bmatrix}, \quad R^4 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}. \quad (34)$$

The detailed version of Theorem 2 in Appendix C guarantees that each of these matrices is TP2. Fig. 2 illustrates these polytopes and hyperplanes η_y defined below.

Let us give some intuition behind Theorem 2. Define the following Y hyperplanes that are subsets of $\Pi(X)$:

$$\eta_y = \{\pi \in \Pi(X) : (c_1 - c_2)' B_y P' \pi = 0\}, \quad y = 1, \ldots, Y. \quad (35)$$

$$\begin{aligned} \mathcal{P}_1 &= \{\pi \in \Pi(X) : (c_1 - c_2)' B_1 P' \pi \geq 0\} \\ \mathcal{P}_l &= \{\pi \in \Pi(X) : (c_1 - c_2)' B_{l-1} P' \pi < 0 \ \cap \ (c_1 - c_2)' B_l P' \pi \geq 0\}, \ l = 2, \ldots, Y \\ \mathcal{P}_{Y+1} &= \{\pi \in \Pi(X) : (c_1 - c_2)' B_Y P' \pi < 0\} \end{aligned} \quad (31)$$

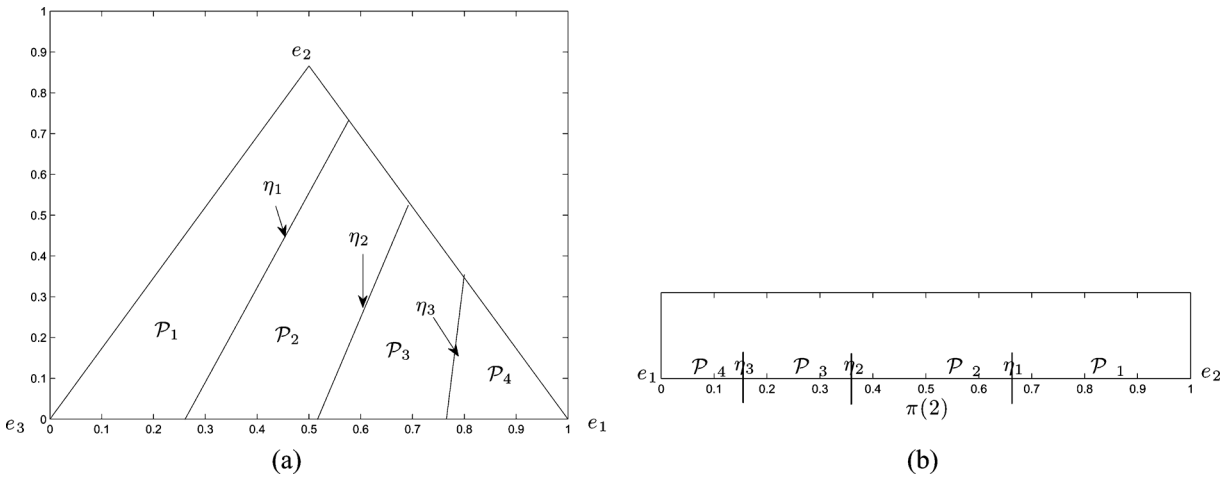

Fig. 2. Illustration of polytopes $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$, and $\mathcal{P}_4$ defined in (31) and hyperplanes η_1, η_2, η_3 defined in (35) for $Y = 4, A = 2$. Theorem 2 ensures that the hyperplanes do not intersect within the simplex $\Pi(X)$ and on each polytope, the local decision likelihoods R^π are a constant. In the figure, $e_2, e_3 \in \mathcal{P}_1$—Assumption (PH)(ii) in Section V ensures this.

The main intuition of the above theorem is that (A1), (A2), and (S) imply that $(c_1 - c_2)' B_y P' \pi$ satisfies a single crossing condition [4] with respect to a, y (see Definition 5 in Appendix A). This means that the set of belief states satisfy the following subset property:

$$\{\pi : (c_1 - c_2)' B_y P' \pi \geq 0\} \subseteq \{\pi : (c_1 - c_2)' B_{y+1} P' \pi \geq 0\}. \tag{36}$$

This implies that the hyperplanes η_y, $y \in \mathbb{Y}$, do not intersect within the simplex $\Pi(X)$. It is nice that straightforward conditions such as (A1), (A2), and (S) ensure this. Otherwise dealing with intersecting hyperplanes in a multidimensional simplex can be a real headache. Theorem 2(iv) in Appendix C shows that each hyperplane η_y partitions $\Pi(X)$ such that vertices $e_1, e_2, \ldots, e_{i_y}$ lie on one side and $e_{i_y+1}, \ldots, e_X$ lie on the other side. In Section V, we will introduce Assumption (PH)(ii) which ensures that $e_2, \ldots, e_X$ always lie in polytope $\mathcal{P}_1$ as illustrated in Fig. 2.

B. Multithreshold Structure of Social-Learning-Based Quickest Detection

The main result (Theorem 3 and Corollary 1) below gives sufficient conditions under which social-learning-based quickest detection has a double threshold policy. Consider the model (P, B, c, C, ρ) in (22) with geometric change time

$$\mathbb{X} = \mathbb{Y} = \mathbb{A} = \{1, 2\}, \quad P = \begin{bmatrix} 1 & 0 \\ \epsilon & 1 - \epsilon \end{bmatrix} \tag{37}$$

with $\mathbf{f} = \mathbf{f} = (0,\ f_2)'$ false-alarm vector in (17) and delay cost (18). Here, the change probability $\epsilon \ll 1$ is a small nonnegative scalar. So the change time τ^0 is geometrically distributed with $\mathbb{E}\{\tau^0\} = 1/\epsilon$.

The analysis in this section proceeds as follows.

Step 1: For $\epsilon = 0$, the problem becomes a simple sequential detection problem for state 1—we explicitly characterize the multithreshold behavior of the optimal decision policy in Theorem 3 below.

Step 2: It is then shown that for small ϵ, the optimal value function is within $O(\epsilon)$ of the value function for the case of zero change probability (Corollary 1). So, the optimal policy computed for zero change probability yields performance that is close to that of the optimal quickest detection policy for small ϵ.

1) Step 1: Sequential Detection of State 1: In line with above plan, consider the sequential detection problem for state 1 with social learning formulated in Section II with

$$\mathbb{X} = \mathbb{Y} = \mathbb{A} = \{1, 2\}, \quad P = I. \tag{38}$$

The state x is a random variable chosen at $k = 0$ with distribution π_0 and remains constant for $k > 0$. The goal is to detect and announce state 1 if $x_0 = 1$ based on noisy observations. The global decision $u_k = \mu(\pi_k) \in \{1\ (\text{stop}), 2\ (\text{continue})\}$ is a function of the public belief π_k. The optimal policy $\mu^*(\pi)$ that optimizes (20) satisfies Bellman's (25).

The 2-D belief state $\pi = [1 - \pi(2), \pi(2)]$ is parameterized by the scalar $\pi(2) \in [0, 1]$, i.e., $\Pi(X)$ is the interval $[0, 1]$. Each hyperplane η_y (35) now is a point on the interval $[0, 1]$; let the 2-D vector $[1 - \eta_y(2), \eta_y(2)]$ denote the belief state corresponding to η_y. The polytopes $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ in Theorem 2 are now intervals which are subsets of $[0, 1]$. If (A1) and (S) hold, then $\mathcal{P}_3 = [0, \eta_2(2))$, $\mathcal{P}_2 = [\eta_2(2), \eta_1(2))$, $\mathcal{P}_1 = [\eta_1(2), 1]$.

To handle the discontinuity in the social learning filter (11), we start with the following lemma that characterizes useful structural properties of the social learning filter. First define the belief state

$$q = T^{\eta_1}(\eta_1, 1). \tag{39}$$

Lemma 2: Consider the social learning filter (11) and assume (A1), (S) hold. Then:

i) $q = T^{\eta_1}(\eta_1, 1) = T^{\eta_2}(\eta_2, 2)$.

ii) If B is symmetric, then η_1 and η_2 are fixed points of the composite Bayesian map

$$\eta_1 = T^q(T^{\eta_1}(\eta_1, 1), 2), \quad \eta_2 = T^q((T^{\eta_2}(\eta_2, 2), 1). \tag{40}$$

■

The implication of the above lemma is that $\Pi(X)$ can be partitioned into 4 intervals, namely $[e_1, \eta_2)$, $[\eta_2, q)$, $[q, \eta_1)$ and

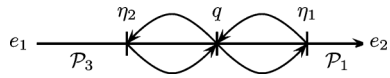

Fig. 3. Structure of social learning filter under the assumptions of Lemma 2 and symmetric B. Right (left) arrows represent evolution of the public belief when $a = 2$ ($a = 1$). As can be seen, η_1, η_2 are fixed points of the composite maps in (40).

$[\eta_1, e_2]$. Fig. 3 illustrates these regions and the dynamics specified in Lemma 2. The main result below characterizes the structure of the optimal global decision policy $\mu^*(\pi)$ on these four intervals. The theorem also characterizes information cascades [12] (more colloquially "herding") which is a salient feature of social learning.

Theorem 3: Consider the sequential detection problem with parameters (38). Suppose agents make local decisions via social learning. Assume (A1), (S) hold. (Note (A2) holds trivially since $P = I$). The optimal global decision policy $\mu^*(\pi)$ has the following properties:

i) For $\pi \in \mathcal{P}_1 \cup \mathcal{P}_3$, the global decision policy has a threshold structure

$$\mu^*(\pi) = \begin{cases} 2, & \text{if } \pi(2) > \pi^*(2) \\ 1, & \text{otherwise} \end{cases}$$
$$\text{where } \pi^*(2) = \frac{d}{f_2(1-\rho)+d}. \tag{41}$$

Also for $\pi \in \mathcal{P}_1 \cup \mathcal{P}_3$, the value function (25) is $V(\pi) = \min\{0, C(\pi,2)/(1-\rho)\}$ where $C(\pi,2)$ is defined in (24).

ii) The intervals $\mathcal{P}_1$ and $\mathcal{P}_3$ are "information cascades" [12]. That is, if $\pi_k \in \mathcal{P}_1 \cup \mathcal{P}_3$, then $\pi_{k+1} = \pi_k$ and social learning ceases.

iii) If B is symmetric, then for $\pi \in \mathcal{P}_2$, the global decision policy has the following structure:
   a) For $\pi \in [\eta_2(2), q(2))$, $V(\pi)$ is concave and there is at most one interval where $\mu^*(\pi) = 1$.
   b) For $\pi \in [q(2), \eta_1(2))$, $V(\pi)$ is concave and there is at most one interval where $\mu^*(\pi) = 1$.

■

The implication of Part (iii) of the above theorem is that the stopping set $\mathcal{S}$ comprises at most three intervals. One of these intervals is $(\pi^*(2), 1)$, with the threshold $\pi^*(2)$ defined in (41). The second claim of the theorem follows, since if public belief $\pi \in \mathcal{P}_1$, then the optimal local decision is $a = 2$ irrespective of the observation y. Similarly, if $\pi \in \mathcal{P}_3$, then the optimal local decision is $a = 1$ irrespective of the observation y. Therefore, when the public belief is in $\mathcal{P}_1 \cup \mathcal{P}_3$, the local decision of an agent reveals no information about its local observation to subsequent agents.

2) Step 2: Quickest Time Detection Bound for Small ϵ: Given the characterization in Theorem 3 of the optimal policy for $\epsilon = 0$, we now consider the quickest change detection problem for small ϵ specified in (37). It is convenient to introduce the following ϵ dependent notation.

Let $V_{\mu^*_\epsilon}(\pi)$ denote the cost incurred by the optimal policy μ^*_ϵ with transition matrix $P^\epsilon = \begin{bmatrix} 1 & 0 \\ \epsilon & 1-\epsilon \end{bmatrix}$. We use the notation $\mathcal{P}^\epsilon_l$ to denote the explicit dependence of the 3 intervals $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$, defined in (31). For $\epsilon = 0$, we denote these intervals as $\mathcal{P}^0_l$. The following result bounds the difference between $V_{\mu^*_0}(\pi)$ and $V_{\mu^*_\epsilon}(\pi)$. Note that $\mu^*_0(\pi)$ is characterized in Theorem 3 and $P^0 = I$ (identity matrix).

Recall from (23) that $\bar{V}(\pi)$ is the actual optimal expected cost associated with optimal decision policy $\mu^*(\pi)$. As mentioned below (25), the transformed value function $V(\pi)$ is more convenient to deal with to prove the existence of optimal threshold policies and the optimal policy remains invariant to the transformation from $\bar{V}(\pi)$ to $V(\pi)$.

Corollary 1: Consider the social-learning-based quickest detection model (P, B, c, C, ρ) in (22) with probability of change specified in (37). Then, for initial belief $\pi \in \mathcal{P}^\epsilon_l \cap \mathcal{P}^0_l$, $l = 1, 2, 3$, the optimal policy μ^*_0 (characterized in Theorem (3)) incurs a total global cost $V_{\mu^*_0}(\pi)$ that constitutes an $O(\epsilon)$ upper bound to the optimal global cost $\bar{V}_{\mu^*_\epsilon}(\pi)$ incurred in the quickest detection problem. More specifically, for $\pi \in \mathcal{P}^\epsilon_l \cap \mathcal{P}^0_l$, $l = 1, 2, 3$,

$$\bar{V}_{\mu^*_0}(\pi) - \bar{V}_{\mu^*_\epsilon}(\pi) \le \frac{4\rho\epsilon}{(1-\rho)^2} \max(d, f_2). \text{QED.} \tag{42}$$

■

Discussion: The implication of (42) is that the simple policy $\mu^*_0(\pi)$ of Theorem 3 is near optimal for quickest time detection with social learning when ϵ is small. Note that (42) compares the optimal costs in regions $\pi \in \mathcal{P}^\epsilon_l \cap \mathcal{P}^0_l$, $l = 1, 2, 3$, so we are omitting intervals where the models have different local decision likelihood probabilities R^π. The regions we are omitting are $O(\epsilon)$ in size. In each region $\pi \in \mathcal{P}^\epsilon_l \cap \mathcal{P}^0_l$, the only difference between the quickest detection model and the simplified model is the transition matrix (P^ϵ versus P^0). This allows us to give a tight bound in the sense that for $\epsilon = 0$, the optimal costs $\bar{V}_{\mu^*_0}(\pi)$ and $\bar{V}_{\mu^*_\epsilon}(\pi)$ coincide. Of course, (42) requires the discount factor $\rho < 1$. We refer the reader to [48] for an alternative and more general approach.

The proof of Corollary 1 follows from Theorem 2 of [42]. In terms of our notation, Theorem 2 of [42] shows that for a POMDP with piecewise linear value function at each iteration of the value-iteration algorithm, for $\pi \in \mathcal{P}^\epsilon_l \cap \mathcal{P}^0_l$

$$\bar{V}_{\mu^*_0}(\pi) \le \bar{V}_{\mu^*_\epsilon}(\pi) + \frac{2\rho}{(1-\rho)^2} \|\bar{C}(\pi, u)\|_\infty \sup_i \|[P^\epsilon - P^0]_{ij} R^l_a\|_1 \tag{43}$$

where the $\|\cdot\|_1$ induced matrix norm is with respect to the (j, a) elements. Since from Theorem 3, the value function is piecewise linear, (43) applies. From the structure of P^ϵ in (37) and since $P^0 = I$, clearly

$$\sup_i \|[P^\epsilon - P^0]_{ij}\ R^l_a\|_1 = \epsilon \max(B_{11} + B_{21},\ B_{12} + B_{22}) \le 2\epsilon.$$

Also $\|\bar{C}(\pi, u)\|_\infty = \max(d, f_2)$. Substituting these in (43) yields the bound (42).

3) Numerical Example: Consider the social learning quickest detection model (P, B, c, C, ρ) with $\mathbb{X} = \mathbb{Y} = \mathbb{A} = \{1, 2\}$

$$B = \begin{bmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{bmatrix}, \quad c = \begin{bmatrix} 1 & 2 \\ -1 & -3.57 \end{bmatrix}$$
$$\rho = 0.8, \quad d = 1.8, \quad f_2 = 2. \tag{44}$$

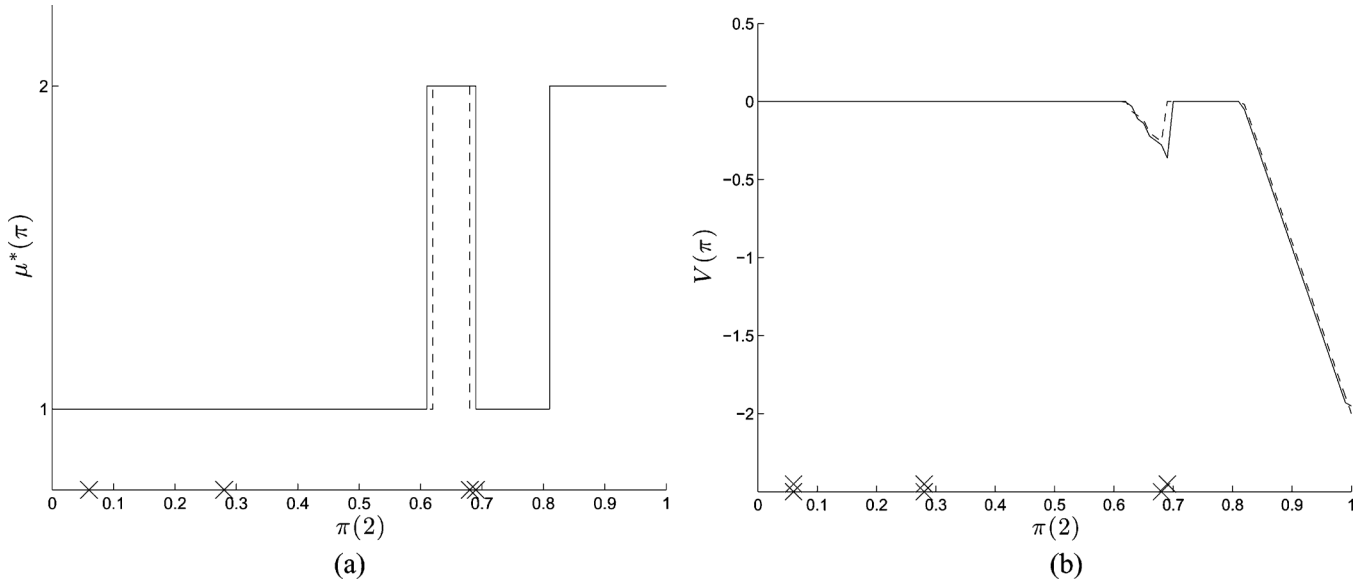

Fig. 4. Optimal decision policy for quickest time change detection with social learning for geometric distributed change time with small probability of change. In each subfigure, the graph with solid lines is for $\epsilon = 0.005$ and the graph with broken lines is for $\epsilon = 0$. The policies and optimal costs are very close for $\epsilon = 0.005$ and $\epsilon = 0$. Equation (42) gives a bound for the difference in the optimal costs. In both cases, the optimal policies are a double threshold and the value functions are nonconcave and discontinuous. (a) Optimal global decision policies $\mu_0^*(\pi)$ and $\mu_\epsilon^*(\pi)$. (b) Value functions for global decision policy.

Fig. 4 shows the optimal policies μ_0^* (see Theorem 3) and μ_ϵ^* (optimal quickest detection policy) together with optimal costs $V_{\mu_0^*}(\pi)$ and $V_{\mu_\epsilon^*}(\pi)$ for change probability $\epsilon = 0.005$. As can be seen the quickest detection optimal policy and costs are very close to the costs and policies specified by Theorem 3. For $\epsilon = 0.002$ the policies μ_0^* and μ_ϵ^* are almost identical and cannot be distinguished in Fig. 4. The policies and optimal costs were obtained by running the value iteration algorithm for horizon 500 with $\Pi(X) = [0, 1]$ discretized to a grid of 100 points.

V. Quickest Time Detection for Geometric and PH-Distributed Change Time

The previous section illustrated the multithreshold behavior of social-learning-based quickest time change detection. What sufficient conditions on the social learning model lead to single threshold behavior? This section gives such conditions for PH-distributed change times τ^0 modeled by a $X \geq 2$-state Markov chain. For geometric change times (i.e., $X = 2$) these conditions yield a threshold that is identical to the classical Kolmogorov–Shiryaev criterion (21).

This section comprises the following results.

1) Section V-A gives sufficient conditions for the optimal global decision policy μ^* to be myopic and characterized by a linear hyperplane threshold.
2) Section V-B gives less restrictive conditions under which the optimal policy is increasing with respect to the MLR order and is characterized by a single threshold curve. Recall that for PH-distributed change time, the belief space $\Pi(X)$ is a multidimensional simplex. To order posterior distributions on this simplex, the MLR stochastic order (which is a partial order) will be used since it is preserved under conditional expectations. The results involve analysis of the structure of the social learning Bayesian filter together with lattice programming. All definitions of these orders and consequences are given in the Appendix.
3) Section V-C describes how sufficient conditions can be given for multiple-threshold policies.
4) Finally, Section V-D characterizes the optimal linear approximation to the MLR increasing policy. It then formulates estimation of the optimal linear approximation to the threshold curve as a stochastic optimization problem.

Assumption (PH): Recall fictitious states $2, \ldots, X$ (corresponding to belief states $e_2, \ldots, e_X$) are used to model the PH-distribution in (4). It, therefore, makes sense to constrain the model parameters so that the global decision policy $\mu^*(\pi)$ at the belief states $e_2, \ldots, e_X$ are identical (and similarly for the local decisions taken in social learning). Throughout this section, when considering PH-distributed change times, we make the following assumption:

(PH) (i) $C' e_i < 0$ for $i = 2, \ldots, X$. (ii) $e_2, \ldots, e_X$ lie in polytope $\mathcal{P}_1$.

Assumption (PH)(i) says that the optimal policy $\mu^*(\pi)$ treats each of the fictitious states $2, \ldots, X$ identically—they all lie outside the stopping set $\mathcal{S}$. In similar vein, (PH)(ii) requires that individual agents making local decisions treat the fictitious states $i = 2, \ldots, X$ identically, i.e., they lie to the left of each hyperplane η_y, $y = 1, \ldots, Y$.

Obviously, (PH) holds trivially for $X = 2$ (geometric case)—otherwise the quickest change problem would be degenerate.

A. Case 1: Myopic Quickest Detection With Linear Hyperplane Threshold

The main result of this section is Theorem 4 which shows that under suitable conditions, the optimal policy $\mu^*(\pi)$ has a myopic structure characterized by $C'\pi = 0$. Recall that C in (24) denotes the transformed costs of the global decision maker with elements C_j, $j = 1, \ldots, X$. Denote the $X - 1$ vertices of the intersection of the linear hyperplane $\{\pi : C'\pi = 0\}$ with

the facets of simplex $\Pi(X)$ as $\nu_j, j = 1, \ldots, X-1$. Then, it is straightforwardly seen that these vertices are

$$\nu_j = \frac{C_{j+1}\, e_1 - C_1\, e_{j+1}}{C_{j+1} - C_1}, \quad j = 1, \ldots, X-1. \tag{45}$$

Now introduce the following assumption:

(C1) $C' R_a^{\nu_j} P' \nu_j \geq 0$ for all $a \in \mathbb{A}$, $j = 1, \ldots, X-1$.

The relevance of (C1) is apparent from the following lemma (proof in Appendix E). Define the set of belief states (polytope)

$$S = \{\pi : C'\pi \geq 0\} \tag{46}$$

Lemma 3: (C1) together with (A1), (A2), (A3), and (PH) is sufficient for the set S defined in (46) to be closed under the social learning filter (11). That is $\pi \in S \Rightarrow T^\pi(\pi, a) \in S$ for all $a \in \mathbb{A}$.

Recall that (A1), (A2), and (A3) were introduced in Section IV-A and (PH) at the beginning of Section V. The main result is as follows. The proof is in Appendix E.

Theorem 4: Consider the social-learning-based quickest time detection model (P, B, c, C, ρ) in (22). Assume (A1), (A2), (A3), (S), (C1), (PH). Then the global decision maker's optimal policy is myopic and is of the form

$$\mu^*(\pi) = \begin{cases} 1\,(\text{stop}), & \text{if } C'\pi \geq 0 \\ 2\,(\text{continue}), & \text{otherwise} \end{cases}, \quad \mathcal{S} = \{\pi : C'\pi \geq 0\}. \tag{47}$$

For the special case $X = 2$ (geometric change time)

$$\mu^*(\pi) = \begin{cases} 1\,(\text{stop}), & \text{if } \pi(2) \leq \pi^*(2) \\ 2\,(\text{continue}), & \text{if } \pi(2) > \pi^*(2) \end{cases}$$
$$\text{where } \pi^*(2) = \frac{d}{d + f_2(1-\rho P_{22})} \tag{48}$$

The above result is similar to the entry fee optimal stopping problem having a myopic policy discussed in [21, pp. 389] and [41, Th. 2.2, p. 54]. It is important to note, however, that even though the optimal policy $\mu^*(\pi)$ in (47) is characterized by a linear threshold, the value function $V(\pi)$ can still be discontinuous and nonconcave (unlike classical stopping time problems). This will be illustrated in the numerical example below.

Let us illustrate what Theorem 4 says. Consider Fig. 5. The shaded region in Fig. 5 denotes the set $S = \{\pi : C'\pi \geq 0\}$. It is clear from Bellman's (25) that the stopping set $\mathcal{S}$ is a *subset* of this shaded region S. What Theorem 4 says is that the stopping set is *equal* to the shaded region, i.e., $\mathcal{S} = S$, if (C1) and (PH) hold. In terms of Fig. 5, (C1) is sufficient for $T^\pi(\pi, a)$ to map the belief states ν_1 and ν_2 (which are the vertices of the line $C'\pi = 0$) to polytope S. (PH)(i) implies that states e_2, e_3 lie to the left of the line $C'\pi = 0$ (which corresponds to the region $C'\pi < 0$). Similarly, (PH)(ii) means that e_2, e_3 lie to the left of each line segment η_y, $y = 1, 2$, i.e., $e_2, e_3 \in \mathcal{P}_1$.

Numerical Example: To illustrate Theorem 4, consider the geometric change time model in (44) except that $P_{22} = 0.75$. Even though the sufficient condition (C1) does not hold, the optimal policy is characterized by a single threshold given by (48). This is shown in Fig. 6. As can be seen in Fig. 6, the value function is nonconcave and discontinuous.

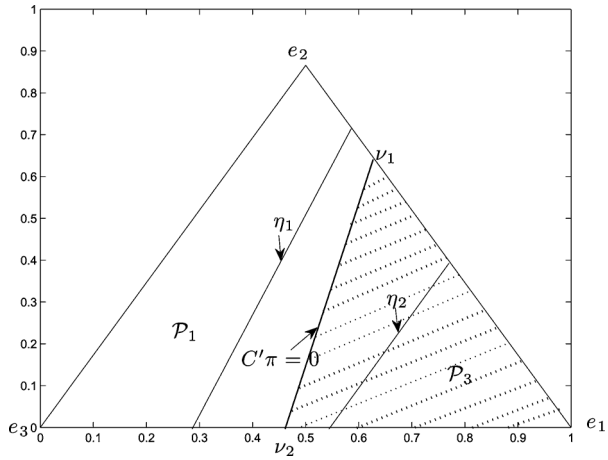

Fig. 5. Illustration of Theorem 4. The shaded region which depicts the polytope $\{\pi : C'\pi \geq 0\}$ is equivalent to the stopping set $\mathbf{S}$ under the assumptions of the theorem.

B. Case 2: Existence of a Single Threshold Switching Curve

In this section, we consider another special case of the social-learning-based quickest detection model (22). Theorem 5 below shows that the stopping set is characterized by a single threshold curve on the belief space. The threshold coincides with the classical quickest time detection problem with noninformative observations. For PH-distributed change times, unlike the previous section, the threshold curve is not necessarily linear. We give a stochastic gradient algorithm to estimate this threshold curve in Section V-D.

1) Structural Result: We make the following assumptions. Recall the global decision maker's cost vector C is defined in (24). Let $\bar{\nu}_j, j = 1, \ldots, X-1$ denote the $X-1$ vertices of the intersection of hyperplane η_Y (defined in (35)) with $\Pi(X)$. These vertices are computed as (45) with C replaced by $PB_Y(c_1 - c_2)$.

(C2) $(c_1 - c_2)' B_Y (P')^2 \bar{\nu}_j \leq 0$ for $j = 1, \ldots, X-1$.

(C3) The linear hyperplane $\{\pi : C'\pi = 0\}$ lies in polytope $\mathcal{P}_{Y+1}$.

The following is the main result. The proof is in Appendix F.

Theorem 5: Consider the social-learning-based quickest detection model (P, B, c, C, ρ) in (22). Assume (A1), (A2), (S), and (PH) hold. The optimal policy $\mu^*(\pi)$ has the following structure

i) Under (C3), $\mu^*(\pi) = 2$ for $\pi \notin \mathcal{P}_1$.

ii) Under (C2) and (C3), the stopping set $\mathcal{S}$ is as convex subset of polytope $\mathcal{P}_{Y+1}$. Therefore, the boundary of $\mathcal{S}$ is differentiable almost everywhere.

iii) For geometric-distributed change time $(X = 2)$, under (C3), the optimal policy is identical to that of the Kolmogorov–Shiryaev criterion (21) with uniformly distributed observation probabilities.

iv) Under (A3), (C2), and (C3). on the polytope $\mathcal{P}_{Y+1}$, $\mu^*(\pi)$ has the following structure:

$$\pi_1, \pi_2 \in \mathcal{P}_{Y+1} \text{ and } \pi_1 \geq_r \pi_2 \text{ implies } \mu^*(\pi_1) \geq \mu^*(\pi_2) \tag{49}$$

(The MLR order $\geq_r$ is defined in (61) in Appendix A). Hence, the boundary of the stopping set $\mathcal{S}$ within $\Pi(X)$ intersects any line segment $\mathcal{L}(e_1, \bar{\pi})$ or $\mathcal{L}(e_X, \bar{\pi})$ at most once (see geometric interpretation below). ■

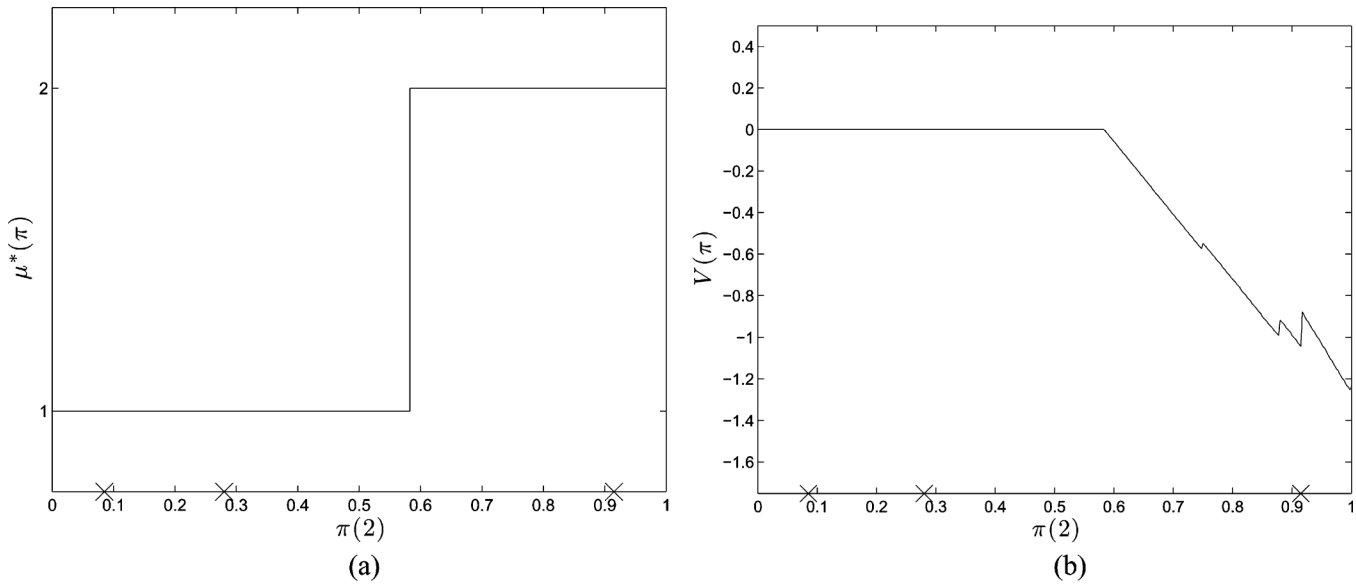

Fig. 6. Numerical example illustrating Theorem 4 which characterizes the optimal decision policy for social-learning-based quickest detection. The example is described in Section V-A. Even though the value function is nonconcave and discontinuous, the optimal policy has a single threshold specified by (48). (a) Optimal global decision policy $\mu^*(\pi)$. (b) Value functions for global decision policy.

Even though the policy $\mu^*(\pi)$ in Theorem 5 coincides with that of classical quickest detection, the optimal cost incurred is always larger as shown in Theorem 1.

2) Discussion of Theorem 5 and Assumptions: Assumption (C3) localizes the decision threshold to polytope $\mathcal{P}_{Y+1}$. As a consequence of (C3), $C'\pi < 0$ on all polytopes except $\mathcal{P}_{Y+1}$. Therefore, on these polytopes, $\mu^*(\pi) = 2$. Thus, statement (i) is obvious.

Assumption (C2) together with (A1), (A2), (S), and (PH) ensures that the polytope $\mathcal{P}_{Y+1}$ is closed under the belief state mapping $T^\pi(\pi, a)$. That is, $\pi \in \mathcal{P}_{Y+1}$ implies $T^\pi(\pi, a) \in \mathcal{P}_{Y+1}$ for all a. Note that Assumption (C2) holds trivially for $X = 2$ as shown in the footnote.[8] (C2) is similar in spirit to (C1) of the Section V-A—the key difference is that (C1) deals with the global cost vector C whereas (C2) deals with local costs c_1, c_2.

Assumptions (C2) and (C3) allow us to show that the value function $V(\pi)$ is concave on $\mathcal{P}_{Y+1}$. Then, statement (ii), namely convexity of the stopping set $\mathcal{S}$, follows from arguments in [32].

Statement (iii) is straightforward to show. The local decision likelihood probabilities on $\mathcal{P}_{Y+1}$ are uniform since the local decision yields no information about the state. Thus under (C3) the threshold is identical to the classical quickest detection threshold for the Kolmogorov–Shiryaev criterion (21) with uniformly distributed observation probabilities.

The proof of Statement (iv) is more involved and is given in Appendix F. The proof uses structural properties of the Bayesian social filter studied in Theorem 10 of Appendix F, along with submodularity, MLR stochastic order and a version defined on line segments $\mathcal{L}(e_1, \bar{\pi})$ and $\mathcal{L}(e_X, \bar{\pi})$ in Appendix A.

3) Interpretation of Statement (iv): Since statement (iv) is nontrivial, let us explain what it says from a geometric point of view. For PH-distributed change times with $X > 2$, statement (iv) says a lot more than convexity of the stopping region $\mathcal{S}$. On the unit simplex $\Pi(X)$ define $\mathcal{L}(e_1, \bar{\pi})$ as the line segment constructed from e_1 to any point $\bar{\pi} \in \{e_2, \ldots, e_X\}$ on the opposite facet of the simplex $\Pi(X)$. Similarly denote $\mathcal{L}(e_X, \bar{\pi})$ as any line segment from e_X to any point $\bar{\pi}$ on the opposite facet $(e_1, \ldots, e_{X-1})$. Statement (iv) implies that the boundary of the stopping set $\mathbf{S}$ within $\Pi(X)$ intersects any such line $\mathcal{L}(e_1, \bar{\pi})$ or $\mathcal{L}(e_X, \bar{\pi})$ at most once. Fig. 7 shows examples of convex sets that violate this condition. Also, statement (iv) leads to the following nice geometrical interpretation. If a belief state $\pi \in \mathcal{S}$ lies on a line $\mathcal{L}(e_X, \bar{\pi})$, then all belief states on this line closer to $\bar{\pi}$ also lie in $\mathcal{S}$. Similarly, if a belief state $\pi \in \mathcal{L}(e_1, \bar{\pi})$ lies outside the stopping set $\mathcal{S}$, then all belief states on the line $\mathcal{L}(e_1, \bar{\pi})$ further away from e_1 also lie outside the stopping set.

Numerical examples are given in Section VII.

C. Extensions of Theorem 5 and Multithreshold Policies

1) Local and Global Costs in Global Decision Making: Theorem 5 can be extended to consider a more general global decision maker's cost function (instead of only false alarm and delay) which takes into account the cost of local decisions in social learning. For example, suppose that the global decision maker's cost for picking decision $u = 2$ (continue) is the delay cost plus an "operating cost." That is

$$\bar{C}(\pi, 2) = de_1'\pi + \beta C_{op}(\pi). \tag{50}$$

Here, $\beta \geq 0$ is a user defined constant and with $\sigma(\cdot)$, $T(\cdot)$, $\mathcal{H}$ defined in (7), (1)

$$\begin{aligned} C_{op}(\pi) &\stackrel{\triangle}{=} \mathbb{E}_y\{\min_{a\in\mathbb{A}} \mathbb{E}\{c(x,a)|\mathcal{H}_k\}\} \\ &= \sum_{y\in\mathbb{Y}} \min_{a\in\mathbb{A}}\{c_a' T(\pi, y)\}\sigma(\pi, y) \\ &= \sum_{y\in\mathbb{Y}} \min_{a\in\mathbb{A}} c_a' B_y P'\pi. \end{aligned} \tag{51}$$

[8]For $X = 2$, the second element of $P'\pi$ is $P_{22}\pi_2$ which is always smaller than π_2, So applying P to any belief state keeps it within the interval $\mathcal{P}_{Y+1}$.

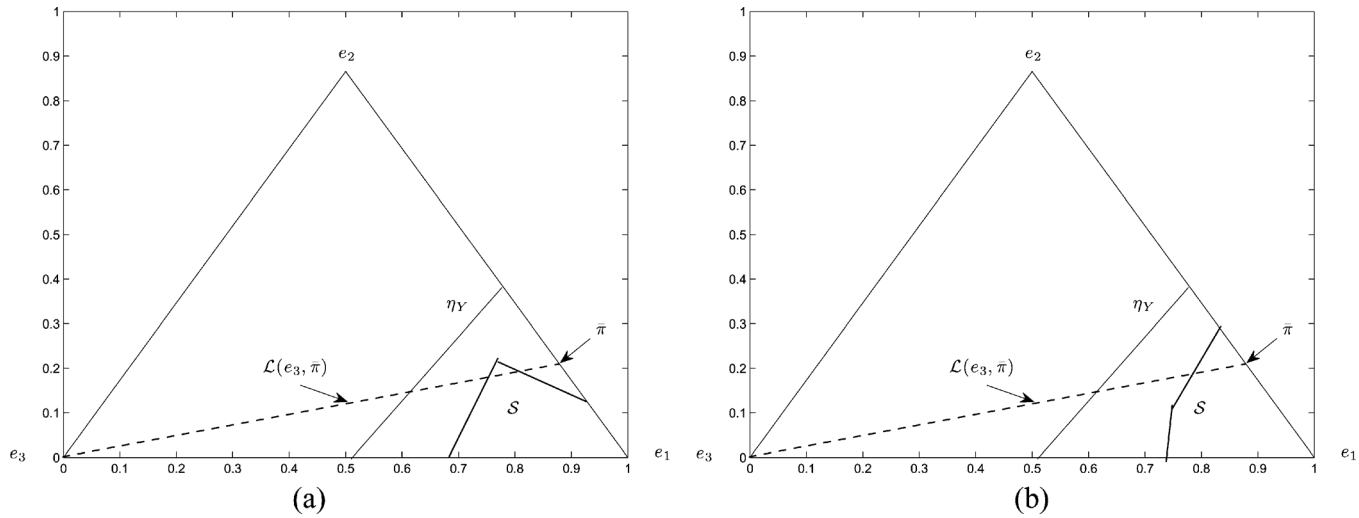

Fig. 7. Stopping set $\mathcal{S}$ in (a) violates Statement (iv) of Theorem 5 since the boundary of stopping set $\mathcal{S}$ within $\Pi(X)$ intersects the line $\mathcal{L}(e_3,\bar{\pi})$ twice. (b) Example of a stopping set $\mathcal{S}$ that satisfies Statement (iv) of Theorem 5. In both figures, the region to the right of line η_Y is the polytope $\mathcal{P}_{Y+1}$.

$C_{op}(\pi)$ is the expected operating cost since it is incurred at each agent k when it makes its local decision via social learning. Note $C_{op}(\pi)$ is the expected local cost from choosing decision $u=2$, receiving signal y, picking recommendation a and broadcasting the information to the network: the probability of the event is $\sigma(\pi,y)$ and the cost is $\min_a \bar{c}_a' T(\pi,y)$. The last equality in (51) follows since $\sigma(\pi,y)$ is a nonnegative scalar independent of a. Actually, the above choice of $C_{op}(\pi)$ is very similar to that used in constrained social learning in [12, Ch. 4].

Then using the same transformation as in (24), the optimal policy is given by the Bellman's (25) with $C(\pi,2) = C'\pi + C_{op}(\pi)$. Assumption (C3), namely, $\{\pi : C(\pi,2)=0\} \in \mathcal{P}_{Y+1}$ is then equivalent to the linear hyperplane $(C+Pc_1)'\pi = 0$ lying in polytope $\mathcal{P}_{Y+1}$. This is because on polytope $\mathcal{P}_{Y+1}$ the optimal local decision $a=1$, see (34), and so $C_{op}(\pi) = c_1' P'\pi$. Suppose assumption (A3) is augmented with the condition that $c(i,a=1)$ is decreasing with i. Then, Theorem 5 continues to hold.

2) Multiple Thresholds: Using a similar proof to Theorem 5, sufficient conditions can be given for the optimal global policy $\mu^*(\pi)$ in social learning-based quickest detection to have multiple thresholds. We describe this as follows.

Suppose the hyperplane $C'\pi = 0$ lies in polytope $\mathcal{P}_{y^*}$ for some $y^* \in \{1,2,\ldots,Y+1\}$. Assume (C2) holds. Also assume the following generalization of (C2) holds.

(C2') The social learning filter maps belief states in polytope $\mathcal{P}_y$ to polytope $\mathcal{P}_{y+1}$ for $y = y^*, y^*+1,\ldots,Y$. That is, $\pi \in \mathcal{P}_y$ implies $T^\pi(\pi,a) \in \mathcal{P}_{y+1}$.

Then, similar to the proof of Theorem 5, the following result can be established (proof omitted).

Theorem 6: Under (A1), (A2), (A3), (S), (PH), (C2), (C2'), the value function $V(\pi)$ is MLR decreasing and therefore optimal policy $\mu^*(\pi)$ is MLR increasing on each polytope $\mathcal{P}_y$, $y = y^*,\ldots,Y+1$.

As a result, $\mu^*(\pi)$ is characterized by up to $Y+2-y^*$ threshold curves, one on each of these polytopes. The reason is that even though $V(\pi)$ is decreasing in each polytope, there is no guarantee that is decreasing between polytopes. Theorem 5 is a special case of the above result when $y^* = Y+1$, and therefore, $\mu^*(\pi)$ is characterized by a single threshold curve.

As an example, consider $X=2$, $Y=2$, $A=2$ and suppose $C'\pi = 0$ lies in $\mathcal{P}_2$, i.e., $y^*=2$. Since $X=2$ (geometric change time), conditions (A2), (A3), (PH), and (C2) hold trivially. (C2') holds if the social learning filter $T^\pi(\cdot)$ maps the belief states in $\mathcal{P}_2$ to $\mathcal{P}_3$. A sufficient condition for this is $T^{\eta_1}(\eta_1,2) \in \mathcal{P}_3$, i.e., the transition matrix satisfies (52) shown at the bottom of the page. If (A1) and (52) hold, then according to the Theorem 6, the optimal policy $\mu^*(\pi)$ is monotone decreasing on each interval $\mathcal{P}_2$ and $\mathcal{P}_3$. So $\mu^*(\pi)$ is characterized by up to 2 thresholds, one in each of these intervals.

D. Optimal Linear Decision Threshold and Algorithms

Theorem 5 showed that under conditions (A1), (A2), (A3), (S), (PH), (C2), and (C3), the optimal decision policy $\mu^*(\pi)$ was MLR increasing in belief state $\pi \in \mathcal{P}_{Y+1}$. In this section, we characterize *linear* threshold hyperplanes that preserve this MLR structure. Such linear thresholds can then be computed via a stochastic approximation algorithm. For geometric distributed change time τ^0, since the thresholds are points, estimation is an obvious special case.

Throughout this section, we assume that the conditions of Theorem 5 hold.

1) Characterization of MLR Increasing Linear Threshold: For $\pi \in \mathcal{P}_{Y+1}$, define the $X-1$-dimensional parameter vector

$$P_{22} \geq \frac{B_{12}}{B_{22}B_{11}} \frac{(c(2,1)-c(2,2))B_{21}B_{12} - (c(1,1)-c(2,1))B_{11}B_{22}}{(c(2,1)-c(2,2))B_{22} - (c(1,1)-c(1,2))B_{21}} \tag{52}$$

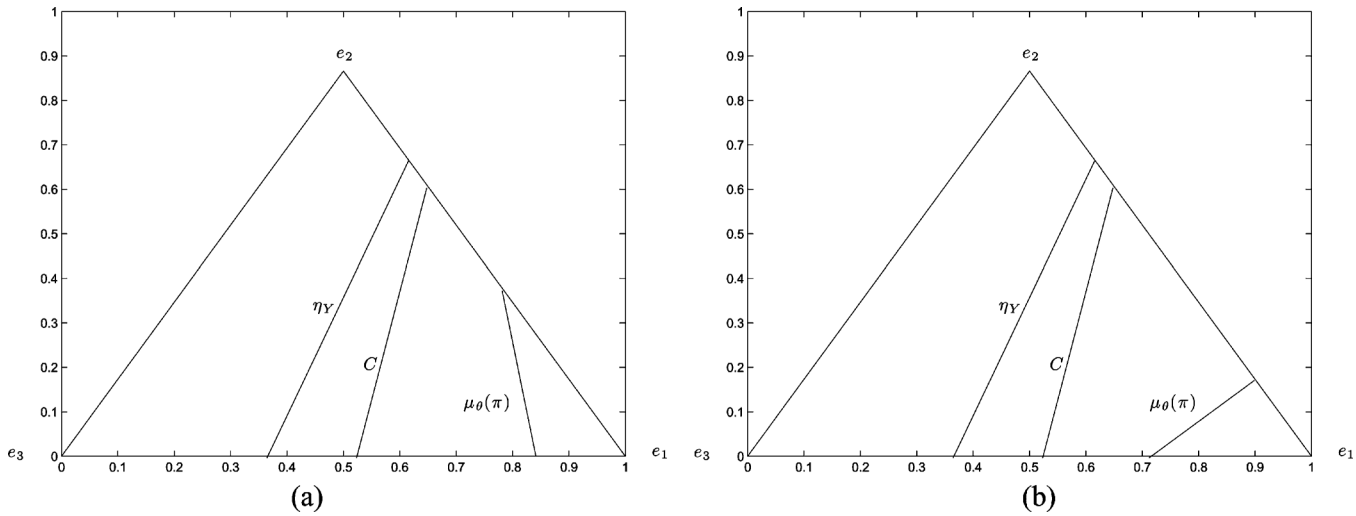

Fig. 8. (a) Valid MLR increasing linear threshold policy. The linear threshold policy in (b) violates the requirement that the policy $\mu_\theta(\pi)$ is MLR increasing since it has a slope less than 60°. In both figures, the region to the right of line η_Y is the polytope $\mathcal{P}_{Y+1}$. In the figures, C denotes the hyperplane $C'\pi = 0$ which lies in $\mathcal{P}_{Y+1}$ by assumption (C3).

$\theta = (\theta(1), \ldots, \theta(X-1))'$. Since $\Pi(X) \subset \mathbb{R}^{X-1}$, a linear hyperplane on $\Pi(X)$ is parameterized by $X-1$ coefficients. Define the linear threshold policy $\mu_\theta(\pi)$ parameterized by the vector θ as

$$\mu_\theta(\pi) = \begin{cases} \text{stop} = 1, & \text{if } \pi(2) + \sum_{i=1}^{X-2} \theta(i)\pi(i+2) \le \theta(X-1) \\ \text{continue} = 2, & \text{otherwise.} \end{cases} \tag{53}$$

Assume conditions (A1), (A2), (A3), (S), (PH), (C2), and (C3) hold for the quickest detection problem (20) so that from Theorem 5, the optimal policy $\mu^*(\pi)$ is MLR increasing on lines $\mathcal{L}(e_X, \bar{\pi})$ and $\mathcal{L}(e_1, \bar{\pi})$. These are defined in Appendix A. The requirement that state 1 lies in the stopping set, means $\mu_\theta(e_1) < 0$ which implies $\theta(X-1) > 0$.

Theorem 7: For belief states $\pi \in \Pi(X)$, the linear threshold policy $\mu_\theta(\pi)$ defined in (53) is

i) MLR increasing on lines $\mathcal{L}(e_X, \bar{\pi})$ iff $\theta(X-2) \ge 1$ and $\theta(i) \le \theta(X-2)$ for $i < X-2$.
ii) MLR increasing on lines $\mathcal{L}(e_1, \bar{\pi})$ iff $\theta(i) \ge 0$, for $i < X-2$.

■

The proof of Theorem 7 is in Appendix G. The constraints in the above theorem are necessary *and* sufficient for the linear threshold policy (53) to be MLR increasing on lines $\mathcal{L}(e_X, \bar{\pi})$ and $\mathcal{L}(e_1, \bar{\pi})$. Under these constraints, (53) defines the set of all MLR increasing linear threshold policies on $\mathcal{L}(e_X, \bar{\pi})$ and $\mathcal{L}(e_1, \bar{\pi})$—it does not leave out any MLR increasing polices; nor does it include any non MLR increasing policies. In this sense, optimizing over the space of MLR increasing linear threshold policies yields the optimal linear approximation to threshold curve.

The conditions imposed on the linear threshold parameters θ in Theorem 7 have a nice interpretation when $X = 3$. Recall in this case $\Pi(X)$ is an equilateral triangle. Let $(\omega(1), \omega(2))$ denote Cartesian coordinates in the equilateral triangle. So $\pi(2) = 2\omega(2)/\sqrt{3}, \pi(1) = \omega(1) - \omega(2)/\sqrt{3}$. Then, the linear threshold satisfies

$$\omega(2) = \frac{\sqrt{3}\theta(1)}{2-\theta(1)}\omega(1) + \big(\theta(2) - \theta(1)\big)\frac{\sqrt{3}}{2-\theta(1)}.$$

So the conditions of Theorem 7 require that $\theta(1) \ge 1$, i.e., the threshold has slope of 60° or larger. When $\theta(1) > 2$, slope becomes negative, i.e., more than 90°.

Fig. 8 shows examples of a valid and invalid linear threshold. Fig. 8(a) illustrates a valid MLR increasing linear threshold policy. Fig. 8(b) is invalid since the threshold is less than 60° meaning that the resulting policy is not MLR increasing on lines. Also shown is the hyperplane $C'\pi = 0$ which by assumption (C3) lies in polytope $\mathcal{P}_{Y+1}$.

2) Computation of Optimal Linear Threshold: As a consequence of Theorem 7, the optimal linear threshold approximation to threshold curve Γ of Theorem 5 is the solution of the following constrained optimization problem:

$$\theta^* = \arg \min_{\theta \in \mathbb{R}^{X-1}} J_{\mu_\theta}(\pi_0) \quad \text{subject to } 0 \le \theta(i) \le \theta(X-2),\ \theta(X-2) \ge 1 \text{ and } \theta(X-1) > 0 \tag{54}$$

where the cost $J_{\mu_\theta}(\pi_0)$ is obtained as in (20) by applying threshold policy μ_θ in (53).

Because the cost $J_{\mu_\theta}(\pi_0)$ in (54) cannot be computed in closed form, we resort to simulation based stochastic optimization. Let $n = 1, 2 \ldots$, denote iterations of the algorithm. The aim is to solve the following linearly constrained stochastic optimization problem:

$$\text{Compute } \theta^* = \arg \min_{\theta \in \Theta} \mathbb{E}\{J_n(\mu_\theta)\} \quad \text{subject to } 0 \le \theta(i) \le \theta(X-2),\ \theta(X-2) \ge 1 \text{ and } \theta(X-1) > 0. \tag{55}$$

Here, for each initial condition π_0, the sample path cost $J_n(\mu_\theta, \pi_0)$ is evaluated as

$$J_n(\mu_\theta, \pi_0) = \sum_{k=1}^{\infty} \rho^{k-1} C(\pi_k, u_k)$$
$$\text{where } u_k = \mu_\theta(\pi_k) \text{is computed via (53)}$$
$$J_n(\mu_\theta) = \frac{1}{L} \sum_{l=1}^{L} J_n(\mu_\theta, \pi_0^{(l)})$$
$$\text{where prior } \pi_0^{(l)} \text{ is sampled uniformly from simplex } \Pi(X). \quad (56)$$

A convenient way of sampling uniformly from $\Pi(X)$ is to use the Dirichlet distribution (i.e., $\pi_0(i) = x_i / \sum_i x_i$, where $x_i \sim$ unit exponential distribution).

The above stochastic optimization problem is solved by stochastic approximation algorithms such as the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm [47] which converges to a local minimum; see [26] for a novel parameterization that deals with the hypersphere constraints. The stochastic gradient algorithm converges to local optima, so it is necessary to try several initial conditions. The computational cost at each iteration is linear in the dimension of θ and is independent of the observation alphabet size Y. Convergence (w.p.1) can be established using techniques in [29] and [30]. More sophisticated methods than SPSA can also be used. For example, Bartlett and Baxter[7] use the score function method to perform gradient-based reinforcement learning. These algorithms are applicable to solve the constrained stochastic optimization problem (55). Also, if the change time distribution (specified by P) and the observation likelihoods (specified by B) are not completely specified, as long as the assumptions Theorem 5 hold, then the reinforcement learning algorithms [7] can be used to solve (55).

VI. Multiagent Quickest Time Detection With Adaptive Sensing

As mentioned in Section I, the social learning protocol is very similar to multiagent quickest time detection with a sensor manager (controller). Motivated by sensor network applications, this section describes the formulation and the main results. The information patterns are similar to social learning and so the results developed in previous sections apply. The observations now can also belong to a continuum.

Consider a countable number of agents indexed by $k = 1, 2, \ldots$. Each agent acts once in a predetermined sequential order indexed by $k = 1, 2, \ldots$ as follows: Based on the current belief state π_{k-1}, agent k acts as follows.

1) Agent k first chooses decision $u_k \in \{1(\text{stop}), 2(\text{continue})\}$. If the agent decides to stop, then as in earlier sections, a false alarm penalty is paid, and the problem terminates.
2) If agent k chooses $u_k = 2$, then it chooses its operating *mode* $a_k \in \{1, 2\}$ according to a built-in micromanager. Agent k then views the world according to this mode—that is, it obtains observation y_k from a distribution that depends on mode a_k. It then communicates its belief state π_k to the next agent.

Remark: An equivalent formulation is as follows: A single smart sensor adapts its operating mode a_k at each time k based on the posterior distribution of the underlying state at the previous time instant. How can quickest detection be achieved with this sensor?

How can such a network of agents, where each agent makes autonomous micromanagement decisions on its mode, achieve quickest time detection? The quickest time detector can be viewed as a macromanager that operates on the belief states and micromanager decisions. Clearly, the micro- and macromanagers interact—the local decisions a_k taken by the micromanager—determines y_k which determines π_k and hence determines decision u_{k+1} of the quickest time macromanager.

A. Micromanager for Agent Mode Selection

1) Costs and Mode Selection: As in (8), let c_a denote the local cost of deploying sensor mode $a \in \mathbb{A} = \{1, 2\}$. To avoid trivial solutions, as in Section IV-A, we make the submodular assumption (S).

Similar to the social learning formulation, the micromanager picks local decision a_k myopically as follows: Based on the belief state π_{k-1} of the previous agent, each agent k picks its mode $a_k \in \mathbb{A} = \{1, 2\}$ of which sensor to deploy by minimizing its expected predicted cost

$$a_k = \arg \min_{a \in \{1,2\}} \mathbb{E}\{c(x_k, a_k) | \mathcal{F}_{k-1}\} = \arg \min_{a \in \{1,2\}} c_a' P' \pi_{k-1} \quad (57)$$

where $\mathcal{F}_k$ denotes the filtration $\sigma(y_l, l \le k)$. Define the convex polytopes $\mathcal{P}_1$ and $\mathcal{P}_2$ that partition $\Pi(X)$ as

$$\mathcal{P}_1 = \{\pi : (c_1 - c_2)' P' \pi \ge 0\}, \quad \mathcal{P}_2 = \{\pi : (c_1 - c_2)' P' \pi < 0\}. \quad (58)$$

Then, from (57) it follows that for $\pi \in \mathcal{P}_1$, $a_k = 2$ and for $\pi \in \mathcal{P}_2$, $a_k = 1$.

2) Mode-Dependent Observations: The agent then makes an observation y_k depending on its choice of mode a_k. Based on its mode a_k in (57), agent k then obtains an observation from conditional probability distribution

$$P(y_k \le \bar{y} | x_k = e_x, a_k = a) = \sum_{y \le \bar{y}} B_{xy}^{(a)}, \quad x \in \mathbb{X}, a \in \{1, 2\}. \quad (59)$$

Here, $\sum_y$ denotes integration with respect to the Lebesgue measure (in which case $\mathbb{Y} \subset \mathbb{R}$ and B_{xy} is the conditional probability density function) or counting measure (in which case $\mathbb{Y}$ is a subset of the integers and B_{xy} is the conditional probability mass function $B_{xy} = P(y_k = y | x_k = x)$). The key point is that unlike classical quickest detection, each agent now views the world based on its selected mode a_k.

Let $T^{(a)}(\pi, y)$ denote the belief state update if mode a is chosen and measurement y obtained. It is given by the HMM filter (7) with mode dependent probabilities $B_y^{(a)} = \text{diag}(P(y|x, a), x \in \mathbb{X})$. That is,

$$T^{(a)}(\pi, y) = B_y^{(a)} P' \pi / \sigma(\pi, y), \quad \sigma(\pi, y) = \mathbf{1}' B_y^{(a)} P' \pi. \quad (60)$$

B. Macromanager for Quickest Time Detection

In the following, we present the assumptions and main result. Based on the above micromanager protocol, the aim is to

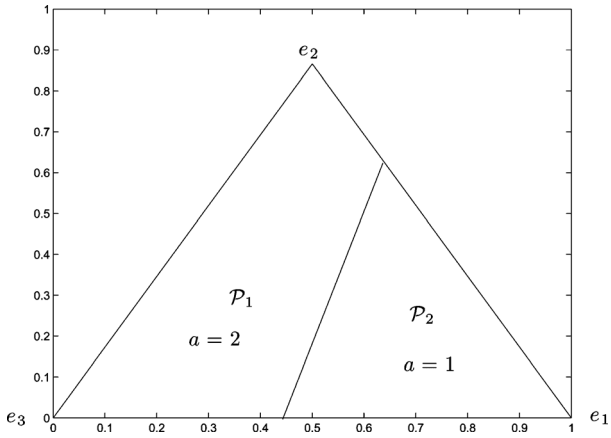

Fig. 9. Illustration of the setup in Section VI. The mode-dependent observation probabilities $B^{(a)}$, $a \in \{1, 2\}$ are chosen depending on the belief state π in polytope $\mathcal{P}_2$ or $\mathcal{P}_1$ defined in (58). The aim is to perform quickest detection given this mode-dependent observation probability constraint.

perform quickest time change detection. So the quickest detection problem can be viewed as optimizing the cost function (24) subject to the constraint that the belief state evolves according to (60). The setup is identical to that in Sections II-B and III-A. For $k < \tau$, agents choose $u = 2(\text{continue})$ and at $k = \tau$, agent k picks $u_k = 1(\text{ declares a change and stop})$. The optimal policy $\mu^*(\pi)$ of the macromanager satisfies Bellman's (25).

The following theorems mimic the results for the social-learning-based quickest detection problem, and their proofs are identical.

1) Blackwell Dominance: Suppose the mode dependent observation matrices are of the form $B^{(a)} = BQ^{(a)}$ where B and $Q^{(a)}$, $a = 1, 2$ are stochastic kernels. Then, an identical proof to Theorem 1 shows that classical quickest detection with observation matrix B always yields a lower cost than mode dependent quickest detection with observation matrices $B^{(a)}$, where the mode a is chosen according to any arbitrary strategy.

2) Threshold Policies: Consider the following assumptions that are similar to (C1) in Section V-A and (C2) in Section V-B. Recall vertices ν_j are defined in (45) and $\bar{\nu}_j$ denote vertices of hyperplane $(c_1 - c_2)'P'\pi = 0$.

(C1) If $\{\pi : C'\pi = 0\}$ lies in one of the polytopes $\mathcal{P}_a$, then $C'B_y^{(a)}P'\nu_j \geq 0, j = 1, \ldots, X-1$ for all $y \in \mathbb{Y}$.

(C2) $(c_1 - c_2)'P'B_y^{(1)}P'\bar{\nu}_j \leq 0, j = 1, \ldots, X-1, y \in \mathbb{Y}$.

We have the following result regarding the structure of $\mu^*(\pi)$ for quickest time detection.

Theorem 8: Theorems 4 and 5 hold for the optimal quickest time decision policy $\mu^*(\pi)$ of the macromanager. Also Theorem 7 holds for MLR policies and computation of the optimal linear threshold can be formulated as the stochastic optimization problem (56).

(C1) and (C2) are relatively easy to check even if $y \in \mathbb{Y}$ is continuum as shown below. For all x, let $y_{\max}$ denote the maximum support of the distribution $B_{xy}^{(1)}$, i.e., $y_{max} = \sup\{y : B_{xy}^{(1)} > 0\}$.

Lemma 4: (C1), (C2) hold if their inequalities hold for $y = y_{\max}$.

Thus only a finite number of inequalities need to be verified. In particular for a Gaussian distribution, since $y_{\max} = \infty$, the filter $T^{(1)}(\pi, \infty)$ becomes the Bayesian predictor $P'\pi$. So it suffices to check that $C'P'\nu_j \geq 0$ for (C1) to hold.

Proof: Consider (C1). $C'B_y^{(a)}P'\nu_j \geq 0$ is equivalent to verifying $C'B_y^{(a)}P'\nu_j/\sigma(\nu_j, y) \geq 0$ since $\sigma(\pi, y)$ is nonnegative for all $\pi \in \Pi(X)$. So we need to check that $C'T^{(a)}(\nu_j, y) \geq 0$ for all $y \in \mathbb{Y}$. But since P and B are TP2 according to Assumptions (A1) and (A2), from Theorem 10(4) in Appendix F, the belief state update $T^{(a)}(\pi, y)$ is MLR increasing in y. Moreover by (A3) C has decreasing elements. Therefore, from Result 1 in Appendix A, $C'T^{(a)}(\pi, y)$ is decreasing in y. So it suffices to check that $C'T^{(a)}(\nu_j, y_{\max}) \geq 0$. ■

VII. Numerical Results

In addition to the numerical examples presented earlier, this section presents two numerical examples. The first example illustrates the multiple threshold policies inherent in social learning (this example was mentioned in Section I). The second example illustrates the optimal threshold curve for a PH-type distributed change time that was proved in Theorem 5.

Example 1. Geometric Distributed Change Time: This examples illustrates the existence of a triple threshold policy for quickest time change detection when the change time τ^0 is geometrically distributed. We chose the social learning model with parameters $\mathbb{X} = \{1, 2\}$ (so $\Pi(X) = [0, 1]$ is a one dimensional simplex), $\mathbb{Y} = \{1, 2, 3\}$, $\mathbb{A} = \{1, 2\}$

$$B = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}, \quad E\{\tau^0\} = 20 \Rightarrow P = \begin{bmatrix} 1 & 0 \\ 0.05 & 0.95 \end{bmatrix}$$
$$c = \begin{bmatrix} 1 & 2 \\ -1 & -3.57 \end{bmatrix}.$$

For the global quickest time detection parameters, we chose $\rho = 0.99$, delay $d = 1.25$, false alarm vector $\mathbf{f} = 3e_2$ (i.e., $f_2 = 3$). It is easily checked that (A1), (A2), and (S) hold.

The optimal policy $\mu^*(\pi)$ is shown in Fig. 1(a) and comprises a triple threshold policy. It was computed by constructing a uniform grid of 500 points for $\pi(2) \in [0, 1]$ and then implementing the value iteration algorithm (27) for a horizon of $N = 200$. The "$\times$" in Fig. 1(a) and (b) represents the values of $\eta_2(2)$, $q(2)$, and $\eta_1(2)$, respectively.

Example 2. Phase Distributed Change Time: This examples illustrates Theorem 5 which proved the existence of a single threshold curve for social-learning-based quickest time change detection with PH-distributed change time. We model the PH-distribution via a three-state Markov chain. So the belief space $\Pi(X)$ is a 2-D simplex (equilateral triangle) and can be visualized easily.

We chose the social learning model with parameters $\mathbb{X} = \{1, 2, 3\}$ $\mathbb{Y} = \{1, 2, 3, 4, 5\}$, $\mathbb{A} = \{1, 2\}$. The observation probabilities and local decision costs were chosen as

$$B_{1,y} \propto \exp(-(y-1)^2/6)$$
$$B_{2,y} = B_{3,y} \propto \exp(-(y-5)^2/6)$$
$$c = (c(i,a)) = \begin{bmatrix} 4 & 50 \\ 2 & 0 \\ 2 & 0 \end{bmatrix}.$$

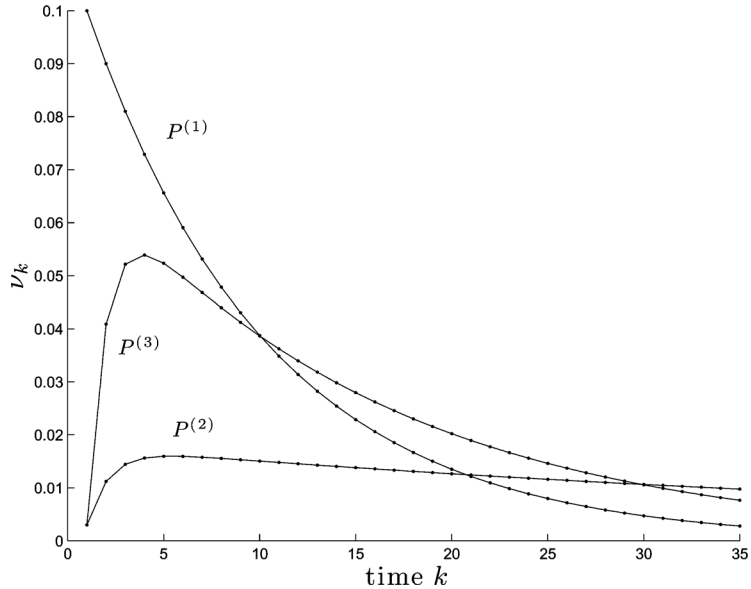

Fig. 10. Plots of change time τ^0 probability mass function ν_k in (4) for $P^{(1)}$ (geometric distribution) and $P^{(2)}$, $P^{(3)}$ (phase-type distributions).

The global costs for quickest detection in (17) and (18) were chosen as $d = 1.5$, $\mathbf{f} = [0\ 20\ 25]'$ and discount factor $\rho = 0.9$.

The PH-distributed change times were modeled by the three-state Markov chain with transition probability P. To illustrate the quickest time detection, we chose four candidate transition probability matrices, namely

$$P^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0.1 & 0.9 & 0 \\ 0.1 & 0.9 & 0 \end{bmatrix}, \quad P^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0.1 & 0.5 & 0.4 \\ 0 & 0.1 & 0.9 \end{bmatrix}$$

$$P^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0.1 & 0.7 & 0.2 \\ 0 & 0.4 & 0.6 \end{bmatrix}, \quad P^{(4)} = \begin{bmatrix} 1 & 0 & 0 \\ 0.1 & 0.45 & 0.45 \\ 0.05 & 0.40 & 0.55 \end{bmatrix}.$$

Note $P^{(1)}$ models the geometric distribution since states 2 and 3 are indistinguishable—in fact it is exactly lumpable [24] into the two-state Markov chain with transition matrix $\begin{bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{bmatrix}$.

Fig. 10 plots the probability mass function ν_k [see (4)] of the PH-distributed change time τ^0 for these four transition matrices for $\bar{\pi}_0 = [0.03,\ 0.97]'$. Fig. 10 shows these PH-distributions are quite different in behavior to a geometric distribution—they are nonmonotone and have heavier tails.

It is easily checked that (A1), (A2), (A3), (S), (PH), (C2), and (C3) of Theorem 5 hold. Fig. 11 shows the optimal decision policies for these four cases with the stopping set $\mathcal{S}$ shaded. The optimal policy was computed as follows. A 50×50 grid of $(\pi(1), \pi(2))$ values was formed within the 2-D unit simplex $\Pi(X)$.Then, the value iteration algorithm (27) was solved for horizon $N = 200$ (in all cases $\|V_{200}(\pi) - V_{199}(\pi)\|_\infty < 10^{-15}$ implying that the value iteration algorithm converged). In all four cases, the optimal decision policy is characterized by a single threshold curve in polytope $\mathcal{P}_6$. This is consistent with Theorem 5.

In each plot of Fig. 11 also shows the hyperplanes $C'\pi = 0$ [defined in (24)] and η_5 [defined in (35)]. The polytope $\mathcal{P}_6$ is to the right of hyperplane η_5. The remaining line segments from left to right are $\eta_1, \ldots, \eta_4$. Note that hyperplane $C'\pi = 0$ lies in $\mathcal{P}_6$, thereby satisfying Assumption (PH) and (C3).

Actually cases $P^{(2)}$ and $P^{(3)}$ satisfy assumptions (C1), and (PH) and so Theorem 4 holds. Therefore, for these two cases, the optimal threshold curve is the linear hyperplane $C'\pi = 0$ as can be seen in Fig. 11.

VIII. Conclusion

Motivated by understanding how local and global decision making interact, this paper has presented structural results for quickest time detection when agents perform social learning. Also, a related model incorporating multiagent sensor scheduling and quickest time detection was considered. Unlike classical quickest detection, the optimal policy can have multiple thresholds. Four main results were presented. First, Theorem 1 showed using Blackwell dominance of measures that social-learning-based quickest detection always results in more expensive cost compared to classical quickest detection. Second, for symmetric observation probabilities and geometric change times, the explicit multithreshold behavior of social-learning-based quickest detection was characterized in Theorem 3 by approximating with a simpler detection problem. Third, quickest time change detection for more general PH-type distributed change times was considered. Theorem 4 gave sufficient conditions for the optimal policy to be characterized by a single linear hyperplane in the multidimensional simplex of posterior distributions. Finally, using lattice programming and likelihood ratio dominance Theorem 5 gave sufficient conditions for the optimal policy to be characterized by a single switching curve. The optimal linear approximation to this curve (that preserves the MLR monotone nature of the policy) was characterized in Theorem 7.

The results of this paper are straightforwardly extended to more general stopping problems where the underlying Markov state does not have an absorbing state, as long as the transition matrix satisfies assumption (A2). In current work, we are using similar social learning models for "order-book" trades in agent

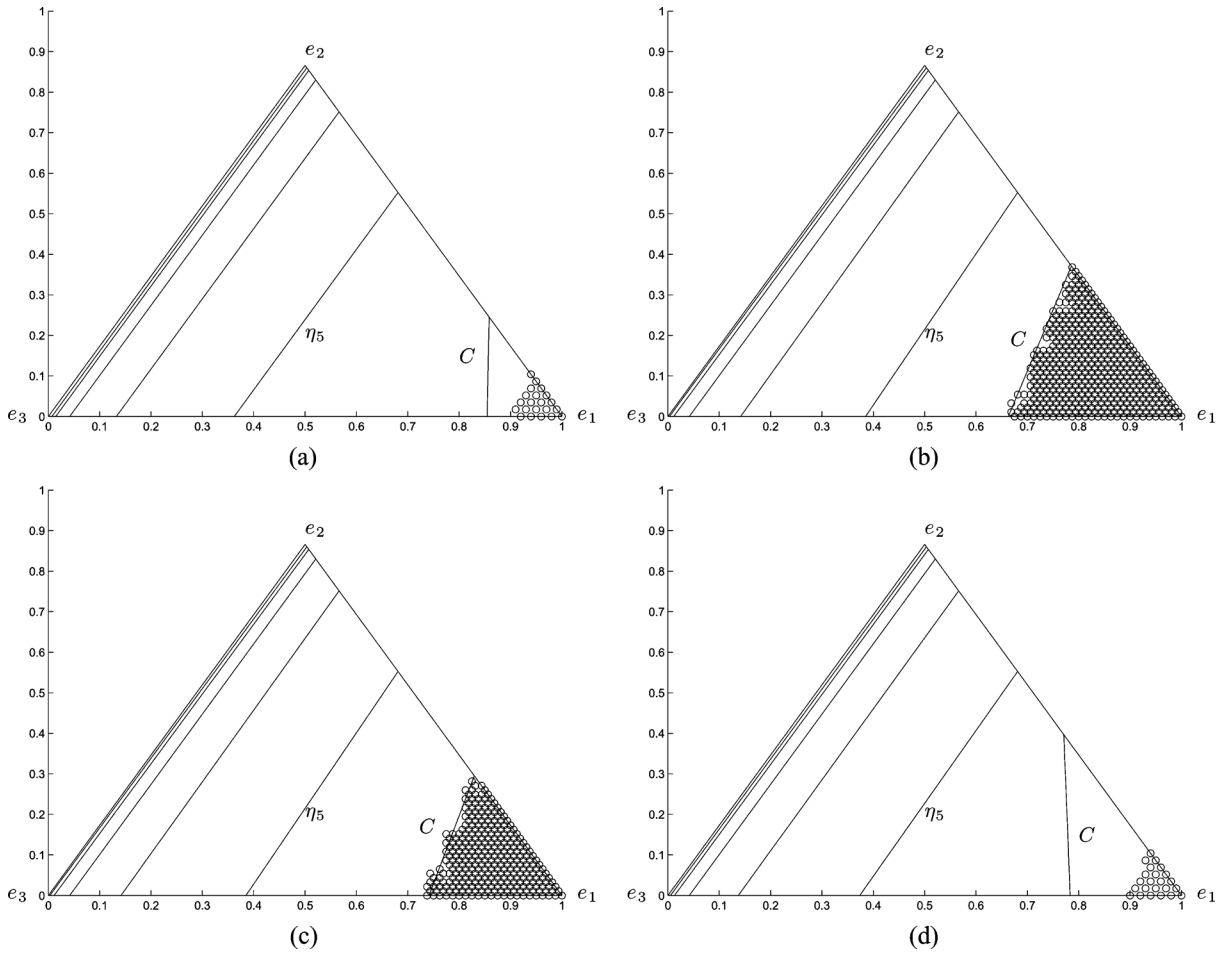

Fig. 11. Optimal decision policy for quickest time change time with geometric probability mass function for geometric distribution (transition probability $P^{(1)}$), and phase-type distributions (transition probabilities $P^{(2)}$, $P^{(3)}$ and $P^{(4)}$). The shaded region depicts the stopping set $\mathcal{S}$ in (26). The parameters are specified in Example 2 of Section VII.

based models for algorithmic market making (see also [38]). Finally, [26] also presents quickest detection with constrained social learning, where agents can either herd or reveal full information about their observation.

APPENDIX

A) Preliminaries: Stochastic Dominance, Submodularity: Background references for stochastic dominance and lattice programming are [23], [25], [35], and [49]. The proofs of Theorem 2 and Theorem 5 require concepts in stochastic dominance. In particular, statement (iv) of Theorem 5 states that the optimal social policy $\mu^*(\pi)$ is monotonically increasing in belief state π. In order to compare belief states π and $\tilde{\pi}$, we will use the MLR stochastic ordering and a specialized version of the MLR order restricted to lines in the simplex $\Pi(X)$. The MLR order is useful for social learning since it is preserved after conditioning [23], [35], [40].

Definition 1 (MLR Ordering, [35, pp. 12–15]: Let $\pi_1, \pi_2 \in \Pi(X)$ be any two belief state vectors. Then π_1 is greater than π_2 with respect to the MLR ordering—denoted as $\pi_1 \geq_r \pi_2$, if

$$\pi_1(i)\pi_2(j) \leq \pi_2(i)\pi_1(j), \quad i < j, i, j \in \{1, \ldots, X\}. \quad (61)$$

Definition 2 (First-Order Stochastic Dominance): Let $\pi_1, \pi_2 \in \Pi(X)$. Then, π_1 first-order stochastically dominates π_2—denoted as $\pi_1 \geq_s \pi_2$—if $\sum_{i=j}^{X} \pi_1(i) \geq \sum_{i=j}^{X} \pi_2(i)$ for $j = 1, \ldots, X$.

Result 1 [35]:

i) $\pi_1 \geq_r \pi_2$ implies $\pi_1 \geq_s \pi_2$. (For $X = 2$, $\geq_r$ and $\geq_s$ are equivalent)
ii) Let $\mathcal{V}$ denote the set of all S dimensional vectors v with nondecreasing components, i.e., $v_1 \leq v_2 \leq \cdots v_X$. Then $\pi_1 \geq_s \pi_2$ iff for all $v \in \mathcal{V}$, $v'\pi_1 \geq v'\pi_2$.
iii) Suppose $f_i \geq g_i$, $i = 1, \ldots, X$ and f_i, g_i are increasing in i. Then $\pi \geq_s \bar{\pi}$ implies $\sum_i f_i\pi_i \geq \sum_i g_i\bar{\pi}_i$. (This follows since from (ii) $\sum_i g_i\pi_i \geq \sum_i g_i\bar{\pi}_i$ and $\sum_i f_i\pi_i > \sum_i g_i\pi_i$ since $f_i > g_i \ \forall i$).

For state-space dimension $X = 2$, MLR is a complete order and coincides with first-order stochastic dominance. For state-space dimension $X > 2$, MLR is a *partial order*, i.e., $[\Pi(X), \geq_r]$ is a partially ordered set (poset) since it is not always possible to order any two belief states $\pi \in \Pi(X)$.

Finally, we define a modification of the MLR order on certain line segments in the simplex which yields a total ordering.

Define the set of belief states $\mathcal{H}_i = \{\pi \in \Pi(X) : \pi(i) = 0\}$. For each belief state $\bar{\pi} \in \mathcal{H}_i$, denote the line segment $\mathcal{L}(e_i, \bar{\pi})$ that connects $\bar{\pi}$ to e_i. Thus

$$\mathcal{L}(e_i, \bar{\pi}) = \{\pi \in \Pi(X) : \pi = (1-\epsilon)\bar{\pi} + \epsilon e_i, \ 0 \leq \epsilon \leq 1\} \quad \bar{\pi} \in \mathcal{H}_i. \quad (62)$$

Definition 3 (MLR Ordering $\geq_{L_1}$ and $\geq_{L_X}$ on Lines): π_1 is greater than π_2 with respect to the MLR ordering on the line $\mathcal{L}(e_1,\bar{\pi})$—denoted as $\pi_1 \geq_{L_1} \pi_2$ if $\pi_1,\pi_2 \in \mathcal{L}(e_1,\bar{\pi})$ for some $\bar{\pi} \in \mathcal{H}_1$ and $\pi_1 \geq_r \pi_2$. Similarly, $\pi_1 \geq_{L_X} \pi_2$, if $\pi_1, \pi_2 \in \mathcal{L}(e_X,\bar{\pi})$) for some $\bar{\pi} \in \mathcal{H}_X$, and $\pi_1 \geq_r \pi_2$.

Note that $[\Pi(X), \geq_{L_1}]$ is a chain, i.e., all elements $\pi, \tilde{\pi} \in \mathcal{L}(e_1,\bar{\pi})$ are comparable, i.e., either $\pi \geq_{L_1} \tilde{\pi}$ or $\tilde{\pi} \geq_{L_1} \pi$. Similarly $[\Pi(X), \geq_{L_X}]$ is a chain. In Lemma 5, we summarize useful properties of $[\Pi(X), \geq_{L_1}]$ that will be used in our proofs.

Lemma 5: Consider $[\Pi(X), \geq_r]$, $[\mathcal{L}(e_X,\bar{\pi}), \geq_{L_1}]$.

i) On $[\Pi(X), \geq_r]$, e_1 is the least and e_X is the greatest element. On $[\mathcal{L}(e_X,\bar{\pi}), \geq_{L_1}]$, $\bar{\pi}$ is the least and e_X is the greatest.
ii) Convex combinations of MLR comparable belief states form a chain. For any $\gamma \in [0,1]$, $\pi \leq_r \tilde{\pi} \Rightarrow \pi \leq_r \gamma\pi + (1-\gamma)\tilde{\pi} \leq_r \tilde{\pi}$.
iii) All points on a line $\mathcal{L}(e_X,\bar{\pi})$ are MLR comparable. Consider any two points π^{γ_1}, $\pi^{\gamma_2} \in \mathcal{L}(e_X,\bar{\pi})$ (62) where $\pi^{\gamma} = \gamma e_X + (1-\gamma)\bar{\pi}$. Then $\gamma_1 \geq \gamma_2$, implies $\pi^{\gamma_1} \geq_{L_1} \pi^{\gamma_2}$. A similar result holds for $\mathcal{L}(e_1,\bar{\pi})$.

Definition 4 (Submodular Function [49]): $f : \mathcal{L}(e_1,\bar{\pi}) \times \{1,2\} \to \mathbb{R}$ is submodular (antitone differences) if $f(\pi,u) - f(\pi,\bar{u}) \leq f(\tilde{\pi},u) - f(\tilde{\pi},\bar{u})$, for $\bar{u} \leq u$, $\pi \geq_{L_1} \tilde{\pi}$.

The following result says that for a submodular function $Q(\pi,u)$, $u^*(\pi) = \operatorname{argmin}_u Q(\pi,u)$ is increasing in its argument π. This implies $\mu^*(\pi)$ is MLR increasing on the line segments $\mathcal{L}(e_x,\bar{\pi})$, which in turn will be used to prove the existence of as threshold decision curve.

Theorem 9 [49]: If $f : \mathcal{L}(e_1,\bar{\pi}) \times \{1,2\} \to \mathbb{R}$ is submodular, then there exists a $u^*(\pi) = \operatorname{argmin}_{u\in\{1,2\}} f(\pi,u)$, that is increasing on $[\mathcal{L}(e_1,\bar{\pi}), \geq_{L_1}]$, i.e., $\tilde{\pi} \geq_{L_1} \pi \Rightarrow u^*(\pi) \leq u^*(\tilde{\pi})$.

Definition 5 (single Crossing Condition [4], [49]): $g : \mathbb{Y} \times \mathbb{A} \to \mathbb{R}$ satisfies a single crossing condition in (y,a) if $g(y,a) - g(y,\bar{a}) \geq 0$ implies $g(\bar{y},a) - g(\bar{y},\bar{a}) \geq 0$ for $\bar{a} > a$ and $\bar{y} > y$. Then $a^*(y) = \operatorname{argmin}_a g(y,a)$ is increasing in y.

Definition 6 (TP2 Ordering and Reflexive TP2 Distributions): Let P and Q denote any two multivariate probability mass functions. Then

i) $P \underset{\text{TP2}}{\geq} Q$ if $P(\mathbf{i})Q(\mathbf{j}) \leq P(\mathbf{i}\vee\mathbf{j})Q(\mathbf{i}\wedge\mathbf{j})$. If P and Q are univariate, then this definition is equivalent to the MLR ordering $P \geq_r Q$ defined above.
ii) A multivariate distribution P is said to be multivariate TP2 if $P \underset{\text{TP2}}{\geq} P$ holds, i.e., $P(\mathbf{i})P(\mathbf{j}) \leq P(\mathbf{i}\vee\mathbf{j})P(\mathbf{i}\wedge\mathbf{j})$.
iii) If $\mathbf{i}, \mathbf{j} \in \{1,\ldots,X\}$ are scalar indices, Statement (ii) is equivalent to saying that a $M \times N$ matrix A is TP2 if all second order minors are nonnegative. if $i \geq j$, then the i-th row of A MLR dominates the j-th row.

B) Proof of Theorem 1: Let $\underline{V}_k(\pi)$ denote the value function at iteration k of the value iteration algorithm (27) associated with the classical quickest detection Bellman (29). Recall $V_k(\pi)$ is the value function associated with the social-learning-based quickest detection problem (25).

We start with the following lemma which is proved at the end of Appendix B

Lemma 6: $\sum_a \underline{V}_k(T^\pi(\pi,a))\sigma(\pi,a) \geq \sum_y \underline{V}_k(T(\pi,y))\sigma(\pi,y)$.

The proof of Theorem 1 then follows by mathematical induction using the value iteration algorithm (27). Assume $V_k(\pi) \geq \underline{V}_k(\pi)$ for $\pi \in \Pi(X)$. Then

$$\begin{aligned} C(\pi,2) + \sum_a V_k(T^\pi(\pi,a))\sigma(\pi,a) &\geq C(\pi,2) + \sum_a \underline{V}_k(T^\pi(\pi,a))\sigma(\pi,a) \\ &\geq C(\pi,2) + \sum_y \underline{V}_k(T(\pi,y))\sigma(\pi,y) \end{aligned}$$

where the second inequality follows from Lemma 6. Thus $V_{k+1}(\pi) \geq \underline{V}_{k+1}(\pi)$. This completes the induction step. Since value iteration converges pointwise, $V(\pi) \geq \underline{V}(\pi)$ thus proving the theorem.

Proof of Lemma 6:

Step 1: First, let us show that $\underline{V}_k(\pi)$ is concave over $\Pi(X)$ for any k by induction. Recall from (27) that $\underline{V}_0(\pi) = -\bar{C}(\pi,1)$ which is linear in $\pi \in \Pi(X)$. Assume $\underline{V}_k(\pi)$ is concave at iteration k. Note that $\underline{V}_k(k)$ is positively homogeneous, i.e., for any $c \geq 0$, $\underline{V}_k(c\pi) = c\underline{V}_k(\pi)$. So the value iteration algorithm (27) associated with Bellman's (29) is

$$\underline{V}_{k+1}(\pi) = \min\{C'\pi + \rho \sum_y \underline{V}_k(B_y P'\pi), 0\}$$

Since the composition of concave function with a linear function preserves concavity, therefore $\sum_y \underline{V}_k(B_y P'\pi)$ is concave and so $\underline{V}_{k+1}(\pi)$ is concave.

Step 2: We then use the Blackwell dominance condition (13). The social learning filter (11) can be expressed in terms of the HMM filter (7) as

$$\begin{aligned} T^\pi(\pi,a) &= \sum_{y\in\mathbb{Y}} T(\pi,y)\frac{\sigma(\pi,y)}{\sigma(\pi,a)}P(a|y,\pi) \\ \text{and } \sigma(\pi,a) &= \sum_{y\in\mathbb{Y}} \sigma(\pi,y)P(a|y,\pi). \end{aligned}$$

Therefore, $\frac{\sigma(\pi,y)}{\sigma(\pi,a)}P(a|y,\pi)$ is a probability measure w.r.t y. Since from Step 1, $\underline{V}_k(\cdot)$ is concave for $\pi \in \Pi(X)$, using Jensen's inequality it follows that

$$\begin{aligned} \underline{V}_k(T^\pi(\pi,a)) &= \underline{V}_k\left(\sum_{y\in\mathbb{Y}} T(\pi,y)\frac{\sigma(\pi,y)}{\sigma(\pi,a)}P(a|y,\pi)\right) \\ &\geq \sum_{y\in\mathbb{Y}} \underline{V}_k(T(\pi,y))\frac{\sigma(\pi,y)}{\sigma(\pi,a)}P(a|y,\pi) \end{aligned}$$

implying

$$\sum_a \underline{V}_k(T^\pi(\pi,a))\sigma(\pi,a) \geq \sum_y \underline{V}_k(T(\pi,y)\sigma(\pi,y).$$

C) Proof of Theorem 2: Here, we present a detailed version of Theorem 2 that was presented in Section IV-A.

Theorem 2 (Detailed Version): Under (A1), (A2), (S),

i) The local decision $a^*(\pi,y) = \arg\min_a c_a' B_y P'\pi$ (see (9)) is increasing in y.
ii) $a^*(\pi,y)$ is MLR increasing in π, i.e., $\pi \geq_r \bar{\pi} \Rightarrow a^*(\pi,y) \geq a^*(\bar{\pi},y)$.

iii) The Y linear hyperplanes $(c_1 - c_2)'B_y P'\pi = 0$, $y = 1, \ldots, Y$ do not intersect within the interior of the belief space $\Pi(X)$. Thus, out of the 2^Y polytopes in (30), there are a maximum of $Y+1$ nonempty polytopes in Π, namely (31).

iv) Let $i_y^* = \max\{i : e_i \in \{\pi : (c_1 - c_2)'B_y P'\pi < 0\}$. Then each of the Y hyperplanes $(c_1 - c_2)'B_y P'\pi = 0$, $y = 1, \ldots, Y$ partitions $\Pi(X)$ such that the vertices $e_1, e_2, \ldots, e_{i_y^*}$ lie in the convex polytope $(c_1 - c_2)'B_y P'\pi < 0$ and the vertices $e_{i_y^*+1}, \ldots, e_X$ lie in the convex polytope $(c_1 - c_2)'B_y P'\pi > 0$.

v) i_y^* decreases with y.

vi) M^π defined in (13) has the following structure:

$$M^\pi = \begin{bmatrix} 1_{Y-l+1} & 0_{Y-l+1} \\ 0_{l-1} & 1_{l-1} \end{bmatrix}, \text{ for } \pi \in \mathcal{P}_l, l = 1, \ldots, Y+1 \tag{63}$$

■

Proof:

i) From [33, Lemma 1.2(1)] if B and P are TP2 (i.e., (A1), (A2) hold) then $\frac{B_y P'\pi}{\mathbf{1}'B_y P'\pi} \leq_r \frac{B_{y+1} P'\pi}{\mathbf{1}'B_{y+1} P'\pi}$. Next MLR dominance implies first-order stochastic dominance. Then since from (S), $c(i,1) - c(i,2)$ is increasing in i, it follows that $\frac{(c_1-c_2)'B_y P'\pi}{\mathbf{1}'B_y P'\pi} < \frac{(c_1-c_2)'B_{y+1} P'\pi}{\mathbf{1}'B_{y+1} P'\pi}$. Since the denominators are nonnegative, this implies that $(c_1 - c_2)'B_y P'\pi \geq 0 \Rightarrow (c_1 - c_2)'B_{y+1} P'\pi \geq 0$. That is, the single crossing condition (36) holds, see Definition 5. So $a^*(\pi, y) \uparrow y$.

ii) To prove $a^*(\pi, y) \uparrow \pi$ w.r.t $\geq_r$, we use a similar approach to Part (i). From [33, Lemma 1.2(2)], assuming (A3), $\pi \leq_r \bar{\pi}$ implies $\frac{B_y P'\pi}{\mathbf{1}'B_y P'\pi} \leq_r \frac{B_y P'\bar{\pi}}{\mathbf{1}'B_y P'\bar{\pi}}$. As in the proof above, using (S) this implies $\frac{(c_1-c_2)'B_y P'\pi}{\mathbf{1}'B_y P'\pi} < \frac{(c_1-c_2)'B_y P'\bar{\pi}}{\mathbf{1}'B_y P'\bar{\pi}}$. Since the denominators are nonnegative, this implies that

$$\pi \leq_r \bar{\pi} \text{ and } (c_1 - c_2)'B_y P'\pi \geq 0 \Rightarrow (c_1 - c_2)'B_y P'\bar{\pi} \geq 0. \tag{64}$$

That is a single crossing condition (see Definition 5) holds w.r.t (π, a) and the partial order $\geq_r$. So $a^*(\pi, y) \uparrow \pi$.

iii) follows immediately from (36).

iv) Since $i_y^* = \max\{i : e_i \in \{\pi : (c_1 - c_2)'B_y P'\pi < 0\}$, clearly $e_{i_y^*+1} \in \{\pi : (c_1 - c_2)'B_y P'\pi > 0\}$. Next since $e_{i_y^*+1} \leq_r e_{i_y^*+2} \cdots \leq_r e_X$, the single crossing condition (64) yields $(c_1-c_2)'B_y P' e_{i_y^*+2} \geq 0, \ldots, (c_1-c_2)'B_y P' e_X \geq 0$.

v) Start with the single crossing condition (36) repeated below for clarity

$$\{\pi : (c_1 - c_2)'B_{y+1} P'\pi \leq 0\} \subseteq \{\pi : (c_1 - c_2)'B_y P'\pi \leq 0\}.$$

Therefore

$$\max\{i : e_i \in \{\pi : (c_1 - c_2)'B_{y+1} P'\pi \leq 0\} \leq \max\{i : e_i \in \{\pi : (c_1 - c_2)'B_y P'\pi \leq 0\}$$

vi) follows by enumerating all matrices M^π that satisfy (i) and (ii); see (33) for an example.

D) Proof of Theorem 3: Similar to the example given below Theorem 2, it can be verified from (13) that there are only 3 possible values for R^π, namely

$$R^\pi = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \pi \in \mathcal{P}_1, \quad R^\pi = B, \pi \in \mathcal{P}_2$$
$$R^\pi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \text{ and } \pi \in \mathcal{P}_3. \tag{65}$$

Thus, Bellman's (25) reads

$$\begin{aligned} V(\pi) = \min\{&C(\pi, 2) + \rho V(\pi) I(\pi \in \mathcal{P}_1) \\ &+ \rho \sum_{a \in \mathbb{A}} V(T^\pi(\pi, a)) \sigma(\pi, a)\, I(\pi \in \mathcal{P}_2) \\ &+ \rho V(\pi) I(\pi \in \mathcal{P}_3), 0\} \end{aligned} \tag{66}$$

Claim (i): For $\pi \in \mathcal{P}_1 \cup \mathcal{P}_3$, $V(\pi) = \min\{C(\pi, 2) + \rho V(\pi), 0\}$. This can be solved explicitly as

$$V(\pi) = \min\{C(\pi, 2)/(1-\rho), 0\}$$
$$\text{implying } \mu^*(\pi) = \begin{cases} 1, & C(\pi, 2) < 0 \\ 2, & C(\pi, 2) \geq 0 \end{cases}.$$

Since $C(\pi, 2)$ is MLR decreasing in π, the optimal policy for $\pi \in \mathcal{P}_1 \cup \mathcal{P}_3$ is a threshold policy with threshold at $C(\pi^*, 2) = 0$. This proves the first claim of the theorem.

Claim (ii): Since $T^\pi(\pi, a) = \pi$ for $\pi \in \mathcal{P}_1 \cup \mathcal{P}_3$, the private belief state update (7) freezes in these regions, i.e., $\pi_{k-1} \in \mathcal{P}_1 \cup \mathcal{P}_3$ implies that, $\eta_k = \pi_{k-1}$. Therefore all agents take the same local decision a according to (9) implying an information cascade.

Claim (iii): The proof of this is more involved.

We need the following property of the social learning Bayesian filter which is a detailed version of Lemma 2 in Section IV-B. Since we are going to partition $\Pi(X)$ into four intervals, namely $[0, \eta_2(2))$, $[\eta_2(2), q(2))$, $[q(2), \eta_1(2))$ and $[\eta_1(2), 1]$, it is convenient to introduce the following notation: Denote these intervals as $\bar{\mathcal{P}}^1$, $\bar{\mathcal{P}}^2$, $\bar{\mathcal{P}}^3$, $\bar{\mathcal{P}}^4$, respectively. Note $\mathcal{P}_1 = \bar{\mathcal{P}}_1$, $\mathcal{P}_2 = \bar{\mathcal{P}}_2 \cup \bar{\mathcal{P}}_3$, $\mathcal{P}_3 = \bar{\mathcal{P}}_4$.

Lemma 2 (Detailed Version): Consider the social learning Bayesian filter (11). Then, $T^{\eta_1}(\eta_1, 1) = q$, $T^{\eta_2}(\eta_2, 2) = q$. Furthermore, if B is symmetric TP2, then $T^q(q, 2) = \eta_1$, $T^q(q, 1) = \eta_2$ and $\eta_2 \leq_r q \leq_r \eta_1$. So

i) $\pi \in \bar{\mathcal{P}}_2$ implies $T^\pi(\pi, 2) \in \bar{\mathcal{P}}_1$ and $T^\pi(\pi, 1) \in \bar{\mathcal{P}}_3$.

ii) $\pi \in \bar{\mathcal{P}}_3$ implies $T^\pi(\pi, 2) \in \bar{\mathcal{P}}_2$ and $T^\pi(\pi, 1) \in \bar{\mathcal{P}}_4$. ■

The proof of Lemma 2 is as follows. Recall from (65) that on interval $\mathcal{P}_2$, $R^\pi = B$. Then, it is straightforwardly verified from (11) that $T^{\eta_1}(\eta_1, 1) = T^{\eta_2}(\eta_2, 2) = q$. Next, using (11) it follows that $B_{12}B_{11} = B_{22}B_{21}$ is a sufficient condition for $T^q(q, 2) = \eta_1$ and $T^q(q, 1) = \eta_2$. Also, since by (A1) B is TP2, applying Theorem 10 (2), implies $\eta_2 \leq_r q \leq_r \eta_1$. So B symmetric TP2 is sufficient for the claims of the lemma to hold. Statements (i) and (ii) then follow straightforwardly. In particular, from Theorem 10(1), $\eta_1 \geq_r \pi \geq_r q$ implies $T^{\eta_1}(\eta_1, 1) = q \geq_r T^\pi(\pi, 1) \geq_r T^q(q, 1) = \eta_2$, which implies Statement (i) of the lemma. Statement (ii) follows similarly.

Returning to the proof of Theorem 3, we use mathematical induction on the value iteration algorithm (27). Clearly $V_0(\pi) =$

$-\bar{C}(\pi,1)$ is linear. Assume now that $V_k(\pi)$ is piecewise linear and concave on each of the four intervals $\bar{\mathcal{P}}_1,\ldots\bar{\mathcal{P}}_4$. That is, for two dimensional vectors γ_{m_l} in the set Γ_l

$$V_k(\pi)=\sum_l \min_{m_l\in\Gamma_l}\gamma'_{m_l}\pi\, I(\pi\in\bar{\mathcal{P}}_l).$$

Consider $\pi\in\bar{\mathcal{P}}_2$. From (65), since $R^\pi_a=B_a$, $a=1,2$, Lemma 2 (i) together with the value iteration algorithm (66) yields

$$V_{n+1}(\pi)$$
$$=\min\{C(\pi,2)+\rho\left[\min_{m_3\in\Gamma_3}\gamma'_{m_3}B_1\pi+\min_{m_1\in\Gamma_1}\gamma'_{m_1}B_2\pi\right],0\}.$$

Note the crucial point in the above equation: as a result of Lemma 2 (i)—the social learning filter maps $\bar{\mathcal{P}}_2$ to only $\bar{\mathcal{P}}_1$ (for $a=1$) and $\bar{\mathcal{P}}_3$ (for $a=2$). Since each of the terms in the above equation are piecewise linear and concave, it follows that $V_{k+1}(\pi)$ is piecewise linear and concave on $\bar{\mathcal{P}}_2$. A similar proof holds for $\bar{\mathcal{P}}_3$ and this involves using Lemma 2 (ii). As a result the stopping set on each interval $\bar{\mathcal{P}}_l$, $l=1,\ldots,4$ is a convex region, i.e., an interval. This proves claim (ii).

E) Proof of Lemma 3 and Theorem 4:

Proof of Lemma 3: Let us introduce the following notation. Define

$$S^+=\{\pi: C'\pi>0\}\text{ and }S^=\{\pi:C'\pi=0\}. \tag{67}$$

The proof comprises three parts.

Statement (i): Under (PH), for every $\pi\in S^+$, there exists a $\bar{\pi}\in S^=$ such that $\bar{\pi}\geq_r\pi$.

Proof: Consider any belief state $\pi\in S^+$. Construct a line segment from e_1 through the belief state π and let this line segment intersect the hyperplane $S^=$. Denote $\bar{\pi}$ as this point of intersection. Clearly, $\bar{\pi}=\alpha e_1+(1-\alpha)\pi$ where $\alpha=C'\pi/(C'\pi-C_1)$. It is straightforwardly established that $\bar{\pi}\geq_r\pi$ if $\alpha\geq 1$ which is clearly true since $C_1>0$ and $C'\pi>0$ for $\pi\in S^+$.

Statement (ii): Under (A1), (A2), and (A3), if $\bar{\pi}\geq_r\pi$, then $C'T^{\bar{\pi}}(\bar{\pi},a)>0\Rightarrow C'T^\pi(\pi,a)>0$.

Proof: Under (A1), (A2), and (A3), it follows from Theorem 10(2) in Appendix F that $T^\pi(\pi,a)$ is MLR increasing, that is, $\bar{\pi}\geq_r\pi$ implies $T^{\bar{\pi}}(\bar{\pi},a)\geq_r T^\pi(\pi,a)$. Under (A3), the elements of C are decreasing. So from Result 1 in Appendix A, it follows that $C'T^{\bar{\pi}}(\bar{\pi},a)\leq C'T^\pi(\pi,a)$. So $C'T^{\bar{\pi}}(\bar{\pi},a)>0$ implies $C'T^\pi(\pi,a)>0$.

Statements (i) and (ii) imply that if the social learning filter (11) maps belief states in $S^=$ to S, then all belief states in $\{\pi: C'\pi>0\}$ are also mapped to S. Since the hyperplane $S^=\{\pi:C'\pi=0\}$ has infinite points, how can we formulate a sufficient condition for belief states $\{\pi:C'\pi=0\}$ to be mapped to the polytope $\mathcal{P}_{Y+1}$? (C1) serves as a sufficient condition as proved in Statement (iii) as follows.

Statement (iii): A sufficient condition for $C'T^{\bar{\pi}}(\bar{\pi},a)<0$ to hold for all $\bar{\pi}\in S^=$ is that $C'T^{\nu_i}(\nu_i,a)<0$ for all $X-1$ vertices ν_j of (45).

Proof: Clearly, every belief state $\pi\in S^=$ is a convex combination of the vertices, i.e., $\pi=\sum_i\alpha_i\nu_i$, for some $\alpha_i\geq 0$ and $\sum_i\alpha_i=1$. Now $C'T^{\nu_i}(\nu_i,a)<0$ is equivalent to $C'R^{\nu_i}P'\nu_i<0$ since the normalization term in $T^\pi(\cdot)$ is nonnegative. This implies $C'\sum_i\alpha_iR^{\nu_i}P'\nu_i<0$, and this is equivalent to $C'R^\pi P'\pi<0$.

Proof of Theorem 4: Define $S=\{\pi:C'\pi\geq 0\}$.

Step 1: We first prove that $V(\pi)=0$ for $\pi\in S$. This is equivalent to saying that for $\{\pi:C'\pi\geq 0\}$, the optimal policy $\mu^*(\pi)=1$.

The proof of Step 1 is by induction on the value iteration algorithm (27). Suppose $V_0(\pi)=0$. Then it trivially satisfies $V_0(\pi)=0$ for $\pi\in S$. Next suppose $V_k(\pi)=0$ for $\pi\in S$. Then for $\pi\in S$, Assumption (C1) implies that $T^\pi(\pi,a)$ belongs to S implying that $V(T^\pi(\pi,a))=0$. So from (27), it follows that $V_{k+1}(\pi)=\min\{C'\pi,0\}=0$ since $C'\pi\geq 0$ for $\pi\in S$. Since $V_k(\pi)$ converges pointwise to $V(\pi)$, Step 1 follows. For initial condition $V_0(\pi)=-\bar{C}(\pi,1)$ [see (27)], $V(\pi)$ obtained as the limit of the value iteration algorithm is identical to that with initial condition $V_0(\pi)=0$.

Step 2: From Bellman's equation it follows trivially that for $\{\pi:C'\pi<0\}$, $\mu^*(\pi)=2$.

From Steps 1 and 2, we have $C'\pi\geq 0$ iff $\mu^*(\pi)=1$.

F) Proof of Theorem 5: This section is in two parts. We start with several preliminary results that are similar to the results in [33]. Then, the proof of Theorem 5 is presented.

7) Structural Properties of Social Learning Filter:

Theorem 10: The following structural properties hold for the public belief update evaluated by the social learning Bayesian filter defined in (11):

1) Under (S), M^π is TP2 for $\pi\in\Pi(X)$, see Definition 6.
2) Under (A1), (A2), (S) if $\pi_1,\pi_2\in\mathcal{P}_l$, then $\pi_1\geq_r\pi_2$ implies $T^{\pi_1}(\pi_1,a)\geq_r T^{\pi_2}(\pi_2,a)$.
3) Under (A1), (A2), (S), if $\pi_1,\pi_2\in\Pi(X)$, then $\pi_1\geq_r\pi_2\Rightarrow\sigma(\pi_1,\cdot)\geq_s\sigma(\pi_2,\cdot)$.
4) Under (A1), (A2), if $\pi\in\Pi(X)$, then $a>\bar{a}$ implies $T^\pi(\pi,a)\geq_r T^\pi(\pi,\bar{a})$.

Proof:

1) We need to show that for fixed $\pi\in\Pi(X)$

$$M^\pi_{ya}M^\pi_{y'a'}\leq M^\pi_{y\wedge y',a\wedge a'}M^\pi_{y\vee y',a\vee a'}. \tag{68}$$

Recall from (13) that M^π is a matrix with a single 1 in each row at $a^*(\pi,y)$ and all other elements zero. So the only nontrivial case to prove is when both terms on the LHS are 1, i.e., $M^\pi_{y,a^*(\pi,y)}=1$ and $M^\pi_{y',a^*(\pi,y')}=1$. Assuming (A2), (S), Theorem 2(i) says that $a^*(\pi,y)\uparrow y$. This means that $y<y'\Rightarrow a^*(\pi,y)<a^*(\pi,y')$ and $y\geq y'\Rightarrow a^*(\pi,y)\geq a^*(\pi,y')$. In either case (68) holds with equality since the RHS is identical to the LHS.

2) Since P is TP2 (A2), we have $P'\pi\geq_r P'\bar{\pi}$ for $\pi\geq_r\bar{\pi}$, see [23]. So it suffices to show that $\frac{R^\pi_a\pi}{\mathbf{1}'B^\pi_a\pi}\geq_r\frac{B^{\bar{\pi}}_a\bar{\pi}}{\mathbf{1}'B^{\bar{\pi}}_a\bar{\pi}}$ for $\pi\geq_r\bar{\pi}$. Moreover, since $\pi,\bar{\pi}$ belong to the same polytope $\mathcal{P}_l$, $R^\pi_a=R^{\bar{\pi}}_a=R^l_a$ (say) [see (30)]. From [23], a sufficient condition for $\frac{R^l_a\pi}{\mathbf{1}'R^l_a\pi}\geq_r\frac{R^l_a\bar{\pi}}{\mathbf{1}'R^l_a\bar{\pi}}$ is that R^l is TP2. Of course we need this to hold on each of the $Y+1$ polytopes, i.e., for $l=1,\ldots,Y+1$.

So under what conditions is R^π TP2 in each of the $Y+1$ polytopes? Note (A2) says B is TP2. Since $R^\pi=BM^\pi$

[see (13)] and the product of TP2 matrices is TP2 [23, p. 471], it only remains to prove that M^π is TP2. This follows from (S) as proved in (i) above.

3) Since P is TP2 (A2), it suffices to prove that $\pi \geq_r \bar{\pi}$ implies $\mathbf{1}' B_a^\pi \pi \geq_s \mathbf{1}' B_a^{\bar{\pi}} \bar{\pi}$, i.e., $\sum_i \sum_{a>\bar{a}} B_{ia}^\pi \pi_i \geq \sum_i \sum_{a \geq \bar{a}} B_{ia}^{\bar{\pi}} \bar{\pi}_i$. From Statement 1, M^π and $M^{\bar{\pi}}$ are TP2 and from (A2) B is TP2. So $B_\pi = BM^\pi$ and $B^{\bar{\pi}} = BM^{\bar{\pi}}$ are TP2. Therefore, from Definition 6(iii), the rows of R^π and $R^{\bar{\pi}}$ are MLR increasing. Since MLR dominance implies first-order stochastic dominance, this means that both $\sum_{a>\bar{a}} R_{ia}^\pi$ and $\sum_{a>\bar{a}} R_{ia}^{\bar{\pi}}$ are increasing with i. Since $\pi \geq_r \bar{\pi}$, Result 1 (i), (ii), and (iii), imply that a sufficient condition for $\pi \geq_r \bar{\pi} \Rightarrow \mathbf{1}' R_a^\pi \pi \geq_s \mathbf{1}' R_a^{\bar{\pi}} \bar{\pi}$ is that $\sum_{a>\bar{a}} R_{ia}^\pi > \sum_{a>\bar{a}} R_{ia}^{\bar{\pi}}$ or equivalently, $\sum_y B_{iy} \sum_{a>\bar{a}} M_{ya}^\pi \geq \sum_y B_{iy} \sum_{a>\bar{a}} M_{ya}^{\bar{\pi}}$. A sufficient condition for this is $\pi \geq_r \bar{\pi} \Rightarrow \sum_{a>\bar{a}} M_{ya}^\pi \geq \sum_{a>\bar{a}} M_{ya}^{\bar{\pi}}$. But this condition holds from the structure of M in (63) and the fact that $a^*(\pi, y)$ is MLR increasing w.r.t π (Statement (ii) of Theorem 2 in Appendix C).

4) Since P is TP2 (A2), it suffices to prove that $\pi \geq_r \bar{\pi} \Rightarrow \frac{R_a^\pi P' \pi}{\mathbf{1}' R_a^\pi P^\pi \pi} \geq_r \frac{R_{a'}^{\bar{\pi}} P' \pi}{\mathbf{1}' R_{a'}^{\bar{\pi}} P' \pi}$ for $a \geq a'$. Since R^π is TP2 (A1), this result follows straightforwardly from [51, Theorem 4]. ■

8) Proof of Theorem 5: Here, we prove Theorem 5. The update of belief state in $\mathcal{P}_{Y+1}$ is simple, since $R_{ia}^\pi = 1/X$ (uniformly distributed) for each i (see (34) for example). In comparison, the sensor management case of Theorem 8 on $\mathcal{P}_2$ with update given by (60) requires an arbitrary TP2 matrix R^π. To allow for this generality, in the proof below, we assume R^π is an arbitrary TP2 matrix on $\mathcal{P}_{Y+1}$.

Part 1: Under (A1), (A2), (A3), (S), (C3), (C2), and (PH), $V(\pi)$ is MLR decreasing on polytope $\mathcal{P}_{Y+1}$:

The proof of Part 1 is by mathematical induction on the value iteration algorithm (27). Start with $V_0(\pi) = -\bar{C}(\pi, 1)$ in (27). Clearly, this is MLR decreasing on $\Pi(X)$ and therefore on polytope $\mathcal{P}_{Y+1}$ since $\mathbf{f}$ is chosen with increasing elements, see (17). Now for the inductive step: Assume at iteration k, $V_k(\pi)$ is MLR decreasing on polytope $\mathcal{P}_{Y+1}$. Then, since $T^\pi(\pi, a)$ is MLR increasing in a [see Theorem 10(4)] and $T(\pi, a) \in \mathcal{P}_{Y+1}$ by (C2), it follows that $V_k(T^\pi(\pi, 1)) \geq V_k(T^\pi(\pi, 2))$.

Consider any $\pi \geq_r \bar{\pi} \in \mathcal{P}_{Y+1}$. Since $\sigma(\pi, .) \geq_s \sigma(\bar{\pi}, .)$ [see Theorem 10(3)]

$$\sum_a V_k(T^\pi(\pi, a)) \sigma(\pi, a) \leq \sum_a V_k(T^\pi(\pi, a)) \sigma(\bar{\pi}, a). \quad (69)$$

Next since $\pi \geq_r \bar{\pi} \Rightarrow T^\pi(\pi, a) \geq_r T^\pi(\bar{\pi}, a)$ (Theorem 10(2)), so $V_k(\pi)$ MLR decreasing in π implies $V_k(T^\pi(\pi, a)) \leq V_k(T^\pi(\bar{\pi}, a))$. So from (69), $\pi \geq_r \bar{\pi}$ implies

$$\begin{aligned} \sum_a V_k(T^\pi(\pi, a)) \sigma(\pi, a) &\leq \sum_a V_k(T^\pi(\pi, a)) \sigma(\bar{\pi}, a) \\ &\leq \sum_a V_k(T^\pi(\bar{\pi}, a)) \sigma(\bar{\pi}, a). \quad (70) \end{aligned}$$

From (A3), $C(\pi, 2)$ is MLR decreasing. So $\pi \geq_r \bar{\pi}$ implies $C(\pi, 2) \leq C(\bar{\pi}, 2)$. Therefore, $\pi \geq_r \bar{\pi}$ implies $Q_{k+1}(\pi, 2) \leq Q_{k+1}(\bar{\pi}, 2)$. Thus, $\min_u Q_{k+1}(\pi, u) \leq \min_u Q_{k+1}(\bar{\pi}, u)$, i.e., $V_{k+1}(\pi) \leq V_{k+1}(\bar{\pi})$. This completes the induction step. Finally, since $V_k \to V$ as $k \to \infty$ pointwise [see discussion below (27)], V is MLR decreasing on polytope $\mathcal{P}_{Y+1}$.

Part 2: Under the above conditions, $\mu^*(\pi)$ is MLR increasing on polytope $\mathcal{P}_{Y+1}$. It suffices to show that $Q(\pi, u)$ is submodular (see Definition 4) on $\mathcal{P}_{Y+1}$ w.r.t the MLR ordering since then Theorem 9 applies implying that $\mu^*(\pi)$ is MLR decreasing in $\pi \in \mathcal{P}_{Y+1}$. To show that $Q(\pi, u)$ in (25) is submodular, we need to show that $Q(\pi, 2)$ is MLR decreasing in π. But this follows from (A3) and Part 1. Thus, from Theorem 9, (49) holds.

I) Proof of Theorem 7: Given any $\pi_1, \pi_2 \in \mathcal{L}(e_X, \bar{\pi})$ with $\pi_2 \geq_{L_X} \pi_1$, we need to prove: $\mu_\theta(\pi_1) \leq \mu_\theta(\pi_2)$ iff $\theta(X-2) \geq 1$, $\theta(i) \leq \theta(X-2)$ for $i < X-2$. But from the structure of (53), obviously $\mu_\theta(\pi_1) \leq \mu_\theta(\pi_2)$ is equivalent to $\begin{bmatrix} 0 & 1 & \theta' \end{bmatrix}' \begin{bmatrix} \pi_1 \\ -1 \end{bmatrix} \leq \begin{bmatrix} 0 & 1 & \theta' \end{bmatrix}' \begin{bmatrix} \pi_2 \\ -1 \end{bmatrix}$, or equivalently, $\begin{bmatrix} 0 & 1 & \theta(1) & \cdots & \theta(X-2) \end{bmatrix} (\pi_1 - \pi_2) \leq 0$.

Now from Lemma 5 (iii), $\pi_2 \geq_{L_X} \pi_1$ implies that $\pi_1 = \epsilon_1 e_X + (1-\epsilon_1)\bar{\pi}$, $\pi_2 = \epsilon_2 e_X + (1-\epsilon_2)\bar{\pi}$ and $\epsilon_1 \leq \epsilon_2$. Substituting these into the above expression, we need to prove

$$(\epsilon_1 - \epsilon_2)\big(\theta(X-2) - \begin{bmatrix} 0 & 1 & \theta(1) & \cdots & \theta(X-2) \end{bmatrix}' \bar{\pi}\big) \leq 0 \quad \forall \bar{\pi} \in \mathcal{H}_X$$

iff $\theta(X-2) \geq 1, \theta(i) \leq \theta(X-2), i < X-2$. This is obviously true.

A similar proof shows that on lines $\mathcal{L}(e_1, \bar{\pi})$ the linear threshold policy satisfies $\mu_\theta(\pi_1) \leq \mu_\theta(\pi_2)$ iff $\theta(i) \geq 0$ for $i < X-2$.

REFERENCES

[1] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games Appl.*, pp. 1–47, 2010.

[2] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, Cascades in networks and aggregate volatility Nat. Bureau Economic Res., Cambridge, MA, 2010, Tech. Rep.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[4] R. Amir, "Supermodularity and complementarity in economics: An elementary survey," *South. Econ. J.*, vol. 71, no. 3, pp. 636–660, 2005.

[5] M. S. Andersland and D. Teneketzis, "Measurement scheduling for recursive team estimation," *J. Optim. Theory Appl.*, vol. 89, no. 3, pp. 615–636, Jun. 1996.

[6] A. Banerjee, "A simple model of herd behavior," *Quat. J. Econ.*, vol. 107, pp. 797–817, 1992.

[7] P. Bartlett and J. Baxter, "Estimation and approximation bounds for gradient-based reinforcement learning," *J. Comput. Syst. Sci.*, vol. 64, no. 1, pp. 133–150, 2002.

[8] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes—Theory and Applications*, ser. Information and System Sciences Series. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[9] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 2000, vol. 1/2.

[10] S. Bikchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as information cascades," *J. Political Economy*, vol. 100, pp. 992–1026, 1992.

[11] L. Bru and X. Vives, "Informational externalities, herding, and incentives," *J. Inst. Theor. Econ.*, vol. 158, no. 1, pp. 91–105, 2002.

[12] C. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge, MA: Cambridge Univ. Press, 2004.

[13] C. Chamley and D. Gale, "Information revelation and strategic delay in a model of investment," *Econometrica*, vol. 62, no. 5, pp. 1065–1085, 1994.

[14] Y. Chen, Q. Zhao, V. Krishnamurthy, and D. Djonin, "Transmission scheduling for optimizing sensor network lifetime: A stochastic shortest path approach," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2294–2309, May 2007.

[15] T. Cover and M. Hellman, "The two-armed-bandit problem with time-invariant finite memory," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 2, pp. 185–195, Mar. 1970.
[16] F. R. Gantmacher, *Matrix Theory*. New York: Chelsea, 1960, vol. 2.
[17] S. Hart, "Adaptive heuristics," *Econometrica*, vol. 73, no. 5, pp. 1401–1430, 2005.
[18] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, 2000.
[19] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, pp. 765–782, 1970.
[20] O. Hernández-Lerma and J. B. Laserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. New York: Springer-Verlag, 1996.
[21] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research*. New York: McGraw-Hill, 1984, vol. 2.
[22] S. Karlin, *Total Positivity*. Stanford, CA: Stanford Univ., 1968, vol. 1.
[23] S. Karlin and Y. Rinott, "Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions," *J. Multivariate Anal.*, vol. 10, pp. 467–498, 1980.
[24] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. New York: Van Nostrand, 1960.
[25] M. Kijima, *Markov Processes for Stochastic Modelling*. London, U.K.: Chapman & Hall, 1997.
[26] V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty—A lattice programming POMDP approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 7096–7124, Oct. 2011.
[27] V. Krishnamurthy, M. Maskery, and G. Yin, "Decentralized activation in a ZigBee-enabled unattended ground sensor network: A correlated equilibrium game theoretic analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6086–6101, Dec. 2008.
[28] V. Krishnamurthy and B. Wahlberg, "POMDP multiarmed bandits—Structural results," *Math. Oper. Res.*, vol. 34, no. 2, pp. 287–302, May 2009.
[29] V. Krishnamurthy and G. Yin, "Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 458–476, Feb. 2002.
[30] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Recursive Algorithms and Applications*, 2nd ed. New York: Springer-Verlag, 2003.
[31] I. Lobel, D. Acemoglu, M. Dahleh, and A. E. Ozdaglar, "Preliminary results on social learning with partial observations," presented at the presented at the 2nd Int. Conf. Performance Evaluation Methodologies Tools, Nantes, France, 2007.
[32] W. S. Lovejoy, "On the convexity of policy regions in partially observed systems," *Oper. Res.*, vol. 35, no. 4, pp. 619–621, Jul.–Aug. 1987.
[33] W. S. Lovejoy, "Some monotonicity results for partially observed Markov decision processes," *Oper. Res*, vol. 35, no. 5, pp. 736–743, Sep.–Oct. 1987.
[34] G. B. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, pp. 1379–1387, 1986.
[35] A. Muller and D. Stoyan, *Comparison Methods for Stochastic Models and Risk*. New York: Wiley, 2002.
[36] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. New York: Marcel Dekker, 1989.
[37] J. Papastravrou, J. Pothiawala, and M. Athans, Designing an Organization in a Hypothesis Testing Framework Lab. Inf. Decision Syst., Massachusetts Inst. Technol., 1989, Tech. Rep.
[38] A. Park and H. Sabourian, "Herding and Contrarian behavior in financial markets," *Econometrica*, vol. 79, no. 4, pp. 973–1026, 2011.
[39] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge: Cambridge Univ. Press, 2008.
[40] U. Rieder, "Structural results for partially observed control models," *Methods Models Oper. Res.*, vol. 35, pp. 473–490, 1991.
[41] S. Ross, *Introduction to Stochastic Dynamic Programming*. San Diego, CA: Academic, 1983.
[42] S. Ross, M. Izadi, M. Mercer, and D. Buckeridge, "Sensitivity analysis of POMDP value functions," in *Proc. IEEE Int. Conf. Mach. Learning Appl.*, 2009, pp. 317–323.
[43] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Prob. Appl.*, vol. 8, no. 22, 1963.
[44] A. N. Shiryayev, *Optimal Stopping Rules*. New York: Springer-Verlag, 1978.
[45] L. Smith and P. Sorensen, "Informational herding and optimal experimentation," in Economics Papers 139, Economics Group. Oxford, U.K., 1997, Nuffield College, Univ. Oxford.
[46] L. Smith and P. Sorenson, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, 2000.
[47] J. Spall, *Introduction to Stochastic Search and Optimization*. New York: Wiley, 2003.
[48] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Prob. Appl.*, vol. 49, no. 3, pp. 458–497, 2005.
[49] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ: Princeton Univ. Press, 1998.
[50] C. C. White and D. P. Harrington, "Application of Jensen's inequality to adaptive suboptimal design," *J. Optim. Theory Appl.*, vol. 32, no. 1, pp. 89–99, 1980.
[51] W. Whitt, "A note on the influence of the sample on the posterior distribution," *J. Amer. Statist. Assoc.*, vol. 74, pp. 424–426, 1979.
[52] B. Yakir, "A note on optimal detection of a change point in distribution," *Ann. Statist.*, vol. 25, pp. 2117–2126, 1997.
[53] B. Yakir, A. M. Krieger, and M. Pollak, "Detecting a change in regression: First-order optimality," *Ann. Statist.*, vol. 27, no. 6, pp. 1896–1913, 1999.

Vikram Krishnamurthy (S'90–M'91–SM'99–F'05) was born in 1966. He received his bachelor's degree from the University of Auckland, New Zealand in 1988, and Ph.D. from the Australian National University in 1992. He currently is a professor and Canada Research Chair at the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada.

Dr Krishnamurthy's current research interests include computational game theory, stochastic control in sensor networks, and stochastic dynamical systems for modeling of biological ion channels and biosensors. Dr. Krishnamurthy currently serves as Editor in Chief of IEEE Journal Selected Topics in Signal Processing. He has served as associate editor for several journals including IEEE Transactions Automatic Control and IEEE Transactions on Signal Processing. In 2009–2010, he served as Distinguished lecturer for the IEEE signal processing society.