How to Schedule Measurements of a Noisy Markov Chain in Decision Making?

Vikram Krishnamurthy, *Fellow, IEEE*

***Abstract*—A decision maker records measurements of a finite-state Markov chain corrupted by noise. The goal is to decide when the Markov chain hits a specific target state. The decision maker can choose from a finite set of sampling intervals to pick the next time to look at the Markov chain. The aim is to optimize an objective comprising of false alarm, delay cost, and cumulative measurement sampling cost. Taking more frequent measurements yields accurate estimates but incurs a higher measurement cost. Making an erroneous decision too soon incurs a false alarm penalty. Waiting too long to declare the target state incurs a delay penalty. What is the optimal sequential strategy for the decision maker? This paper shows that under reasonable conditions, the optimal strategy has the following intuitive structure: when the Bayesian estimate (posterior distribution) of the Markov chain is away from the target state, look less frequently; while if the posterior is close to the target state, look more frequently. Bounds are derived for the optimal strategy. Also the achievable optimal cost of the sequential detector as a function of transition dynamics and observation distribution is analyzed. The sensitivity of the optimal achievable cost to parameter and strategy variations is bounded in terms of the Kullback divergence. Also structural results are obtained for joint optimal sampling and measurement control (active sensing).**

***Index Terms*—Change detection, optimal sequential sampling, quickest state estimation, decision making, active sensing, Bayesian filtering, stochastic dominance, submodularity, stochastic dynamic programming, partially observed Markov decision process.**

I. Introduction and Examples

A. The Problem

CONSIDER the following quickest detection optimal sampling problem which is a special case of the problem considered in this paper. Let $\tau_k, k = 0, 1, \ldots$ denote integer-valued time instants at which decisions to observe a noisy finite state Markov chain are made. Assume $\tau_0 = 0$. As it accumulates measurements over time, a decision-maker needs to announce when the Markov chain hits a specific absorbing target state. At each decision time τ_k, the decision maker chooses its decision u_k from the action set $\mathcal{U} = \{0, 1, \ldots, L\}$ where we have the following.

1) Decision $u_k = 0$ made at time τ_k corresponds to "announce the target state and stop." When this decision is made, the problem terminates at time τ_k with possibly a false alarm penalty (if the Markov chain was not in the target state).
2) Decision $u_k \in \{1, 2, \ldots, L\}$ at time τ_k corresponds to: "Look at noisy Markov chain next at time $\tau_{k+1} = \tau_k + D_{u_k}$." Here, $D_1 < D_2 < \cdots < D_L$ are fixed positive integers. They denote the set of possible time intervals to sample the Markov chain next.

Suppose the Markov chain hits the absorbing target state at time Γ. Given the history of past measurements and decisions, how should the decision-maker choose its decisions $u_k, k = 0, 1, \ldots$? Let τ_{k^*} denote the time at which the decision maker chooses $u_{k^*} = 0$, i.e., announces that the Markov chain has hit the target state. The decision-maker considers the following costs:

i) *False alarm penalty:* If $\tau_{k^*} < \Gamma$, i.e., the Markov chain is not in the target state, but the decision-maker chooses $u^*_k = 0$ and so falsely announces that the chain has hit the target state, it pays a unit nonnegative false alarm penalty.
ii) *Delay penalty*: Suppose $\tau_{k^*} > \Gamma$, i.e., the Markov chain hits the target state at time Γ and the decision-maker announces $u_{k^*} = 0$ at a later time τ_{k^*}. Then, it pays a nonnegative delay penalty d at each time instant between Γ and τ_{k^*}. So the total delay penalty is $d \max(\tau_{k^*} - \Gamma, 0)$.
iii) *Sampling cost*: At each decision time τ_k, if the decision maker takes action $u_k \in \{1, 2, \ldots, L\}$, then it pays a nonnegative measurement (sampling) cost m to look at the noisy Markov chain at time $\tau_{k+1} = \tau_k + D_{u_k}$. (More generally, m can depend on the Markov state and decision u.)

Suppose the Markov chain starts with initial distribution π_0 at time 0. What is the optimal sampling strategy μ^* for the decision-maker to minimize the following combination of the false alarm rate, delay penalty, and measurement cost? That is, determine $\mu^* = \arg\inf_\mu J_\mu(\pi_0)$ where[1]

$$J_\mu(\pi_0) = d\, \mathbb{E}^\mu_{\pi_0}\{\max(\tau_{k^*} - \Gamma, 0)\} + \mathbb{P}^\mu_{\pi_0}(\tau_{k^*} < \Gamma) + m\, \mathbb{E}^\mu_{\pi_0}\left\{\sum_{k=0}^{k^*-1} I(u_k \neq 0)\right\} \tag{1}$$
$$\text{subject to } \tau_{k^*} = u_0 + u_1 + \cdots + u_{k^*-1}, \quad u_0 = \mu(\pi_0).$$

[1]By definition, the sampling cost for looking at the noisy Markov chain at time τ_{k^*} is paid at time τ_{k^*-1}. Hence, the summation in the last term of (1) is up to $k^* - 1$.

Manuscript received August 10, 2012; revised February 27, 2013; accepted March 07, 2013. Date of publication March 19, 2013; date of current version June 12, 2013. This work was supported in part by NSERC, 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 G. V. Moustakides, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2013.2253352

Here, $I(\cdot)$ denotes the indicator function. Also, μ denotes a stationary strategy of the decision maker. $\mathbb{P}^{\mu}$ and $\mathbb{E}^{\mu}_{\pi_0}$ are the probability measure and expectation of the evolution of the observations and Markov state which are strategy dependent (These are defined in Section II). Taking frequent measurements yields accurate estimates but incurs a higher measurement cost. Making an erroneous decision too soon incurs a false alarm penalty. Waiting too long to declare the target state incurs a delay penalty.

B. Generalizations

In the special case when the Markov chain has two states (equivalently, the change time Γ is geometrically distributed), action space $\mathcal{U} = \{0\ (\text{announce change}), 1\ (\text{continue})\}$, measurement cost $m = 0$, then (1) becomes the classical Kolmogorov–Shiryayev quickest detection problem [28], [31]. This paper generalizes this in three nontrivial ways.

1. *Quickest state estimation:* First, the Markov chain can jump multiple random number of times into and out of the target state (unlike quickest detection where the process jumps once into an absorbing target state). Such "quickest state" estimation problems seek to announce when the Markov chain is in the target state as quickly as possible. They arise in financial and active sensing applications (see discussion below). The aim is to determine the optimal sampling strategy that minimizes a combination of false alarm rate, delay penalty, and measurement cost to announce if the Markov chain is currently in the target state.
2. *Multiple continue actions:* Second, unlike classical quickest detection, there are now multiple "continue" actions $u \in \{1, 2, \ldots, L\}$ corresponding to different sampling intervals $\{D_1, D_2, \ldots, D_L\}$. (In quickest detection, there is only one continue action and one stop action.) Each of these "continue" actions results in different dynamics of the posterior distribution and incurs different costs. Also, the measurement costs can be state, action, and observation dependent.
3. *Multiple state Markov chain:* Finally, the underlying Markov chain can have more than two states. As described in [25], a phase-distributed (PH-distributed) change time Γ can be modeled as a multistate Markov chain with an absorbing state. PH-distributions form a dense subset for the set of all distributions; see [18] for quickest detection. PH-distributed change times are used widely to model discrete event systems [25] and are a natural candidate for modeling arrival/demand processes for services that have an expiration date [11].

C. Context: Optimality of Monotone Strategies

This paper analyzes the structure of the optimal sampling strategy. The problem is an instance of a partially observed Markov decision process (POMDP) [9]. In general (worst case), solving a POMDP is computationally intractable [26]. However, the optimal sampling problem results in a POMDP that has a *monotone* optimal strategy and hence a finite dimensional characterization. To illustrate this structure via a numerical example, assume the decision maker observes a two state Markov chain via a binary erasure channel (parameters specified in Example 1, Section VII). Suppose the action space is $\{0\ \ (\text{announce change}), 1, 2, 3, 4\}$ where actions $1, 2, 3, 4$ correspond to choosing sampling intervals of $D_1 = 1, D_2 = 3, D_3 = 5, D_4 = 10$, respectively. (So at each decision time, the decision maker can either stop or look at the Markov chain in 1, 3, 5 or 10 time points.) Fig. 1(a) and (c) shows the optimal sampling strategy (computed via stochastic dynamic programming) for quickest detection (where one state is absorbing) and quickest state estimation (with no absorbing state). The horizontal axis denotes the posterior distribution $\pi(1)$ of the target state while the vertical axis denotes the optimal action taken. For example, in Fig. 1(a), for $\pi(1) < \pi_4^*$, it is optimal to look every ten time points at the noisy Markov chain, for $\pi(1)$ in the interval $[\pi_4^*, \pi_3^*]$ look every five points at the noisy Markov chain, etc. The key point in Fig. 1(a) and (c) is that the optimal strategy is *monotone decreasing* in the posterior $\pi(1)$.

Intuition suggests that under suitable assumptions, such a monotone strategy is sensible: Since the decision maker is interested in detecting when the Markov chain hits the target state, there is little point in incurring a measurement cost by looking at the Markov chain when its posterior suggests that it is far away

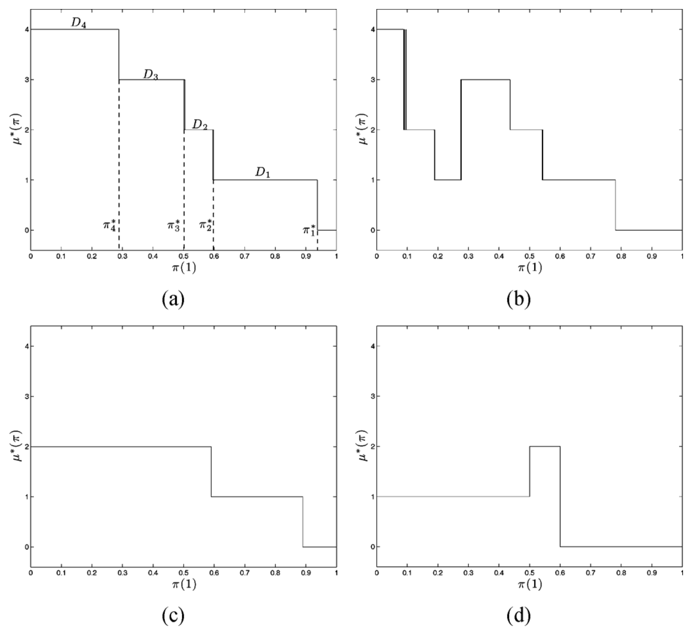


Fig. 1. Optimal sampling strategy $\mu^*(\pi)$ for a quickest-change detection and quickest estimation problem. (a) and (b) are for quickest detection where a two-state Markov chain jumps once with geometric distribution to the target state. (c) and (d) are for a quickest estimation problem where a two-state Markov chain jumps in and out of the target state. The action space is $u \in \{0\ (\text{announce change})\ , 1, 2, 3, 4\}$ where actions $1, 2, 3, 4$ correspond to choosing sampling intervals $D_1 = 1, D_2 = 3,\ D_3 = 5,\ D_4 = 10$, respectively. The noisy observations are from a binary erasure channel and the parameters are specified in Example 1 of Section VII. (a) and (c) depict monotone decreasing optimal strategies in posterior $\pi(1)$. Theorems 1 and 2 give sufficient conditions so that the optimal sampling strategy $\mu^*(\pi)$ has this monotone structure. The threshold values in (a), namely, $\pi_1^*, \pi_2^*, \pi_3^*, \pi_4^*$ give a finite dimensional characterization of the optimal strategy. (b) and (d) give examples where the conditions of Theorem 2 are violated and the optimal strategy is no longer monotone in $\pi(1)$.

from the target state. (The target state has posterior $\pi(1) = 1$.) If posterior $\pi(1)$ gets close to 1, then the decision maker should pay a higher sampling cost and look more frequently. Finally, if $\pi(1)$ is sufficiently close to 1, the decision maker should announce the target state has been reached to avoid paying a delay penalty.

This paper shows that under reasonable conditions, the optimal sampling strategy has this monotone structure. The usefulness of the main results (see Theorems 1 and 2) are enhanced by noting that without introducing appropriate conditions, the optimal strategy is not necessarily monotone. Fig. 1(b) and (d) gives examples where the sufficient conditions of Theorem 2 are violated and the optimal strategy is no longer monotone. In particular, Fig. 1(d) (quickest state estimation) shows that the optimal policy can be nonmonotone if the transition matrix does not satisfy the totally positive (TP2) assumption of Section III. It shows that the transition probabilities play an important role in the structure of the optimal strategy in quickest state estimation. This highlights an important difference between quickest state estimation and quickest detection. In quickest detection with geometric distributed change time, the transition probability matrix (with absorbing state) satisfies the TP2 assumption automatically, and so is irrelevant in the existence of a monotone optimal strategy.

D. Main Results, Organization, and Related Works

Main Results and Organization: This paper establishes the following structural results.

i) For two-state Markov chains observed in noise, since the elements of the 2-D posterior probability mass function add to 1, it suffices to consider one element of this posterior, denoted $\pi(1)$ – note $\pi(1)$ is a probability and lies in the interval $[0, 1]$. Theorem 1 shows that for quickest detection optimal sampling problems, the optimal sampling strategy of the decision-maker has a monotone structure in the posterior distribution. More generally, Theorem 2 shows that for quickest state estimation problems (where the target state is not necessarily absorbing), the optimal sampling strategy continues to have this monotone structure on a subset of the space of posterior distributions. The monotone structure of Theorems 1 and 2 reduces a dynamic programming problem on the space of posterior distributions to a finite dimensional optimization, since a monotone strategy with L possible actions has at most $L-1$ thresholds in the space of posterior distributions. For example, in Fig. 1, one only needs to compute/estimate the threshold values $\pi_1^*, \pi_2^*, \pi_3^*, \pi_4^*$ to determine the optimal strategy. The threshold values can be estimated via simulation based stochastic approximation.
The main idea is to give sufficient conditions so that the optimal strategy $\mu^*(\pi) = \operatorname{argmin}_u Q(\pi, u)$ is "monotone" in posterior π, where Q is obtained from stochastic dynamic programming, and "monotone" is with respect to a partial order (monotone likelihood ratio (MLR) stochastic order) of posterior distributions π. Lattice programming and "monotone comparative statics" [2], [3], [33] provide a general set of sufficient conditions for the existence of monotone strategies.

ii) For general-state Markov chains observed in noise, the posterior lies in a multidimensional unit simplex and so the posteriors can only be partially ordered. For three or more states, determining sufficient conditions for the optimal sampling strategy to have a monotone structure is still an open problem [23], [29]. This motivates the question: Can the optimal strategy be lower and upper bounded by monotone policies? Theorem 4 shows that the optimal sampling strategy can be lower- and upper-bounded by judiciously chosen myopic strategies on the unit simplex of posterior distributions. Such myopic strategies form easily computable bounds to the optimal strategy. Sufficient conditions are also given for the myopic strategy to have a monotone structure with respect to the MLR stochastic order. MLR stochastic dominance is ideally suited for Bayesian problems since it is preserved under conditional expectations. Theorem 5 illustrates the result for quickest detection with PH-distributed change time.

iii) How does the optimal expected sampling cost vary with transition matrix and noise distribution? Is it possible to order these parameters such that the larger they are, the larger the optimal sampling cost? For multi-state Markov chains observed in noise, Theorem 6 examines how the cost achieved by the optimal sampling strategy varies with transition matrix (state dynamics) and observation matrix (noise distribution). In particular, dominance measures are introduced for the transition matrix and observation distribution (Blackwell dominance) that result in the optimal cost increasing with respect to this dominance order. Theorem 6 shows that for optimal sampling problems, certain PH-distributions for the change time result in larger total optimal cost compared to other distributions.

iv) Theorem 7 derives explicit sensitivity bounds on the total cost for sampling with a mismatched model and mismatched strategy. By elementary use of the Pinsker inequality [10], Theorem 7 shows that the sensitivity is a linear function of the Kullback–Leibler divergence between the two models. Also, the bounds are tight in the sense that if the difference between the two models goes to zero, so does the performance degradation.

v) To prove the results in this paper, several stochastic dominance properties of the Bayesian filter are presented in Theorem 9. How does the posterior distribution computed by the Bayesian filter vary with observation, prior, transition matrix and observation matrix? Is it possible to order these so that the posterior distribution increases with respect to this ordering? The theorem gives sufficient conditions for the Bayesian filtering recursion to preserve the MLR stochastic order, and for the normalization measure to be submodular. It also shows that if starting with two different transition matrices but identical priors, then the optimal predictor with the larger transition matrix [in terms of the order introduced in (36)] MLR dominates the predictor with the smaller transition matrix.

vi) Finally, we consider the problem of joint optimal sampling with measurement control. The problem is motivated by the question: Should the decision maker sample less frequently but more accurately or more frequently but less accurately to achieve quickest detection/estimation? Theorem 10 gives sufficient conditions for the optimal sampling/measurement control strategy to be lower bounded by a monotone policy. It says that given the current posterior (belief), if the instantaneous expected cost of sampling less frequently but more accurately is smaller, then it is optimal to do this.

Related Works: Several examples in active/smart/cognitive sensing [17] use measurement-sampling control. The recent paper [4] considers a measurement control problem for geometric-distributed change times (two-state Markov chain with an absorbing state), where at each time the decision is made whether to take a measurement or not.[2] A constrained version of the problem in [4] can be formulated in terms of our optimal sampling problem. We discuss this in Section VI-C.

We also refer to the seminal work of Moustakides (see [35] and references therein) in event triggered sampling. Quickest detection has been studied widely (see [28] and [32] and references therein). The recent preprint [12] deals with a non-Bayesian formulation of quickest detection with sampling (and quantization) constraints. Dayanik and Goulding [11] give a unified formulation of several sequential detection problems including those with a fixed (deterministic) number of jumps as a stochastic dynamic programming problem. We have considered recently a POMDP approach to quickest detection with social learning [19] and nonlinear penalties [18] and phase-distributed change times. However, in these papers, there is only one continue and one stop action. The results in this paper are considerably more general due to the propagation of different dynamics for the multiple continue actions, general observation noise distributions (Gaussians, exponentials, discrete memoryless channels) and the possibility of a nonabsorbing target state. This is a useful feature of the lattice programming approach [1], [23], [29] used in this paper.

Finally, a brief comment on quickest state estimation where (unlike quickest detection) the target state is not absorbing and the Markov chain jumps in and out of this state. Such problems arise in cued sensing applications in body area networks [8], [27] and also active/adaptive radar systems [6]. Suppose the kinematic coordinates of a moving target are classified into a finite number of states, e.g., "in range" or "out of range." A sensor obtains noisy measurements of this kinematic state. Due to constraints in its agility, the sensor can only revisit a target at specified time intervals D_u, $u \in \{1, 2, \ldots, L\}$. If the state estimate is "in range," then the sensor cues a high resolution sensor (radar) to examine the target more carefully. The high resolution sensor is a resource that needs to be allocated amongst several targets [17], [21]. If the sensor falsely cues the high resolution sensor when the target is out of range, it pays a penalty (analogous to the false alarm) and the procedure terminates. The longer the sensor waits to cue the high resolution sensor, the more the delay penalty accumulated (due to increased uncertainty and threat levels). Given a measurement cost, how often should the sensor look at a target in order to detect if the target is in range? In radar resource management [20], this is called the revisit time problem.

[2]The author is grateful to Dr. V. Veeravalli of University of Illinois at Urbana Champaign for sharing the results in and useful discussions

II. Formulation of Optimal Sampling Problem

Let $t = 0, 1, \ldots$ denote discrete time and x_t denote a Markov chain on the finite state space

$$\{e_1, \ldots, e_X\} \tag{2}$$

where e_i is the X-dimensional unit vector with 1 in the ith position. Here, state "1" (corresponding to e_1) is labeled as the "target state." Denote

$$\mathbb{X} = \{1, 2, \ldots, X\}. \tag{3}$$

Denote $X \times X$ transition probability matrix A and the $X \times 1$ initial distribution π_0, where

$$\begin{aligned} &A = (A_{ij}, i, j \in \mathbb{X}),\ A_{ij} = P(x_{t+1} = e_j | x_t = e_i) \\ &\pi_0 = (\pi_0(i), i \in \mathbb{X}),\ \pi_0(i) = P(x_0 = e_i). \end{aligned} \tag{4}$$

A. Measurement Sampling Protocol

Let $\tau_0, \tau_1, \ldots, \tau_{k-1}$ denote previous discrete time instants at which measurement samples were taken, where by convention $\tau_0 = 0$. Let τ_k denote the current time-instant at which a measurement is taken. The measurement sampling protocol proceeds according to the following steps:

: *Step 1. Observation:* A noisy measurement $y_k \in \mathbb{Y}$ at time τ_k, $k = 1, 2, \ldots$ of the Markov chain is obtained with conditional probability distribution

$$P(y_k \leq \bar{y} | x_{\tau_k} = e_x) = \sum_{y \leq \bar{y}} f_{xy}, \quad x \in \mathbb{X}. \tag{5}$$

Here, $\sum_y$ denotes integration with respect to the Lebesgue measure (in which case, $\mathbb{Y} \subset \mathbb{R}$ and f_{xy} is the conditional probability density function) or counting measure (in which case, $\mathbb{Y}$ is a subset of the integers and f_{xy} is the conditional probability mass function $f_{xy} = P(y_k = y | x_{\tau_k} = e_x)$).

: *Step 2. Sequential Decision Making:* Let $\mathcal{I}_k = (y_1, \ldots, y_k, u_0, u_1, \ldots, u_{k-1})$ denote the history of past decisions and available observations. At times τ_k, a decision u_k is taken where

$$u_k = \mu(\mathcal{I}_k) \in \mathcal{U} = \{0\ (\text{announce change}), 1, 2, \ldots, L\} \tag{6}$$

and $u_k = l$ denotes: obtain next measurement after D_l time points, $l \in \{1, 2, \ldots, L\}$. The initial decision at time $\tau_0 = 0$ is $u_0 = \mu(\pi_0)$, where π_0 is the initial distribution specified in (4). In (6), the strategy μ belongs to the class of stationary decision strategies denoted $\boldsymbol{\mu}$. It is well known [7] that for stochastic shortest path problems that terminate with probability one in finite time,[3] strategies can be restricted to nonrandomized stationary policies. Also, $D_1 < D_2 < \cdots < D_L$ are L distinct positive integers that

[3]Such problems are equivalent to discounted infinite horizon problems [5].

denote the set of possible sampling time intervals. Thus, the decision u_k specifies the next time τ_{k+1} to make a measurement as follows:

$$\tau_{k+1} = \tau_k + D_{u_k}, \quad u_k \in \{1, 2, \dots, L\}, \quad \tau_0 = 0. \tag{7}$$

: *Step 3. Costs:* If decision $u_k \in \{1, 2, \dots, L\}$ is chosen, a *decision cost* $c(x_t, u_k)$ is incurred by the decision-maker at each time $t \in [\tau_k, \dots, \tau_{k+1} - 1]$ until the next measurement is taken at time τ_{k+1}. Also at each time τ_k, $k = 0, 1, \dots, k^* - 1$, the decision maker pays a nonnegative measurement (sampling) cost $\bar{m}(x_{\tau_k}, x_{\tau_{k+1}}, y_{k+1}, u_k)$ to observe the noisy Markov chain at time $\tau_{k+1} = \tau_k + D_{u_k}$. In terms of the policy $\mu(\mathcal{I}_k)$, this is equivalent to choosing the measurement cost as (see, e.g., [9, p. 31])

$$m(x_{\tau_k} = e_i, u_k) = \sum_j A^{D_{u_k}}|_{ij} f_{jy}\, \bar{m}(x_{\tau_k} = e_i, x_{\tau_{k+1}} = e_j, y_{k+1} = y, u_k) \tag{8}$$

where $A^{D_u}|_{ij}$ denotes the (i, j) element of matrix A^{D_u}.

: *Step 4:* If at time $t = \tau_{k^*}$ the decision $u_{k^*} = 0$ is chosen, then a terminal cost $c(x_{\tau_{k^*}}, 0)$ is incurred and the problem terminates. Recall that no sampling cost is paid at stopping time τ_{k^*} since the cost of looking at the Markov chain at time τ_{k^*} is paid at time τ_{k^*-1}.

If decision $u_k \in \{1, 2, \dots, L\}$, set k to $k + 1$ and go to Step 1. ■

Belief State Formulation: It is convenient to re-express Step 2 of the above protocol in terms of the belief state. Since μ is nonrandomized, (6) is equivalent to $u_k = \mu(\mathcal{F}_k)$, where $\mathcal{F}_k$ is the σ-algebra generated by $(y_1, \dots, y_k)$. It is well known from elementary stochastic control [23] that the belief state (posterior) constitutes a sufficient statistic for $\mathcal{I}_k$. So it suffices to choose $u_k = \mu(\pi_k)$ where $\pi_k = \mathbb{E}\{x_{\tau_k} | \mathcal{F}_k\}$ denotes the belief state. Since the state space (2) comprises of unit indicator vectors, conditional probabilities and conditional expectations coincide. So

$$\pi_k = (\pi_k(i),\, i \in \mathbb{X}) \text{ where } \pi_k(i) = P(x_{\tau_k} = e_i | \mathcal{I}_k), \tag{9}$$

initialized by π_0. It is easily proved that the belief state is updated via the Bayesian (Hidden Markov Model) filter

$$\begin{aligned} \pi_k &= T(\pi_{k-1}, y_k, u_{k-1}), \text{ where } T(\pi, y, u) = \frac{f_y (A')^{D_u} \pi}{\sigma(\pi, y, u)}, \\ \sigma(\pi, y, u) &= \mathbf{1}'_X f_y (A')^{D_u} \pi \\ f_y &= \operatorname{diag}(f_{1y}, f_{2y}, \dots, f_{Xy}), \text{ where } f_{xy} \text{ is defined in (5).} \end{aligned} \tag{10}$$

Here, $\sigma(\pi, y, u)$ is the normalization measure of the Bayesian update with $\sum_y \sigma(\pi, y, u) = 1$. Also, A' denotes transpose of matrix A and $\mathbf{1}_X$ denotes the X-dimensional vector of ones. Note that π in (10) is an X-dimensional probability vector. It belongs to the $(X - 1)$-dimensional unit-simplex denoted as

$$\Pi(X) \stackrel{\triangle}{=} \left\{\pi \in \mathbb{R}^X : \mathbf{1}'_X \pi = 1, \quad 0 \le \pi(i) \le 1 \text{ for all } i \in \mathbb{X}\right\}. \tag{11}$$

For example, $\Pi(2)$ is a 1-D simplex (unit line segment), $\Pi(3)$ is a 2-D simplex (equilateral triangle); $\Pi(4)$ is a tetrahedron, etc. Note that the unit vector states $e_1, e_2, \dots, e_X$ defined in (2) of the Markov chain x are the vertices of $\Pi(X)$.

Step 2 in the above protocol expressed in terms of the belief state reads: At decision time τ_k

1) *Step 2(a)*. Update belief state π_k according to Bayesian filter (10)
2) *Step 2(b)*. Make decision $u_k \in \mathcal{U}$ using stationary strategy μ as (see (6))

$$u_k = \mu(\pi_k) \in \mathcal{U} = \{0 \text{ (announce change)}, 1, 2, \dots, L\}. \tag{12}$$

B. Sequential Decision-Maker's Objective and Stochastic Dynamic Programming

Given the above protocol with measurement-sampling strategy μ in (12), we now define the objective of the sequential decision maker. Let $(\Omega, \mathcal{F})$ be the underlying measurable space, where $\Omega = (\mathbb{X} \times \mathcal{U} \times \mathbb{Y})^\infty$ 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 strategy $\mu \in \boldsymbol{\mu}$, there exists a (unique) probability measure $\mathbb{P}^\mu_{\pi_0}$ on $(\Omega, \mathcal{F})$; see [13] for details. Let $\mathbb{E}^\mu_{\pi_0}$ denote the expectation with respect to the measure $\mathbb{P}^\mu_{\pi_0}$.

Define the $\{\mathcal{F}_k, k \ge 1\}$ measurable stopping time k^* as

$$k^* = \{\inf k : u_k = 0 \text{ (announce target state and stop) }\}. \tag{13}$$

That is, k^* is the epoch and τ_{k^*} is the time at which the decision maker declares the target state has been reached and the problem terminates. For initial distribution $\pi_0 \in \Pi(X)$, and strategy μ, the decision maker's global objective function is

$$J_\mu(\pi_0) = \mathbb{E}^\mu_{\pi_0} \left\{ \sum_{k=0}^{k^*-1} \left[m(x_{\tau_k}, u_k) + \sum_{t=\tau_k}^{\tau_{k+1}-1} c(x_t, u_k) \right] + c(x_{\tau_{k^*}}, u_{k^*}) \right\}. \tag{14}$$

Recall that the decision cost $c(x, u)$ and measurement sampling cost $m(x, u)$ are defined in Step 3 of the protocol. Using the smoothing property of conditional expectations, (14) can be expressed in terms of the belief state π as

$$J_\mu(\pi_0) = \mathbb{E}^\mu_{\pi_0} \left\{ \sum_{k=0}^{k^*-1} C(\pi_k, u_k) + C(\pi_{k^*}, u_{k^*} = 0) \right\} \tag{15}$$

where $C(\pi, u) = C'_u \pi$ for $u \in \mathcal{U}$

$$\begin{aligned} C_u &= \begin{cases} m_u + (I + A + \cdots + A^{D_u - 1}) c_u & u \in \{1, 2, \dots, L\} \\ c_0 & u = 0 \end{cases} \\ c_u &= [\, c(e_1, u), \dots, c(e_X, u)\,]', \\ m_u &= [\, m(e_1, u), \dots, m(e_X, u)\,]'. \end{aligned} \tag{16}$$

The decision-maker aims to determine the optimal strategy $\mu^* \in \boldsymbol{\mu}$ to minimize (16), i.e.,

$$J_{\mu^*}(\pi_0) = \inf_{\mu \in \boldsymbol{\mu}} J_\mu(\pi_0). \tag{17}$$

The existence of an optimal stationary strategy μ^* follows from [5, Prop.1.3, Ch. 3].

Considering the global objective (16), the optimal stationary strategy $\mu^* : \Pi(X) \to \mathcal{U}$ and associated optimal objective $J_{\mu^*}(\pi)$ are the solution of the following Bellman's stochastic dynamic programming equation

$$\mu^*(\pi) = \arg\min_{u\in\mathcal{U}} Q(\pi,u),\ J_{\mu^*}(\pi) = V(\pi) = \min_{u\in\mathcal{U}} Q(\pi,u),$$
$$Q(\pi,u) = C(\pi,u) + \sum_{y\in\mathbb{Y}} V(T(\pi,y,u))\sigma(\pi,y,u),\ u=1,\ldots,L,$$
$$Q(\pi,0) = C(\pi,0). \tag{18}$$

Recall $T(\pi,y,u)$ and $\sigma(\pi,y,u)$ were defined in (10). The above formulation is a POMDP, where the observation space $\mathbb{Y}$ can be discrete or continuous (see (5)).

Define the set of belief states where it is optimal to apply action $u = 0$ as

$$\begin{aligned}\mathcal{S} &= \{\pi \in \Pi(X) : \mu^*(\pi) = 0\} \\ &= \{\pi \in \Pi(X) : Q(\pi,0) \le Q(\pi,u),\ u \in \{1,2,\ldots,L\}\}.\end{aligned} \tag{19}$$

$\mathcal{S}$ is called the *stopping set* since it is the set of belief states to "declare target state and stop."

Since the belief state space $\Pi(X)$ is an uncountable set, Bellman's equation (18) does not translate directly into numerical algorithms. We will exploit the structure of Bellman's equation to prove various structural results about the optimal strategy μ^* using lattice programming.

C. *Quickest Change Detection With Optimal Sampling*

We now formulate the quickest detection problem with optimal sampling—this serves as an example to illustrate the above general model. Recall that decisions (whether to stop, or continue and take next observation sample after D_l time points) are made at times $\tau_1, \tau_2, \ldots$. In contrast, the state of the Markov chain (which models the change we want to detect) can change at any time t. We need to construct the delay and false alarm penalties to take this into account.

1. *Phase-Distributed (PH) Change time:* In quickest detection, the target state (labeled as state 1 by convention) is absorbing. States $2,\ldots,X$ (corresponding to unit vectors $e_2,\ldots,e_X$) are fictitious states that the Markov chain x_t resides in before jumping into the absorbing state. So the transition matrix (4) is

$$A = \begin{bmatrix} 1 & 0 \\ \underline{A}_{(X-1)\times 1} & \bar{A}_{(X-1)\times(X-1)} \end{bmatrix}. \tag{20}$$

The "change time" Γ denotes the time at which x_t enters the absorbing state 1, i.e.,

$$\Gamma = \inf\{t > 0 : x_t = e_1\}. \tag{21}$$

Of course, if $X = 2$, then the change time Γ in (21) is geometrically distributed.

For a multistate Markov chain, to ensure that Γ is finite, assume states $2,3,\ldots X$ are transient. This is equivalent to $\bar{A}$ in (20) satisfying $\sum_{n=1}^{\infty} \bar{A}^n|_{ii} < \infty$ for $i = 1,\ldots,X-1$ (where $\bar{A}^n|_{ii}$ denotes the (i,i) element of the nth power of matrix $\bar{A}$). With the transition probabilities (20), the distribution of the change time Γ is given by the PH-distribution

$$P(\Gamma = 0) = \pi_0(1), \quad P(\Gamma = t) = \bar{\pi}_0' \bar{A}^{t-1}\underline{A}, \quad t \ge 1 \tag{22}$$

where $\bar{\pi}_0 = [\pi_0(2),\ldots,\pi_0(X)]'$. By choosing (π_0, A) and state space dimension X, one can approximate any given change-time distribution on $[0,\infty)$ by PH-distribution (22); see [25, pp. 240–243]. Indeed, PH-distributions form a dense subset for the set of all distributions.

2. *Observations*: Since states $2,3,\ldots,X$ are fictitious states that shape the PH-distributed change time (22), they are indistinguishable in terms of the observation y. That is,

$$f_{2y} = f_{3y} = \cdots = f_{Xy} \quad \text{for all } y \in \mathbb{Y}. \tag{23}$$

3. *Costs*: Associated with the quickest detection problem are the following costs.
   i) *False Alarm:* Let τ_{k^*} denote the time at which decision $u_{k^*} = 0$ (stop and announce target state) is chosen, so that the problem terminates. If the decision to stop is made before the Markov chain reaches the target state 1, i.e., $\tau_{k^*} < \Gamma$, then a unit false alarm penalty is paid. So the false alarm penalty at epoch $k = k^*$ is $\sum_{i\neq 1} I(x_{\tau_k} = e_i, u_k = 0)$. The expected false alarm penalty based on the accumulated history at epoch $k = k^*$ is

$$\sum_{i\neq 1} \mathbb{E}\{I(x_{\tau_k} = e_i, u_k = 0)|\mathcal{F}_k\} = (\mathbf{1}_X - e_1)'\pi_k I(u_k = 0). \tag{24}$$

   Recall $\mathbf{1}_X$ denotes the X-dimensional vector of ones.
   ii) *Delay cost of continuing:* Suppose decision $u_k \in \{1,2,\ldots,L\}$ is taken at time τ_k. So the next sampling time is $\tau_{k+1} = \tau_k + D_{u_k}$. Then, for any time $t \in [\tau_k, \tau_{k+1}-1]$, the event $\{x_t = e_1, u_k\}$ signifies that a change has occurred but not been announced by the decision maker. Since the decision maker can make the next decision (to stop or continue) at τ_{k+1}, the delay cost incurred in the time interval $[\tau_k, \tau_{k+1}-1]$ is $d\sum_{t=\tau_k}^{\tau_{k+1}-1} I(x_t = e_1, u_k)$, where d is a nonnegative constant. For $u_k \in \{1,2,\ldots,L\}$, the expected delay cost in interval $[\tau_k, \tau_{k+1}-1] = [\tau_k, \tau_k + D_{u_k} - 1]$ is

$$d \sum_{t=\tau_k}^{\tau_{k+1}-1} \mathbb{E}\{I(x_t = e_1, u_k)|\mathcal{F}_k\} = d e_1'(I + A + \cdots + A^{D_{u_k}-1})'\pi_k. \tag{25}$$

   iii) *Measurement Sampling Cost:* Suppose decision $u_k \in \{1,2,\ldots,L\}$ is taken at time τ_k. As in (16), let $m_{u_k} = (m(x_{\tau_k} = e_i, u_k), i \in \mathbb{X})$ denote the nonnegative measurement cost vector for choosing to take a measurement. Next, since in quickest detection, states $2,\ldots,X$ are fictitious states that are indistinguishable in terms of cost, choose $m(e_2,u) = \ldots = m(e_X,u)$.

Choosing a constant measurement cost at each time (i.e., $m(e_i,u)$ independent of state i and action u), still results in nontrivial global costs for the decision maker. This is because choosing a smaller sampling interval will result in more measurements until the final decision to stop, thereby incurring a higher total measurement cost for the global decision maker.

Remarks:

i) *Quickest state estimation:* The setup is identical to above, except that unlike (20), the transition matrix A no longer has an absorbing target state. Therefore, the Markov chain can jump in and out of the target state. To avoid pathological cases, we assume A is irreducible. Also there is no requirement for the observation probabilities to satisfy (23).

ii) *Summary*: In the notation of (16), the costs for quickest detection/estimation optimal sampling are $C(\pi, u) = C_u'\pi$ where $C_0 = c_0 = \mathbf{1}_X - e_1$ and

$$C_u = m_u + (I + A + \cdots + A^{D_u - 1})c_u, \quad c_u = de_1, \quad (26)$$

for $u \in \{1, 2, \ldots, L\}$.

iii) *Kolmogorov–Shiryayev criterion*: For constant measurement cost $m(e_i, u) = m$, the quickest detection optimal sampling objective (16) with costs (26) can be expressed as (1) where the PH-distributed change time Γ and stopping time τ_{k^*} are defined in (21), (13). For the special case $\mathcal{U} = \{0 \text{ (stop)}, D_1 = 1\}$, measurement cost $m_u = 0$, geometrically distributed Γ (so $X = 2$), then (1) becomes the Kolmogorov–Shiryayev criterion for detection of disorder [30].

III. Structural Results for Optimal Sampling Policy $\mu^*(\pi)$ for Two-State Case

This section analyzes the structure of the optimal sampling strategy $\mu^*(\pi)$ [solution of Bellman's equation (18)] for two-state Markov chains ($X = 2$). Recall that two-state Markov chains model geometric distributed change times in quickest detection problems.

We list the following assumptions that will be used subsequently.

(A0) The target state e_1 belongs to the stopping set $\mathcal{S}$ defined in (19). Either state 1 is absorbing and all other states are transient, or A is irreducible.

(A1) The costs $C(e_i, u)$ in (16) are increasing with $i \in \mathbb{X}$ for each $u \in \mathcal{U}$.

($\overline{\text{A1}}$) The costs $C(e_i, u)$ in (16) are decreasing with $i \in \mathbb{X}$ for each $u \in \mathcal{U}$.

(A2) The transition matrix A is totally positive of order 2 (TP2), that is, all second-order minors are nonnegative.[4]

(A3) The observation distribution f is TP2, that is, $f_{i+1,y}/f_{iy}$ is increasing in y for each state i.

(A4) $C(e_i, u)$ is submodular in (i, u), where $i \in \mathbb{X}$ and $u \in \{1, 2, \ldots, L\}$, i.e., $C(e_i, u+1) - C(e_i, u)$ is decreasing[5] in $i \in \mathbb{X}$.

[4]The definitions of TP2 in (A2) and (A3) are equivalent. Both are equivalent to each row of A and f being MLR dominated by subsequent rows; see [15].

[5]Throughout this paper, we use the term "decreasing" in the weak sense. That is "decreasing" means nonincreasing. Similarly, the term "increasing" means nondecreasing.

(A0) ensures that the stopping problem is well posed. It says that if it was known with certainty that the target state e_1 has been reached, then it is optimal to stop. For quickest time detection, it holds trivially since $C(e_1, 0) \leq C(\pi, u)$ for $u \in \{1, \ldots, L\}, \pi \in \Pi(X)$. The second part of (A0) ensures that Γ is finite and state 1 is recurrent. (A0) is assumed throughout the paper and not will be repeated subsequently. The remaining assumptions are discussed below in Section III-C.

A. Optimality of Threshold Policy for Quickest Detection With Sampling

Consider quickest detection with optimal sampling for geometric distributed change time. From (20), the transition matrix is $A = \begin{bmatrix} 1 & 0 \\ 1 - A_{22} & A_{22} \end{bmatrix}$ and expected change time is $\mathbb{E}\{\Gamma\} = \frac{1}{1 - A_{22}}$, where Γ is defined in (21). For a two-state Markov chain since $\pi(1) + \pi(2) = 1$, it suffices to represent π by its first element $\pi(1) \in [0, 1]$. That is, the belief space $\Pi(X)$ is the interval $[0, 1]$.

Theorem 1: Consider the quickest detection optimal sampling problem of Section II-C with geometric-distributed change time and costs (26). Assume the measurement cost $m(e_i, u)$ satisfies (A1), (A4) and the observation distribution satisfies (A3). Then, there exists an optimal strategy $\mu^*(\pi)$ with the following monotone structure: There exist up to L thresholds denoted $\pi_1^*, \ldots, \pi_L^*$ with $0 \leq \pi_L^* \leq \pi_{L-1}^* \leq \cdots \leq \pi_1^* \leq 1$ such that, for $\pi(1) \in [0, 1]$,

$$\mu^*(\pi) = \begin{cases} L, & \text{if } 0 \leq \pi(1) \leq \pi_L^* \\ L - 1, & \text{if } \pi_L^* < \pi(1) \leq \pi_{L-1}^* \\ \vdots & \vdots \\ 1, & \text{if } \pi_2^* < \pi(1) \leq \pi_1^* \\ 0 \text{ (announce change)}, & \text{if } \pi_1^* < \pi(1) \leq 1. \end{cases} \quad (27)$$

Here, the sampling intervals are ordered as $D_1 < D_2 < \ldots < D_L$. So the optimal sampling strategy (27) makes measurements less frequently when the posterior $\pi(1)$ is away from the target state and more frequently when closer to the target state. (Recall the target state is $\pi(1) = 1$.)

The proof of Theorem 1 is given in Section III-C. Theorem 1 is a special case of a more general result, Theorem 2, that we will present below (where state 1 is not necessarily absorbing).

There are two main conclusions regarding Theorem 1. First, for constant measurement cost, (A1) and (A4) hold trivially. Then, Theorem 1 only requires (A3) which holds for several classes of discrete and continuous observation distributions as discussed in Section III-C. For the general measurement cost $\bar{m}(x_{\tau_{k+1}} = e_j, y_{k+1}, u_k)$ [see (8)] that depends on the state at epoch $k + 1$, $m(e_i, u)$ in (8) automatically satisfies (A4) if A satisfies (A2) and $\bar{m}$ is decreasing in j.

Second, the optimal strategy $\mu^*(\pi)$ is monotone in posterior $\pi(1)$ and, therefore, has a finite dimensional characterization. To determine the optimal strategy, one only needs to determine (estimate) the values of the L thresholds $\pi_1^*, \ldots, \pi_L^*$. These can be estimated via a simulation-based stochastic optimization algorithm. We will give bounds for these threshold values in Section IV.

B. Optimality of Threshold Strategy for Sequential Optimal Sampling

In this section, we consider the general optimal sampling problem where the two-state Markov chain can jump in and out of the target state 1. (Recall quickest detection is a special case where the target state is absorbing). We will give sufficient conditions for the optimal strategy to be monotone on a subset denoted $\Pi_{\mathcal{M}}$ of the belief space $\Pi(X) = [0, 1]$. Define

$$\Pi_{\mathcal{M}} = \{\pi(1) : 0 \le \pi(1) \le \pi_{\mathcal{M}}\},$$
$$\text{where } \pi_{\mathcal{M}} = \frac{A_{22} - A^2|_{22}}{A_{22} - A^2|_{22} + A_{11} - A^2|_{11}}. \quad (28)$$

For transition matrices that satisfy (A2), it will be shown in Lemma 1, Section IV-A, that $\pi_{\mathcal{M}} \in [0, 1]$ and so $\Pi_{\mathcal{M}}$ is nonempty. For the quickest detection problem since $A_{11} = A^2|_{11} = 1$, clearly $\pi_{\mathcal{M}} = 1$, and so $\Pi_{\mathcal{M}} = \Pi(X)$ meaning that the optimal strategy is monotone on the entire belief space $[0, 1]$. (This is why Theorem 1 holds on $\Pi(X)$.)

The intuition for specifying $\Pi_{\mathcal{M}}$ is that for $\pi \in \Pi_{\mathcal{M}}$, the filtering update (10) satisfies the following property under (A2): the first element $\pi(1)$ of $T(\pi, y, u)$ always is smaller than that of $T(\pi, y, u+1)$ (equivalently, $T(\pi, y, u)$ first-order stochastically dominates $T(\pi, y, u+1)$). This property is crucial to prove that the optimal strategy has a monotone structure. Section V-C presents this and several other important properties of the Bayesian filtering update.

The following is the main result of this section (and includes Theorem 1 as a special case).

Theorem 2: Consider the optimal sampling problem of Section II with state dimension $X = 2$ and action space $\mathcal{U}$ in (6). Then, the optimal strategy $\mu^*(\pi)$ in (18) has the following structure:

i) The optimal stopping set $\mathcal{S}$ (19) is a convex subset of $\Pi(X)$. Therefore, the stopping set is the interval $\mathcal{S} = (\pi_1^*, 1]$ where the threshold $\pi_1^* \in [0, 1]$.

ii) Under (A1)–(A4), there exists an optimal sampling strategy $\mu^*(\pi)$ defined in (18) that is decreasing in $\pi(1)$ for $\pi(1) \in \Pi_{\mathcal{M}} \cup \mathcal{S}$, where $\Pi_{\mathcal{M}}$ is defined in (28).

iii) As a consequence of (i) and (ii), there exist up to L thresholds denoted $\pi_1^*, \ldots, \pi_L^*$ in $\Pi_{\mathcal{M}} \cup \mathcal{S}$ with $0 \le \pi_L^* \le \pi_{L-1}^* \le \cdots \le \pi_1^*$ such that, for $\pi(1) \in \Pi_{\mathcal{M}} \cup \mathcal{S}$, the optimal strategy has the monotone structure of (27).

Corollary 1: Consider the quickest state estimation problem with setup identical to the quickest detection problem of Theorem 1 except that the transition matrix A does not necessarily have an absorbing state. Assume A satisfies (A2). Then, the conclusions of Theorem 1 hold on $\Pi_{\mathcal{M}} \cup \mathcal{S}$ where $\Pi_{\mathcal{M}}$ is defined in (28). ■

The proof of Theorem 2 is in Appendix B. The proof of Corollary 1 is in Section III-C.

Fig. 2 illustrates and compares what Theorem 1 and Theorem 2 say for $L = 2$. As shown in Fig. 2(b), for quickest detection, $\Pi_{\mathcal{M}} = [0, 1]$ and so the optimal strategy $\mu^*(\pi)$ is decreasing in belief $\pi(1) \in [0, 1]$ (see Theorem 1). For more general optimal sampling problems, Fig. 2(a) illustrates that $\mu^*(\pi)$ is monotone decreasing in $\pi(1) \in [0, \pi_{\mathcal{M}}] \cup \mathcal{S}$ (see Theorem 2), where $\pi_{\mathcal{M}}$ is defined in (28).

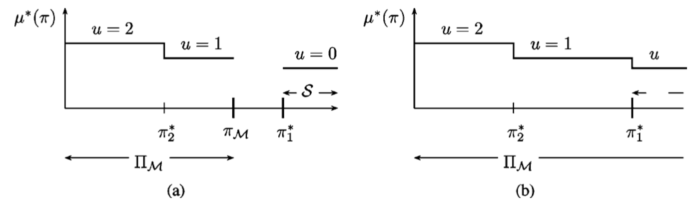


Fig. 2. Illustration of Theorems 1 and 2. (a) illustrates Theorem 2 which states that the optimal strategy $\mu^*(\pi)$ is monotone decreasing in $\pi(1)$ for $\pi(1) \in \Pi_{\mathcal{M}} \cup \mathcal{S}$. It illustrates the case when $\mathcal{S}$ and $\Pi_{\mathcal{M}}$ are disjoint sets, i.e., $\pi_{\mathcal{M}} < \pi_1^*$. In this case, for $\pi(1)$ in the interval $[\pi_{\mathcal{M}}, \pi_1^*]$, the theorem cannot say that $\mu^*(\pi)$ is monotone. (b) illustrates Theorem 1 for the quickest detection problem with optimal sampling. In this case, $\pi_{\mathcal{M}} = 1$, i.e., $\mathcal{S} \subset \Pi_{\mathcal{M}}$. So the optimal strategy is monotone for all $\pi(1) \in [0, 1]$.

A short word on the proof of Theorem 2 is presented in Appendix B. It involves analyzing the structure of Bellman's equation (18). It will be shown that $Q(\pi, u)$ in (18) is a submodular function (defined in Appendix B) on the partially ordered set $[\Pi_{\mathcal{M}}, \ge_r]$ which constitutes a lattice. Here $\ge_r$ denotes the MLR stochastic order defined in Section V-C. For $X = 2$, $\Pi(X)$ is the unit interval $[0, 1]$, and in this case, $[\Pi_{\mathcal{M}}, \ge_r]$ is a chain (totally ordered set) and $\ge_r$ is equivalent to first-order stochastic dominance (denoted as $\ge_s$). For $X \ge 2$ considered in Section IV, a similar idea is used to bound the optimal strategy on $[\Pi(X), \ge_r]$.

C. Discussion of Assumptions A1–A4

As mentioned previously, Theorem 1 for quickest detection and Corollary 1 for quickest estimation are special cases of Theorem 2. To illustrate the assumptions of Theorem 2, we now prove Theorem 1 and Corollary 1 by showing that assumptions that (A1)–(A4) hold. Recall from (20) that for quickest detection with geometric change time, the transition matrix is

$$A = \begin{bmatrix} 1 & 0 \\ 1 - A_{22} & A_{22} \end{bmatrix}. \text{ So } A^{D_u} = \begin{bmatrix} 1 & 0 \\ 1 - A_{22}^{D_u} & A_{22}^{D_u} \end{bmatrix}. \quad (29)$$

1) Assumption (A1): This requires that the elements of the all cost vectors $C(e_i, u)$ are increasing in i for all u. However, quickest detection/estimation is complicated by the fact that (A1) does not hold since from (26), $C(e_i, 0)$ is increasing in i while $C(e_i, u)$, $u \ge 1$ are decreasing in i if the measurement cost m is a constant. The trick is to transform these costs so that they are either all increasing without altering the optimal strategy. Below, we transform the costs so that they are all increasing [satisfy (A1)], satisfy (A4) and yet keep the optimal strategy unchanged.

Theorem 3: For any $\phi \in \mathbb{R}^X$, define the transformed costs $\underline{C}(\pi, u)$ as follows:

$$\underline{C}(\pi, 0) = C(\pi, 0) + \phi'\pi \quad (30)$$
$$\underline{C}(\pi, u) = C(\pi, u) + \phi'\pi - \phi' {A'}^{D_u} \pi, \quad u \in \{1, 2, \ldots, L\}.$$

i) Bellman's equation (18) applied to optimize the global objective (16) with transformed costs $\underline{C}(\pi, u)$ yields the same optimal strategy $\mu^*(\pi)$ as the global objective with original costs $C(\pi, u)$. Also the value function is $\underline{V}(\pi) = V(\pi) + \phi'\pi$.

ii) Consider the quickest detection/estimation problem with costs defined in (26) and $X = 2$. Assume the sampling cost $m(e_i, u)$ satisfies (A1) and (A4). Then, choosing $\phi = -\alpha C(\pi, L)$, where $\alpha = 1/(1 - A^{D_L}|_{11} + A^{D_L}|_{21})$, implies $\underline{C}(e_i, u)$ in (30) satisfies (A1) and (A4). ■

Theorem 3 is proved in Appendix C. As a consequence of Statement (ii), it follows that Theorem 2 holds and the optimal strategy for the transformed costs is monotone in the posterior distribution. Since the strategy $\mu^*(\pi)$ is unchanged by the transformation, Theorem 2 holds for the original costs $C(\pi, u)$, thereby proving Theorem 1.

2) Assumption (A2): From the structure of transition matrix A in (29), clearly (A2) holds automatically for the quickest detection problem. For numerous examples of TP2 transition matrices, see [15]. Also, A does not need to have an absorbing state for Theorem 2 to hold.

The reader familiar with monotone policies for Markov decision processes might question why we do not require a submodular assumption on the transition probabilities for Theorem 2 to hold. The reason is that in the sampling control formulation, A satisfying (A2) implies that for each $q \in \mathbb{X}$, $\sum_{j \geq q} A^{D_u}|_{ij}$ is submodular in (i, u). That is, (A2) implies that $\sum_{j \geq q} A^{D_{u+1}}|_{ij} - A^{D_u}|_{ij}$ is decreasing in $i \in \mathbb{X}$ for $u = 1, 2, \ldots, L-1$. This is proved in Theorem 9(3).

3) Assumption (A3): Numerous continuous and discrete noise distributions satisfy the TP2 property; see [15]. Examples include Gaussians, Exponential, Binomial, Poisson, etc. Examples of discrete observation distribution satisfying (A3) include binary erasure channels—see Section VII. A binary symmetric channel with error probability less than 0.5 also satisfies (A3).

4) Assumption (A4): In general, Theorem 2 requires the costs $C(e_i, u)$ to be submodular. However, for the special case of quickest detection with optimal sampling, Theorem 3 only needs the measurement cost $m(e_i, u)$ to be submodular, i.e., $m(e_i, u+1) - m(e_i, u)$ is decreasing in i. This holds trivially if the measurement cost is independent of the state or action.

IV. Monotone Bounds to Optimal Strategy for Multistate Markov Chain

We now consider optimal sampling for multistate Markov chains ($X \geq 2$) observed in noise. Since $\Pi(X)$ is an $(X-1)$-dimensional simplex, for $X \geq 3$ substantial complications arise as the belief state vectors are only partially orderable. For $X \geq 3$, determining sufficient conditions for the optimal strategy to have a monotone structure is an open problem [23], [29]. This motivates the question: Can the optimal strategy be lower and upper bounded by monotone strategies? This section shows that the optimal strategy can be indeed be bounded by monotone strategies (with respect to the MLR order) that are myopic. Such judiciously chosen myopic strategies provide rigorous and easily computable bounds in *strategy space* to an intractable POMDP problem. They apply to quickest detection optimal sampling problems with PH-distributed change times and multistate quickest estimation. The costs associated with such myopic strategies form an upper bound to the optimal achievable cost. Finally, these monotone myopic strategies are used in Section VI-C to truncate the action space $\mathcal{U}$.

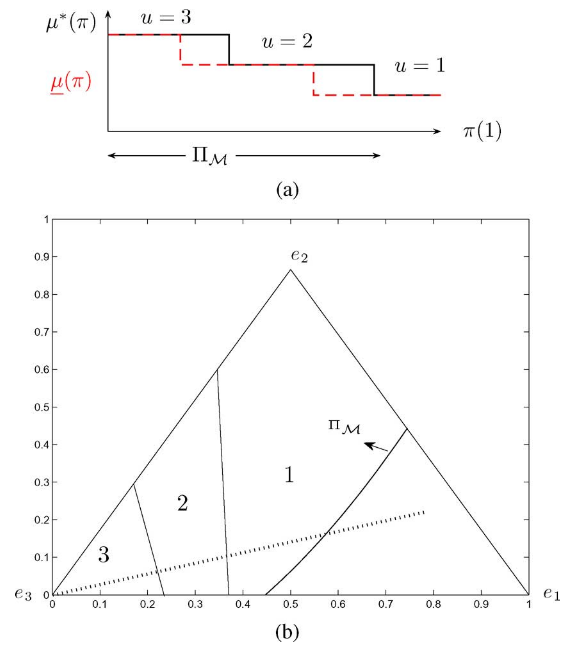


Fig. 3. (a) Illustration of myopic strategy $\underline{\mu}(\pi)$ (in dashed line) and optimal strategy $\mu^*(\pi)$ (solid line). Theorem 4 asserts that $\underline{\mu}(\pi) \leq \mu^*(\pi)$ for $\pi \in \Pi_{\mathcal{M}}$ defined in (33). (b) illustrates $\underline{\mu}(\pi)$ for $X = 3$. The region to the left of the curve is $\Pi_{\mathcal{M}}$. Note also that $\underline{\mu}(\pi)$ is increasing in π with respect to the MLR order. This implies that on any line segment terminating in e_3 (dotted line shown), $\underline{\mu}(\pi)$ increases. (a) $X = 2$. (b) $X = 3$.

Fig. 3 illustrates the main result of this section, namely a monotone strategy $\underline{\mu}(\pi)$ that lower bounds the optimal strategy $\mu^*(\pi)$ on a subset $\Pi_{\mathcal{M}}$ of $\Pi(X)$ defined in (33) below. For example, if $\mathcal{U} = \{0, 1, 2, 3\}$, then Theorem 4 below states that $\underline{\mu}(\pi) = 3$ implies $\mu^*(\pi) = 3$, $\underline{\mu}(\pi) = 2$ implies $\mu^*(\pi) \in \{2, 3\}$, etc. The myopic strategy $\underline{\mu}(\pi)$ is increasing in π with respect to the MLR order. So for $X = 2$ in Fig. 3(a), it comprises step functions decreasing in $\pi(1)$. For $X \geq 3$, the MLR order is a partial order. Any line segment terminating in e_X constitutes a chain (totally ordered subset) and $\underline{\mu}(\pi)$ is increasing on such a line [dotted line in Fig. 3(b)] toward e_3.

To compare belief states in $\Pi(X)$ for $X \geq 2$, we use the MLR stochastic order.

Definition 1 (MLR Order; see [24]): 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

$$\pi_1(i)\pi_2(j) \leq \pi_2(i)\pi_1(j), \quad i < j, i, j \in \{1, \ldots, X\}. \quad (31)$$

Similarly $\pi_1 \leq_r \pi_2$ if $\leq$ in (31) is replaced by $\geq$.

For $X > 2$, the MLR order is a partial order, indeed $[\Pi(X), \geq_r]$ forms a poset (partially ordered set). The MLR stochastic order is useful since it is closed under conditional expectations. That is, $X \geq_r Y$ implies $\mathbb{E}\{X|\mathcal{F}\} \geq_r \mathbb{E}\{Y|\mathcal{F}\}$ for any two random variables X, Y and sigma-algebra $\mathcal{F}$ [15], [24], [34]; see also [23] and [29] for extensive use in partially observed stochastic control.

A. Monotone Bounds to Optimal Sampling Policy

The main result below shows that the optimal sampling strategy $\mu^*(\pi)$ is lower bounded by a monotone strategy that is myopic. Define myopic strategy $\underline{\mu}(\pi)$ and myopic stopping set $\underline{\mathcal{S}}$ by

$$\underline{\mu}(\pi) = \arg\min_{u \in \mathcal{U}} C(\pi, u) \tag{32}$$
$$\underline{\mathcal{S}} = \{\pi \in \Pi(X) : \underline{\mu}(\pi) = 0\}$$
$$= \{\pi \in \Pi(X) : C(\pi, 0) < C(\pi, u),\ u \in \{1, 2, \ldots, L\}\}.$$

So, $\underline{\mathcal{S}}$ is the set of belief states for which the myopic strategy declares stop.

Theorem 4: Consider the sequential sampling problem of Section II with optimal strategy specified by (18). Then

1) The stopping set $\mathcal{S}$ defined in (19) is a convex subset of the belief state space $\Pi(X)$.
2) $\underline{\mathcal{S}} \subset \mathcal{S}$ where $\underline{\mathcal{S}}$ is the myopic stopping set defined in (32).
3) Under (A1), (A2), and (A3), the myopic strategy $\underline{\mu}(\pi)$ defined in (32) forms a lower bound to the optimal strategy $\mu^*(\pi)$, i.e., $\mu^*(\pi) \geq \underline{\mu}(\pi)$ for all $\pi \in \Pi_{\mathcal{M}} - \mathcal{S}$. Here
$$\Pi_{\mathcal{M}} = \{\pi : A'\pi \geq_r A^{2'}\pi\}. \tag{33}$$
Under ($\overline{\text{A1}}$) (A2), and (A3), $\underline{\mu}(\pi)$ forms an upper bound to optimal strategy $\mu^*(\pi)$, i.e., $\mu^*(\pi) \leq \underline{\mu}(\pi)$ for all $\pi \in \Pi_{\mathcal{M}} - \underline{\mathcal{S}}$.
4) If (A4) holds, then the myopic strategy is $\underline{\mu}(\pi)$ is increasing with π for $\pi \in \Pi(X)$ with respect to the MLR stochastic order.

The proof is in Appendix D.

Theorem 4 gives a lot of analytical mileage in terms of characterizing the optimal strategy. Statement (1) characterizes the convexity of the stopping set and Statement (2) gives an easily computable subset of $\mathcal{S}$. Statements 3 and 4 assert that the myopic strategy $\underline{\mu}(\pi)$ comprising of increasing step functions[6] lower bounds the optimal strategy $\mu^*(\pi)$. The myopic strategy defined in (32) is computed trivially on the simplex $\Pi(X)$.

Remark: $\Pi_{\mathcal{M}}$ in (33) is the set of belief states for which a one step ahead optimal predictor MLR dominates a two step ahead optimal predictor. Since the MLR order is transitive, if (A2) holds, it will be shown in Theorem 9 that for $\pi \in \Pi_{\mathcal{M}}$, $T(\pi, y, u) \geq_r T(\pi, y, u+1)$. That is, the belief state update for choosing a smaller sampling interval D_u MLR dominates that of choosing a larger sampling interval D_{u+1}. Note that the above definition specializes to the definition of $\Pi_{\mathcal{M}}$ given previously in (28) for $X = 2$.

The following lemma summarizes some important properties of $\Pi_{\mathcal{M}}$.

Lemma 1: The set of belief states $\Pi_{\mathcal{M}}$ defined in (33) has the following properties:

i) $\Pi_{\mathcal{M}}$ always contains the belief state e_X and so is nonempty.
ii) A sufficient condition for $\Pi_{\mathcal{M}} = \Pi(X)$ is $A_{ij} A^2|_{m,j+1} \leq A^2|_{ij} A_{m,j+1}$, $i, j+1, m \in \mathbb{X}$. For $X = 2$, the condition reads: $A_{12} \geq A^2|_{12}$ and $A_{22} \geq A^2|_{22}$.
iii) For $X = 2$, $\Pi_{\mathcal{M}}$ is an interval of the form $[0, \pi_{\mathcal{M}}]$. (This is consistent with definition (28).)

The proof is in Appendix B.

[6]For $X = 2$, MLR increasing with respect to π is equivalent to decreasing with respect to $\pi(1)$. That is why, Fig. 3(a) shows $\underline{\mu}(\pi)$ decreasing with respect to $\pi(1)$.

B. Quickest Detection/Estimation With Optimal Sampling

We now illustrate Theorem 4 by constructing bounds for the optimal sampling strategy in quickest detection/estimation (e.g., with PH-distributed change time (20)). As described in Section III-C, quickest detection/estimation is complicated by the fact that $C(e_i, 0)$ is increasing, while $C(e_i, u), u \geq 1$, is decreasing in i meaning that neither (A1) nor ($\overline{\text{A1}}$) hold. Similar to Theorem 3, the main idea is to transform the cost so that either ($\overline{\text{A1}}$) or (A1) holds. Recall from Theorem 3 that for any vector $\bar{\phi} \in \mathbb{R}^X$, the following transformed cost

$$\bar{C}(\pi, u) = C(\pi, u) + \bar{\phi}'\pi - \bar{\phi}' A'\pi, \quad \bar{C}(\pi, 0) = C(\pi, u) + \bar{\phi}'\pi \tag{34}$$

results in identical optimal strategy $\mu^*(\pi)$ to that of the original cost $C(\pi, u)$ with associated value function $\bar{V}(\pi) = V(\pi) + \bar{\phi}'\pi$. Define the myopic strategy associated with $\bar{C}(\pi, u)$ as

$$\bar{\mu}(\pi) = \arg\min_{u \in \mathcal{U}} \bar{C}(\pi, u). \tag{35}$$

The aim is to choose the vector $\bar{\phi}$ so that the transformed costs $\bar{C}(e_i, u)$ are decreasing in i [satisfy ($\overline{\text{A1}}$)] and submodular (A4). Since ($\overline{\text{A1}}$) and (A4) impose linear constraints on $\bar{\phi}$, it is straightforward to check if a feasible $\bar{\phi}$ exists and compute it using an LP solver.

Theorem 5: Consider the quickest detection/estimation optimal sampling problem for $X \geq 2$ defined in Section II-C with costs in (26). Then

1) Statements 1 and 2 of Theorem 4 apply.
2) Assume (A2), (A3) and there exists $\bar{\phi} \in \mathbb{R}^X$ such that transformed costs $\bar{C}(e_i, u)$ in (34) satisfy ($\overline{\text{A1}}$) and (A4). Then the myopic policy (32) satisfies $\mu^*(\pi) \leq \bar{\mu}(\pi)$, $\pi \in \Pi_{\mathcal{M}} - \underline{\mathcal{S}}$. Moreover, $\bar{\mu}(\pi)$ is increasing in $\pi \in \Pi(X)$ with respect to the MLR order.
3) Assume $X = 2$, (A2), (A3) hold, and the measurement cost m satisfies (A4). Then, for any $\bar{\phi} \in \mathbb{R}^2$ satisfying $\bar{\phi}_1 - \bar{\phi}_2 \geq \max\{1, \frac{m(e_2,u) - m(e_1,u)}{A_{12} + A_{21}}\}$, the myopic policy (32) satisfies $\bar{\mu}(\pi) \geq \mu^*(\pi)$ for all $\pi \in \Pi_{\mathcal{M}} - \underline{\mathcal{S}}$. Also $\bar{\mu}(\pi)$ is decreasing in $\pi(1) \in [0, 1]$. (Recall for quickest detection (A2) always holds and $\Pi_{\mathcal{M}} = \Pi(X)$.)

The proof is in Appendix E.

Remark (i): Statements (2) and (3) assert that the optimal strategy μ^* can be upper bounded by the monotone myopic strategy $\bar{\mu}$ for multistate quickest detection (PH-distributions) and quickest estimation problems. As mentioned earlier, for $X \geq 3$, if a feasible $\bar{\phi} \in \mathbb{R}^X$ exists so that ($\overline{\text{A1}}$), (A4) hold, it is easily computed via an LP solver. Actually we have the following explicit construction for $\bar{\phi}$ which guarantees ($\overline{\text{A1}}$) and (A4) hold.

Lemma 2: Assume (A2), (A3) and that the measurement cost $m(e_i, u)$ is state independent. Suppose the elements $\bar{\phi}_i$ of $\bar{\phi} \in \mathbb{R}^X$ are chosen as integer concave and decreasing in i with

$$\bar{\phi}_1 - \bar{\phi}_2 \geq \sum_{i=1}^{X} (A_{1i} - A_{Xi})\bar{\phi}_i, \quad \bar{\phi}_1 \geq \bar{\phi}_2 + 1.$$

Then, $\bar{C}(e_i, u)$, $i \in \mathbb{X}, u \in \mathcal{U}$ satisfy $(\overline{\text{A1}})$ and (A4), and therefore, Theorem 5 applies. ■

The proof is in Appendix E.

Remark (ii): Combining Statement (3) with Theorem 1 implies that $\mu^*(\pi)$ is monotone in $\pi(1)$, and the optimal threshold values $\pi_1^*, \pi_2^*, \ldots, \pi_L^*$ are upper bounded by those of the myopic strategy $\bar{\mu}(\pi)$. These upper bounds on the optimal thresholds can be used to initialize a stochastic optimization algorithm to estimate the thresholds of the optimal monotone strategy.

Remark (iii): Regarding choosing $\bar{\phi}$ for the transformed cost $\bar{C}(e_i, u)$ to satisfy (A1) rather than $(\overline{\text{A1}})$, it turns out that the resulting transformed cost always satisfies $\bar{C}(e_i, 1) \leq \bar{C}(e_i, u), u > 1$, implying that $\bar{\mu}(\pi) = 1$ for all $\pi \in \Pi_{\mathcal{M}} - \mathcal{S}$. Then, $\bar{\mu}(\pi)$ forms a trivial lower bound to $\mu^*(\pi)$ on $\Pi_{\mathcal{M}} - \mathcal{S}$ and is not useful.

V. Performance and Sensitivity of Optimal Strategy

In previous sections, we have presented structural results on monotone optimal *strategies*. In comparison, this section focuses on *achievable costs* attained by the optimal strategy. This section presents two results. First, we give bounds on the achievable performance of the optimal strategies by the decision maker. This is done by introducing a partial ordering of the transition and observation probabilities—the larger these parameters with respect to this order, the larger the optimal cost incurred. Second, we give explicit bounds on the sensitivity of the total sampling cost with respect to misspecified model and misspecified strategy—these bounds can be expressed in terms of the Kullback–Leibler divergence. Such robustness is useful since even if a model violates the assumptions of the previous section, as long as the model is sufficiently close to a model that satisfies the conditions, then the optimal strategy is close to a monotone strategy.

A. How Does Total Cost of the Optimal Sampling Strategy Depend on State Dynamics?

Consider the optimal sampling problem formulated in Section II. How does the optimal expected cost J_{μ^*} defined in (17) vary with transition matrix A and observation distribution f? Can the transition matrices and observation distributions be ordered so that the larger they are, the larger the optimal sampling cost?

Consider two distinct optimal sampling problems with transition matrices $\theta = A$ and $\bar{\theta} = \bar{A}$, respectively. Alternatively, consider two distinct optimal sampling problems with observation distributions $\theta = f$ and $\bar{\theta} = \bar{f}$. Let $C(\pi, u; \theta)$ and $C(\pi, u; \bar{\theta})$ denote the associated costs. Recall, the costs (16) depend explicitly on the transition matrix. Let $\mu^*(\theta)$ and $\mu^*(\bar{\theta})$ denote, respectively, the optimal strategies for the two different models. Let $J_{\mu^*(\theta)}(\pi; \theta) = V(\pi; \theta)$ and $J_{\mu^*(\bar{\theta})}(\pi; \bar{\theta}) = V(\pi; \bar{\theta})$ denote the optimal value functions (18) corresponding to applying the respective optimal strategies.

Define the following ordering of two arbitrary transition matrices A and $\bar{A}$:

$$A \succeq \bar{A} \text{ if } A_{ij}\bar{A}_{m,j+1} \leq \bar{A}_{ij}A_{m,j+1}, \quad i, j+1, m \in \mathbb{X}. \quad (36)$$

Introduce the following reverse Blackwell ordering [29] of observation distributions: $\bar{f}$ reverse Blackwell dominates f denoted as

$$f \succeq_B \bar{f} \text{ if } f = \bar{f}R \quad (37)$$

where $R = (R_{lm})$ is a stochastic kernel, i.e., $\sum_m R_{lm} = 1$. (Recall from (5) that $\sum_m$ denotes integration or summation.) This means that $\bar{f}$ yields more accurate measurements of the underlying state than f.

The question we pose is: How does the optimal cost $J_{\mu^*(\theta)}(\pi; \theta)$ vary with transition matrix A and observation distribution f? For example, in the quickest detection optimal sampling problem, do certain phase-type distributions for the change time result in larger total optimal cost compared to other phase-type distributions?

Theorem 6:

1) Consider two distinct optimal sampling problems with transition matrices A and $\bar{A}$, respectively, where $A \succeq \bar{A}$ with respect to ordering (36).
   i) If $C(\pi, u; A) \geq C(\pi, u; \bar{A})$ and (A1), (A2), (A3) hold, then the expected total costs incurred by the optimal sampling strategies satisfy $J_{\mu^*(A)}(\pi; A) \geq J_{\mu^*(\bar{A})}(\pi; \bar{A})$.
   ii) If $C(\pi, u; A) \leq C(\pi, u; \bar{A})$ and $(\overline{\text{A1}})$, (A2), (A3) hold, $J_{\mu^*(A)}(\pi; A) \leq J_{\mu^*(\bar{A})}(\pi; \bar{A})$.
2) Consider two distinct optimal sampling problems with observation distributions f and $\bar{f}$, respectively, where $f \succeq_B \bar{f}$ with respect to ordering (37). If $C(\pi, u; f) \geq C(\pi, u; \bar{f})$, then $J_{\mu^*(f)}(\pi; f) \geq J_{\mu^*(\bar{f})}(\pi; \bar{f})$.

The proof is in Appendix F. Computing the optimal strategies and costs of a POMDP is intractable. Yet, the above theorem facilitates comparison of these optimal costs for different transition and observation matrices.

Remark (i): As a trivial consequence of Statement 2 of the theorem, the optimal cost incurred with perfect measurements is always smaller than that with noisy measurements. Since the optimal sampling problem with perfect measurements is a full observed MDP (or equivalently, infinite signal to noise ratio), the corresponding optimal cost forms a easily computable lower bound to the achievable cost.

Remark (ii): Here are examples of transition matrices $A, \bar{A}$ that satisfy (A3) and $A \succeq \bar{A}$.

Example 1: Geometric distributed change time: $A = \begin{bmatrix} 1 & 0 \\ 1 - A_{22} & A_{22} \end{bmatrix}, \bar{A} = \begin{bmatrix} 1 & 0 \\ 1 - \bar{A}_{22} & \bar{A}_{22} \end{bmatrix}$, where $A_{22} \geq \bar{A}_{22}$.

Example 2: PH-distributed change time:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0.3 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}, \bar{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0.9 & 0.1 & 0 \\ 0.8 & 0.2 & 0 \end{bmatrix}.$$

Example 3: Markov chain without absorbing state: $A = \begin{bmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{bmatrix}, \bar{A} = \begin{bmatrix} 0.8 & 0.2 \\ 0.7 & 0.3 \end{bmatrix}$.

B. Sensitivity to Misspecified Model and Strategy

How sensitive is the total sampling cost to the choice of sampling strategy? The aim of this section is to establish the explicit bounds on the sensitivity of the total cost with respect to misspecified model and misspecified strategy.

1) Notation and Assumption: Define

$$y^*_{\theta,\bar{\theta}} = \inf\Big\{ y : (C_{\bar{u}} - C_0)' T(e_X, y, u; \theta) \le 0 \text{ and } (C_{\bar{u}} - C_0)' T(e_X, y, u; \bar{\theta}) \le 0 \; \forall u, \bar{u} \in \{1, 2, \ldots, L\} \Big\}. \quad (38)$$

The set depicted in (38) represents a subset of the observation space $\mathbb{Y}$ for which the optimal decision is to stop. We assume that

(A5) $P(y \le y^*_{\theta,\bar{\theta}}) > 0$.

Assumption (A5) is discussed after the statement of the theorem below. It holds trivially if the observation distribution f_{xy} [defined in (5)] is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, i.e., if the density has support on $\mathbb{R}$ such as Gaussian noise. (A5) is relevant for cases when the observation space is finite or a subset of $\mathbb{R}$.

2) Main Result: Consider two distinct models $\theta = (A, f)$ and $\bar{\theta} = (\bar{A}, \bar{f})$ of the optimal sampling problem. Recall $J_{\mu^*(\theta)}(\pi; \theta)$ and $J_{\mu^*(\theta)}(\pi; \bar{\theta})$ denote the total costs (16) incurred by these models when using strategy $\mu^*(\theta)$. Similarly, $J_{\mu^*(\bar{\theta})}(\pi; \theta)$ and $J_{\mu^*(\bar{\theta})}(\pi; \bar{\theta})$ denote the total costs (16) incurred by these models when using strategy $\mu^*(\bar{\theta})$.

Theorem 7: Consider two optimal sampling problems with models $\theta = (A, f)$ and $\bar{\theta} = (\bar{A}, \bar{f})$, respectively. Assume θ and $\bar{\theta}$ satisfy (A2), (A3), (A4), (A5). Then, for misspecified model and misspecified strategy, the following sensitivity bounds hold:

Misspecified Model:

$$\sup_{\pi \in \Pi(X)} |J_{\mu^*(\theta)}(\pi; \theta) - J_{\mu^*(\theta)}(\pi; \bar{\theta})| \le K \|\theta - \bar{\theta}\|. \quad (39)$$

Misspecified Strategy:

$$J_{\mu^*(\bar{\theta})}(\pi, \theta) \le J_{\mu^*(\theta)}(\pi, \theta) + 2K \|\theta - \bar{\theta}\| \quad (40)$$

$$\text{where } K = \frac{\max_i C(e_i, 0)}{1 - \rho_{\theta,\bar{\theta}}},$$

$$\rho_{\theta,\bar{\theta}} = \max_u \sum_{y \ge y^*_{\theta,\bar{\theta}}} \sigma(e_X, y, u; \theta) \quad (41)$$

$$\|\theta - \bar{\theta}\| = \max_{i,u} \sum_{j,y} \left| f_{jy} A^{D_u}|_{ij} - \bar{f}_{jy} \bar{A}^{D_u}|_{ij} \right|.$$

If $A = \bar{A}$, and $D(f_j \| \bar{f}_j) = \sum_y f_{jy} \ln(f_{jy}/\bar{f}_{jy})$ denotes the Kullback–Leibler Divergence,

$$\|\theta - \bar{\theta}\| \le \sqrt{2} \max_{i,u} \sum_j A^{D_u}|_{ij} \left[D(f_j \| \bar{f}_j) \right]^{1/2}. \quad (42)$$

If the observation distributions are Gaussians with variance σ^2, $\bar{\sigma}^2$, respectively, then

$$\|\theta - \bar{\theta}\| \le \left(\frac{\sigma}{\bar{\sigma}} - \ln \frac{\sigma}{\bar{\sigma}} - 1 \right)^{1/2}.$$

The proof of Theorem 7 is in Appendix G. The bounds (39), (40) are tight since $\|\theta - \bar{\theta}\| = 0$ implies that the performance degradation is zero. Also (42) follows from Pinsker's inequality [10] that bounds the total variation norm $\|\theta - \bar{\theta}\|$ by the Kullback–Leibler Divergence.

For optimal sampling problems where the transition matrix or observation distribution does not satisfy assumptions (A2), (A3), or (A4) but is ϵ close to satisfying these conditions, the above result ensures that a monotone strategy yields near optimal behavior with explicit bound on the performance given by (39) and (40).

It is instructive to compare the above theorem with Theorem 6 of Section V-A. Theorem 6 compared optimal costs for different models—it showed that $\theta \succeq \bar{\theta} \implies J_{\mu^*(\theta)}(\pi; \theta) \ge J_{\mu^*(\bar{\theta})}(\pi; \bar{\theta})$, where $\mu^*(\theta)$ and $\mu^*(\bar{\theta})$ denote the optimal sampling strategies for models θ and $\bar{\theta}$, respectively (where the ordering $\succeq$ is specified in Section V-A). In comparison, (39) applies the optimal strategy $\mu^*(\theta)$ for model θ to the decision problem with a different model $\bar{\theta}$. Also (40) is a lower bound for the cost of applying the optimal strategy for a different model $\bar{\theta}$ to the true model θ—this bound is in terms of the cost of the optimal strategy for true model $\bar{\theta}$. What Theorem 7 says is that if the "distance" between the two models $\theta, \bar{\theta}$ is small, then the sub-optimality is small, as described by (39) and (40).

3) Discussion of Assumption (A5): The proof of Theorem 7 is nontrivial since there is no discount factor[7] in the cost (16). However, because the sampling problem terminates with probability one in finite time, it has an implicit discount factor—this is typical in stochastic shortest path problems that terminate in finite time [5]. It is here that Assumption (A5) is used. (A5) implies that $\rho_{\theta,\bar{\theta}} < 1$. The term $1 - \rho_{\theta,\bar{\theta}}$ can be interpreted as a lower bound to the probability of stopping at any given time. Since this is nonzero, the term $\rho_{\theta,\bar{\theta}}$ in (41) serves as this implicit discount factor.

(A5) says that for models θ and $\bar{\theta}$, if the underlying Markov chain is in state e_X, then there is a nonzero probability that a noisy observation pulls the belief state to the stopping region. As shown in the proof, the choice of state e_X is because it is furthest away from the stopping state e_1 (with respect to the MLR order).

C. Stochastic Dominance Properties of the Bayesian Filter

This section presents structural properties of the Bayesian filter (10), which determines the evolution of the belief state π.

[7] Instead of (16), if the cost was $J_\mu(\pi_0) = \mathbb{E}^\mu_{\pi_0} \left\{ \sum_{k=0}^{k^*-1} \rho^k C(\pi_k, u_k) + \rho^{k^*} C(\pi_{k^*}, u_{k^*} = 0) \right\}$, where the user defined discount factor $\rho \in [0, 1)$, then establishing a bound such as (39) is straightforward. An artificial discount factor ρ is unnatural in our problem and unnecessary as shown in Theorem 7 since the problem terminates in finite time with probability one and hence has an implicit discount factor denoted as $\rho_{\theta,\bar{\theta}}$.

Indeed, the proofs of Theorems 2–7 presented in previous sections depend on Theorem 9 given below. The results in Theorem 9 are also of independent interest in Bayesian filtering and prediction. To compare posterior distributions of Bayesian filters, we first start with some background definitions on stochastic orders.

1) Stochastic Orders: We already introduced the MLR stochastic order in Definition 1 of Section IV.

Definition 2 (First-Order Stochastic Dominance, [24]): 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 $j = 1, \ldots, X$.

The following result is well known [24]. It says that MLR dominance implies first-order stochastic dominance and gives a necessary and sufficient condition for stochastic dominance.

Theorem 8 (see [24]):

i) Let $\pi_1, \pi_2 \in \Pi(X)$. Then, $\pi_1 \geq_r \pi_2$ implies $\pi_1 \geq_s \pi_2$.
ii) Let $\mathcal{V}$ denote the set of all X-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 $v \in \mathcal{V}$, $v'\pi_1 \geq v'\pi_2$.

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 $\pi \in \Pi(X)$.

2) Main Result: With the above definitions, we are now ready to state the main result regarding the stochastic dominance properties of the Bayesian filter.

Theorem 9: The following structural properties hold for the Bayesian filtering update $T(\pi, y, u)$ and normalization measure $\sigma(\pi, y, u)$ defined in (10):

1) Under (A2), $\pi_1 \geq_r \pi_2$ implies $T(\pi_1, y, u) \geq_r T(\pi_2, y, u)$.
2) Under (A2), (A3), $\pi_1 \geq_r \pi_2$ implies $\sigma(\pi_1, \cdot, u) \geq_s \sigma(\pi_2, \cdot, u)$.
3)
   a) Under (A2), $A^{D_u}{}'e_X \geq_r A^{D_{u+1}}{}'e_X$ and $A^{D_{u+1}}{}'e_1 \geq_r A^{D_u}{}'e_1$ for all $u \in \{1, 2, \ldots, L\}$. So for $X = 2$, $A^{D_u}|_{i2}$ is submodular. That is, $A^{D_{u+1}}|_{12} - A^{D_u}|_{12} \geq A^{D_{u+1}}|_{22} - A^{D_u}|_{22}$.
   b) Under (A2), (A3), for $X = 2$, $\sigma(\pi, \cdot, u)$ is submodular with respect to $\geq_s$ (recall for $X = 2$ that $\geq_s$ and $\geq_r$ coincide):

$$\sum_{y \geq \bar{y}} [\sigma(\pi, y, u+1) - \sigma(\pi, y, u)] \leq \sum_{y \geq \bar{y}} [\sigma(\bar{\pi}, y, u+1) - \sigma(\bar{\pi}, y, u)] \quad \text{for } \pi \geq_s \bar{\pi}.$$

4) For $y, \bar{y} \in \mathbb{Y}$, $y > \bar{y}$ implies $T(\pi_1, y, u) \geq_r T(\pi_1, \bar{y}, u)$ iff (A3) holds.
5) Consider the ordering of transition matrices $A \succeq \bar{A}$ defined in (36).
   a) If $A \succeq \bar{A}$, then $A'\pi \geq_r \bar{A}'\pi$, i.e., the one-step Bayesian predictor with transition matrix A MLR dominates that with transition matrix $\bar{A}$.
   b) If $A \succeq \bar{A}$ and (A2) holds, then $(A^l)'\pi \geq_r (\bar{A}^l)'\pi$ for any positive integer l. That is, the l-step Bayesian predictor preserves this MLR dominance.
6) If (A2) holds, then for $\pi \in \Pi_{\mathcal{M}}$ defined in (28) or (33), $T(\pi, y, u) \geq_r T(\pi, y, u+1)$ and $\sigma(\pi, \cdot, u) \geq_s \sigma(\pi, \cdot, u+1)$.
7) Let $T(\pi, y, u; A)$ and $\sigma(\pi, y, u; A)$ denote, respectively, the Bayesian filter update and normalization measure using transition matrix A. Then, they satisfy the following stochastic dominance property with respect to the ordering of A defined in (36):
   a) $A \succeq \bar{A}$ implies $T(\pi, y, u; A) \geq_r T(\pi, y, u; \bar{A})$.
   b) Under (A3), $A \succeq \bar{A}$ implies $\sigma(\pi, \cdot, u; A) \geq_s \sigma(\pi, \cdot, u; \bar{A})$.

The proof of Theorem 9 is in Appendix H.

In words, Part 1 of the theorem implies that the Bayesian filtering recursion preserves the MLR ordering providing that the transition matrix is TP2 (A2). Part 2 says that the normalization measure preserves first-order stochastic dominance providing (A2) and (A3) hold. Part 3 shows that the normalization measure is submodular. Part 4 shows that under (A3), the larger the observation value, the larger the posterior distribution (wrt MLR order). Part 5 shows that if starting with two different transition matrices but identical priors, then the optimal predictor with the larger transition matrix [in terms of the order introduced in (36)] MLR dominates the predictor with the smaller transition matrix. Part 6 says address the two-state case. It says that for all belief states in $\Pi_{\mathcal{M}}$, as long as the transition matrix is TP2 (satisfies A2), then the belief state update for smaller sampling intervals dominates that of larger sampling intervals. Finally, Part 7 says that starting with two different transition matrices but identical priors, the filtering recursion $T(\pi, y, u)$ and the normalization measure $\sigma(\pi, y, u)$ with the larger transition matrix [in terms of the order introduced in (36)] dominate the predictor with the smaller transition matrix.

VI. Optimal Sampling With Measurement Control

So far we have considered optimal sampling where the control action affects the transition probabilities. In this section, we discuss optimal sampling with measurement control where the action affects both the transition probabilities and observation distribution. Section VI-A shows that a similar structural result to Theorem 4 holds for the optimal policy. That is, for $X \geq 2$, the optimal strategy can be lower bounded by a monotone strategy. Section VI-B discusses general measurement control problems and shows that the techniques developed in this paper for proving the optimal strategy is monotone cannot be applied to such problems. Finally, Section VI-C discusses the special case of measurement control with noninformative observations considered in [4].

A. Structural Results for Optimal Sampling With Measurement Control

Consider the optimal sampling problem of a noisy Markov chain in Section II, but now the observation probabilities in (5) are action dependent. That is, for each $u \in \{1, 2, \ldots, L\}$,

$$P(y_{k+1} \leq \bar{y} | x_{\tau_k} = e_x, u_k = u) = \sum_{y \leq \bar{y}} f_{xyu}, \quad x \in \mathbb{X}. \quad (43)$$

Denote $f_u = (f_{xyu}, x \in \mathbb{X}, y \in \mathbb{Y})$. Choosing action $u_k = u$ is equivalent to looking at the noisy Markov chain with observation distribution f_u at time $\tau_{k+1} = \tau_k + D_u$. Assume that the observation distributions are reverse Blackwell ordered [defined in (37)] as

$$f_u \succeq_B f_{u+1}, \quad u \in \{1, \ldots, L-1\}. \tag{44}$$

Thus, action u yields less accurate measurements but samples more frequently than action $u+1$. Should the decision maker sample less frequently but more accurately or more frequently but less accurately?

Define the observation dependent measurement cost $\bar{m}(x, y, u)$ as the cost of taking action $u_k = u$, obtaining observation $y_{k+1} = y$ when the state is $x_{k+1} = e_x$. As in (8), this measurement cost can be expressed in terms of the measurement cost of state $x_k = e_i$ and action $u_k = u$ as

$$m(e_i, u) = \sum_{y \in \mathbb{Y}} \sum_{j \in \mathbb{X}} \bar{m}(j, y, u) A^{D_u}|_{ij} f_{jyu}. \tag{45}$$

Consider the myopic strategy $\underline{\mu}(\pi)$ and myopic stopping set $\underline{\mathcal{S}}$ defined in (32).

Theorem 10: Consider the optimal sampling problem with measurement control. Then, the conclusions of Theorem 4 hold. Statement 3 requires (A1), (A2), (A3) and that the observation distributions satisfy the Blackwell dominance conditions (44).

The proof is in Appendix I. It combines Blackwell dominance of observation measures together with MLR monotone structure of the value function.

To illustrate Statement 3 of Theorem 10, suppose $\mathcal{U} = \{0, 1, 2\}$. Then, the theorem asserts the following: Given current belief π, if the current expected cost satisfies $C(\pi, 2) < C(\pi, 1)$, then it is optimal to choose action 2, that is, it is optimal to look less frequently but more accurately. Thus, for such belief states, the myopic strategy $\underline{\mu}(\pi)$ coincides with the optimal strategy $\mu^*(\pi)$. This is a nontrivial statement since, in general, just because the expected instantaneous costs satisfy $C(\pi, 2) < C(\pi, 1)$, does not necessarily imply that the myopic strategy coincides with the optimal strategy.

Let us comment on (A1) in the context of the above cost. Suppose the measurement costs $\bar{m}(x, y, u)$ are increasing in x implying that measurements before hitting the target state 1 are more expensive than after hitting the target state. A similar assumption is nicely motivated in [4] in terms of minimizing the average number of observations used before hitting the target state. It is easily shown that under (A2) and (A3) if $\bar{m}(x, y, u)$ is increasing in x, then $m(e_i, u)$ in (45) is increasing in i. Therefore, if $\bar{m}(x, y, u)$ increases sufficiently fast with x, then (A1) holds.

B. Discussion: Monotone Policies for Measurement Control

Section VI-A shows that in measurement control, the optimal strategy can be lower bounded by a monotone strategy. Can the optimal strategy be shown to be monotone using similar techniques to Theorem 2? The answer is no. The reason is that it is not possible to find two nontrivial stochastic matrices (observation distribution kernels) f and $\bar{f}$ such that the belief updates satisfy (i) $T(\pi, y; f) \geq_r T(\pi, y; \bar{f})$ and normalization measure satisfies (ii) $\sigma(\pi, \cdot, f) \geq_s \sigma(\pi, \cdot; \bar{f})$. In [23] and [29], it is claimed that if f TP2 dominates $\bar{f}$, then (i) and (ii) hold. However, we have found that the only examples of stochastic kernels that satisfy the TP2 dominance are the trivial example $f = \bar{f}$.

C. Measurement Control and Interpretation of [4]

As mentioned above, the methods developed here do not apply to proving monotone optimal policies for general measurement control problems. However, for special cases, measurement control is equivalent to a sampling control and our monotone results apply. We now discuss this.

The recent paper [4] considers quickest detection with measurement control for a two-state Markov chain observed in noise. At each time t, a decision is made whether to take a measurement or not. The timing is identical to sensor scheduling problems [17]: Given belief $\pi_t(i) = P(x_t = e_i | y_1, \ldots, y_t, \bar{u}_1, \ldots, \bar{u}_t)$, the action $\bar{u}_{t+1}$ at time $t+1$ is chosen as

$$\bar{u}_{t+1} = \mu(\pi_t) \in \bar{\mathcal{U}} \stackrel{\triangle}{=} \{0 \text{ (stop at time } t), m, \bar{m}\},$$

where m denotes take a measurement at $t+1$ and $\bar{m}$ denotes do not take a measurement at $t+1$.

Below, we show that a constrained version of the problem considered in [4] is identical to the optimal sampling problem with quickest detection formulated in Section II-C.

Lemma 3: Consider the above measurement control problem. Introduce the constraint[8] that stopping is allowed at time $t+1$ only if a measurement is taken at time $t+1$. Then

1) Measurement control with action space $\bar{\mathcal{U}}$ is equivalent to the sampling control with finite action space $\mathcal{U} = \{0, 1, 2, \ldots, L\}$ and sampling intervals $D_u = u$, where $L = \min\{u : C(e_i, u) > C(e_i, 0), i \in \mathbb{X}\}$.
2) Consider quickest detection with geometric distributed change time ($X = 2$). If (A3) holds, then measurement control is equivalent to sampling control with finite action space $\mathcal{U} = \{0, 1, \ldots, L^*\}$, where $L^* > \frac{\log d - \log(d+1-A_{22})}{\log A_{22}}$ and d is the delay penalty. ■

Proof:

1. Introducing the above constraint to the action space $\bar{\mathcal{U}}$ yields the equivalent action space $\tilde{\mathcal{U}} = \{0, \bar{m}^{l-1}m, l = 1, 2, \ldots\}$, where compound action $\bar{m}^{l-1}m$ denotes $l-1$ successive "don't take measurements" followed by "take a measurement." Next, choosing $\bar{u} = \bar{m}^{l-1}m \in \tilde{\mathcal{U}}$ is identical to the choice of action $u = l \in \mathcal{U} = \{0, 1, 2, \ldots\}$ with sampling interval $D_u = u$. Since the elements of $C(e_i, u) - m(e_i, u)$, $u \geq 1$ in (26) are strictly increasing in u and $m(e_i, u)$ is bounded, it follows that $L = \min\{u : C(e_i, u) \geq C(e_i, 0), i \in \mathbb{X}\}$ is finite. Actions $u > L$ are never chosen since stopping ($u = 0$) is cheaper. For each action in $\tilde{\mathcal{U}}$, the belief state update given by (10) is identical to the corresponding action in $\mathcal{U}$ since the observations obtained with action $\bar{m}$ are

[8] Banerjee and Veeravalli [4] do not impose the constraint that stopping is allowed only after a measurement is taken. Then the problem is different. In [4], the optimality of threshold policies for such measurement control problems is shown. In [4], a two-three threshold strategy is shown to be optimal for quickest detection and also threshold type policies are shown to be asymptotically optimal. Paper [4] also contains a nice performance analysis of sub-optimal and nearly optimal strategies. It is an interesting future work to extend the structural results and analysis in [4] to quickest state estimation problems.

noninformative. So if the costs for the actions are chosen as in (16), the problem is equivalent to optimal sampling.

2. Start with the optimal sampling problem with action space $\mathcal{U} = \{1, 2, \ldots\}$. Theorem 5 shows that for geometric change times, if (A3) holds, then the optimal sampling strategy $\mu^*(\pi) \leq \bar{\mu}(\pi)$ for all $\pi \in \Pi(X)$. Here, $\bar{\mu}(\pi) = \operatorname{argmin}_u \bar{C}(\pi, u)$ is the myopic strategy defined in (35), and furthermore, $\bar{\mu}(\pi)$ is increasing in π. So $\mu^*(\pi) \leq \bar{\mu}(\pi) \leq \bar{\mu}(e_X)$ for all $\pi \in \Pi(X)$. So the maximum action ever chosen by $\mu^*(\pi) \in \mathcal{U}$ is upper bounded by $\bar{\mu}(e_X) = \operatorname{argmin}_u \bar{C}(e_X, u)$. To establish a bound L^* on this maximum action, it suffices to show that $\bar{C}(e_X, u) \geq \bar{C}(e_X, u+1)$ for $u < L^*$ and $\bar{C}(e_X, L^*) \leq \bar{C}(e_X, L^*+1)$. For the costs (35), $X = 2$ and absorbing state 1, choosing $\bar{\phi} = Ae_1/A_{22}$ satisfies the conditions of Statement 3 of Theorem 5. Straightforward algebraic manipulations yield $L^* > \frac{\log d - \log(d+1-A_{22})}{\log A_{22}}$. Thus, L^* is a positive integer by construction.

■

VII. Numerical Examples

Example 1. Optimal Sampling With Binary Erasure Channel Measurements: Consider $X = 2, Y = 3, L = 5$, $d = 0.0235$, $m(e_1, u) = 0$, $m(e_2, u) = 0.1647$, $\{D_1, D_2, D_3, D_4\} = \{1, 3, 5, 10\}$.

(a) Quickest Detection: Suppose

$$A = \begin{bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{bmatrix}, \; f = \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0 & 0.2 & 0.8 \end{bmatrix}.$$

The noisy observations of the Markov chain specified by observation probabilities f models a binary nonsymmetric erasure channel [10]. Note that a binary erasure channel is TP2 by construction (all second-order minors are nonnegative) and so (A3) holds.

The optimal strategy was computed by forming a grid of 1000 values in the 2-D unit simplex, and then solving the value iteration algorithm (46) over this grid on a horizon N such that $\sup_\pi |V_N(\pi) - V_{N-1}(\pi)| < 10^{-6}$. Fig. 1(a) shows that when the conditions of Theorem 2 are satisfied, the strategy is monotone decreasing in posterior $\pi(1)$. To show that the sufficient conditions of Theorem 2 are useful, Fig. 1(b) gives an example of when these conditions do not hold, the optimal strategy is no longer monotone. Here, $m(e_1, u) = 0.1647$, $m(e_2, u) = 0$ and therefore violates (A1) of Theorem 1.

(b) Quickest Estimation: Next consider the quickest state estimation problem with identical parameters to above except that $A = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$. The problem satisfies the assumptions of Theorem 2 and the optimal strategy $\mu^*(\pi)$ is monotone decreasing in $\pi(1) \in [0, \pi_{\mathcal{M}}]$, where $\pi_{\mathcal{M}} = 0.8$ is computed via (28). The optimal strategy is illustrated in Fig. 1(c). Next consider $A = \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$. This violates (A2). The optimal strategy is nonmonotone, as illustrated in Fig. 1(d). Fig. 1(c) and (d) highlights a significant difference between quickest detection and quickest state estimation. In the former, the transition matrix satisfies (A2) by definition and so is irrelevant in terms of the existence of a monotone strategy. In the latter, choice of a suitable transition matrix is important as shown in the figures.

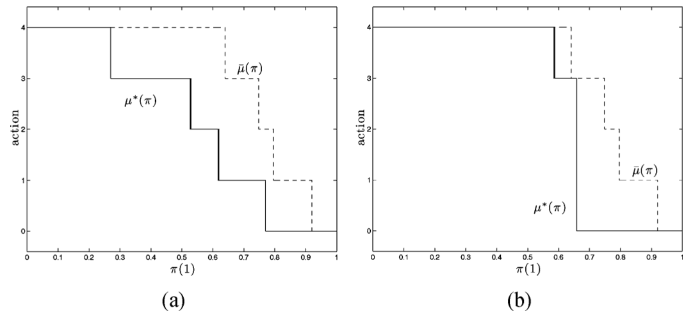


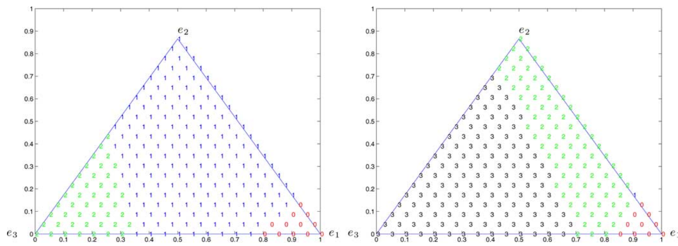
Fig. 4. Optimal sampling strategy for action $u \in \{0 \;(\text{announce change}), 1, 2, 3, 4\}$ for a quickest-change detection problem with geometric change time. The parameters are specified in Examples 2 and 3 in Section VII. The optimal strategy $\mu^*(\pi)$ is monotone decreasing in $\pi(1)$ and is upper bounded by myopic strategy $\bar{\mu}$ according to Theorem 5. (a) Gaussian Observation Probabilities. (b) Poisson Observation Probabilities.

Example 2. Optimal Sampling Quickest Detection With Gaussian Noise Measurements: Here, we consider identical parameters to Example 1 except that the observation distribution is Gaussian with $f_{1y} \sim \mathcal{N}(1, 1)$, $f_{2y} \sim \mathcal{N}(2, 1)$ and measurement costs are $m(e_i, u) = 1$ for all $i \in \mathbb{X}, u \in \{1, 2, 3, 4\}$. Since the measurement cost is a constant, (A1) and (A4) of Theorem 1 hold trivially. As mentioned in Section III-C, (A3) holds for Gaussian distribution. Therefore, Theorem 1 applies and the optimal strategy $\mu^*(\pi)$ is monotone decreasing in $\pi(1)$. Fig. 4 illustrates the optimal strategy. Next, using Theorem 5, the myopic strategies $\bar{\mu}(\pi)$ form an upper bound to the optimal strategy $\mu^*(\pi)$ for actions $u \in \{1, \ldots, 4\}$. We chose $\bar{\phi} = Ae_1/A_{22}$ which clearly satisfies the conditions of Theorem 5(3) for constructing an upper bound $\bar{\mu}(\pi)$ with myopic cost in (35). As a bound for the optimal stopping region, we used the myopic stopping set $\underline{\mathcal{S}}$ defined in (32). These are plotted in Fig. 4(a).

Example 3. Optimal Sampling Quickest Detection With Markov Modulated Poisson Measurements: The parameters here are identical to Example 2 except that the observations are generated by a discrete time Markov Modulated Poisson process. That is, at each time τ_k, noisy observations of the Markov chain are obtained from the Poisson distribution $f_{xy} = (\lambda_x)^{y-1} \frac{e^{-\lambda_x}}{(y-1)!}, x \in \{1, 2\}$, with rates $\lambda_1 = 1, \lambda_2 = 1.5$. Since (A3) holds for Poisson distribution, Theorem 1 applies. Fig. 4(b) illustrates the optimal strategy and upper bound myopic strategy $\bar{\mu}(\pi)$ computed as in Example 2.

Example 4. Optimal Sampling With Phase-Distributed Change Time: Here, we consider optimal sampling quickest detection with PH-distributed change time. Consider a three-state ($X = 3$) Markov chain observed in noise with parameters $d = 0.04, m(e_i, u) = 0.1, \{D_1, D_2, D_3\} = \{2, 4, 5\}$,

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0.7 & 0.3 & 0 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}, f = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.1 & 0.9 \end{bmatrix}.$$



(a) (b)

Fig. 5. Optimal sampling strategy for action $u \in \{0 \text{ (announce change)}, 1, 2, 3\}$ for a quickest-change detection problem with PH-distributed time specified by 3-state Markov chain in Example 4 of Section VII. The belief space $\Pi(X)$ is a 2-D unit simplex (equilateral triangle). The optimal strategy is upper bounded by myopic strategy $\bar{\mu}(\pi)$ according to Theorem 5. (a) Optimal Policy $\mu^*(\pi)$. (b) Myopic Upper Bound $\bar{\mu}(\pi)$.

So $\Pi(X)$ is a 2-D unit simplex depicted in Fig. 5. The optimal strategy was computed by forming a grid of 8000 values in the 2-D unit simplex, and then solving the value iteration algorithm (46) over this grid on a horizon N such that $\sup_\pi |V_N(\pi) - V_{N-1}(\pi)| < 10^{-6}$. Fig. 5(a) shows the optimal strategy.

It can be verified that the transition matrix A satisfies (36) and (A2) so that $\Pi_{\mathcal{M}} = \Pi(X)$ by Lemma 1. The observation distribution f satisfies (A3). Also the conditions of Theorem 5(2) are satisfied for $\bar{\phi} = dAe_1/100$ in (35). Therefore, Theorem 5 holds and the optimal strategy is upper bounded by the myopic strategy $\bar{\mu}(\pi)$ defined in (35). Fig. 5(b) shows the myopic strategy $\bar{\mu}(\pi)$. As a bound for the optimal stopping region, we used the myopic stopping set $\underline{\mathcal{S}}$ defined in (32). In Fig. 5(b), these are represented by "0."

VIII. Discussion

The paper presented structural results for the optimal sampling strategy of a Markov chain given noisy measurements. Examples dealing with quickest detection and quickest estimation with optimal sampling were presented. The main results were: Theorems 1 and 2 gave sufficient conditions for the existence of a monotone optimal sampling strategy (with respect to the posterior distribution) when the underlying Markov chain had two states ($X = 2$). It justified the intuition that in both quickest detection and quickest state estimation problems, one should make measurements less frequently when the posterior estimate of the state is away from the target state; see Fig. 1. For multistate Markov chains ($X \geq 2$), Theorems 4 and 5 gave sufficient conditions for the myopic sampling strategy to form a lower bound or upper bound to the optimal sampling strategy. Theorem 6 gave a partial ordering for the transition matrix and noise distributions so that the expected cost of the optimal sampling strategy increased as these parameters increased. Theorem 7 gave explicit bounds on the sensitivity of the total sampling cost with respect to sampling strategy in terms of the Kullback–Leibler divergence between the noise distributions. Theorem 9 gave several useful structural properties of the optimal Bayesian filtering update including sufficient conditions that preserve monotonicity of the filter with observation, prior distribution, transition matrix, and noise distribution. Finally, Theorem 10 shows that for joint sampling and measurement control, the optimal strategy can be lower bounded by a myopic strategy.

The assumptions (A1)–(A5) used in this paper are set valued; so even if the precise parameters (transition probabilities, observation distribution, costs) are not known, as long as they belong to the appropriate sets, the structural results hold.

Appendix

A) Value Iteration Algorithm: The proof of the structural results in this paper will use the value iteration algorithm [13]. Let $n = 1, 2, \ldots,$ denote iteration number. The value iteration algorithm proceeds as follows: $V_{n+1}(\pi) = \min_{u \in \mathcal{U}} Q_n(\pi, u)$,

$$Q_n(\pi, u) = C(\pi, u) + \sum_{y \in \mathbb{Y}} V_n(T(\pi, y, u))\, \sigma(\pi, y, u),$$
$$\text{and } Q_n(\pi, 0) = C(\pi, 0) \text{ initialized by } V_0(\pi) = 0. \tag{46}$$

Let $\mathcal{B}(X)$ denote the set of bounded real-valued functions on $\Pi(X)$. Then, for any V and $\tilde{V} \in \mathcal{B}(X)$, define the sup-norm metric $\sup \|V(\pi) - \tilde{V}(\pi)\|, \pi \in \Pi(X)$. The value iteration algorithm (46) will generate a sequence of value functions $\{V_k\} \subset \mathcal{B}(X)$ that will converge uniformly (sup-norm metric) as $k \to \infty$ to $V(\pi) \in \mathcal{B}(X)$, the optimal value function of Bellman's equation. However, since the belief state space $\Pi(X)$ is an uncountable set, the value iteration algorithm (46) does not translate into practical solution methodologies as $V_k(\pi)$ needs to be evaluated at each $\pi \in \Pi(X)$, an uncountable set. Nevertheless, the value iteration algorithm provides a natural method for proving our results on the structure of the optimal strategy via mathematical induction.

B) Proof of Theorem 2: Recall the definition of the MLR order $\geq_r$ in Section IV and first-order dominance $\geq_s$ in Section V-C. To prove the existence of a monotone optimal strategy, we show that $Q(\pi, u)$ in (18) is a submodular function on the partially ordered set (poset) $[\Pi(X), \geq_r]$. Note that $[\Pi(X), \geq_r]$ is a lattice since given any two belief states $\pi_1, \pi_2 \in \Pi(X)$, $\sup\{\pi : \pi \leq_r \pi_1, \pi \leq_r \pi_2\}$ and $\inf\{\pi : \pi \geq_r \pi_1, \pi \geq_r \pi_2\}$ lie in $\Pi(X)$. For $X = 2$, $\Pi(X)$ is the unit interval $[0, 1]$ and in this case $[\Pi(X), \geq_r]$ is a chain (totally ordered set) and $\geq_r$ coincides with $\geq_s$.

Definition 3 (Submodular Function [33]): $\phi : \Pi(X) \times \{1, 2\} \to \mathbb{R}$ is submodular (antitone differences) if $\phi(\pi, u) - \phi(\pi, \bar{u}) \leq \phi(\tilde{\pi}, u) - \phi(\tilde{\pi}, \bar{u})$, for $\bar{u} \leq u, \pi \geq_r \tilde{\pi}$.

The following result says that for a submodular function $Q(\pi, u)$, $\mu^*(\pi) = \operatorname{argmin}_u Q(\pi, u)$ is increasing in its argument π. This will be used to prove the existence of a monotone optimal strategy in Theorem 2.

Theorem 11 (see [33]): If $\phi : \Pi(X) \times \mathcal{U} \to \mathbb{R}$ is submodular, then there exists a $\mu^*(\pi) = \operatorname{argmin}_{u \in \mathcal{U}} \phi(\pi, u)$, that is MLR increasing on $\Pi(X)$, i.e., $\tilde{\pi} \geq_r \pi \implies \mu^*(\pi) \leq \mu^*(\tilde{\pi})$. ■

Finally, we state the following result.

Theorem 12: The sequence of value function $\{V_n(\pi), n = 1, 2, \ldots\}$, generated by the value iteration algorithm (46), and optimal value function $V(\pi)$ defined in (18) satisfy:

i) $V_n(\pi)$ and $V(\pi)$ are concave in $\pi \in I$.
ii) Under (A1), (A2), (A3), $V_n(\pi)$ and $V(\pi)$ are increasing in π with respect to the MLR stochastic order on $\Pi(X)$. ■

Statement (i) is well known for POMDPs; see [9] for a tutorial description. Statement (ii) is proved in [23, Proposition 1] using mathematical induction on the value iteration algorithm.

Finally, we present the proof of Lemma 1.

Proof of Lemma 1: (i) and (ii): Start with the following lemma:

Lemma 4: If A is TP2 [i.e., satisfies (A2)], then $A'e_X \geq_r A^{2'}e_X$. Therefore, $A^{D_u{}'}e_X \geq_r A^{D_{u+1}{}'}e_X$. For $X = 2$, this implies, $A_{22} \geq A^2|_{22}$ and $A^{D_u}|_{22} \geq A^{D_{u+1}}|_{22}$.

Proof: From elementary matrix operations, it follows that

$$\begin{bmatrix} A_{X,1:X} \\ e'_X \end{bmatrix} A = \begin{bmatrix} e'_X A^2 \\ e'_X A \end{bmatrix}.$$

The first matrix is TP2 by construction, and A is TP2 by (A2). The product of TP2 matrices is TP2 [16, Th. 3.1, p. 107]. Therefore, the right hand side is TP2. This is equivalent to the first row being MLR dominated by the second row, i.e., $A'e_X \geq_r A^{2'}e_X$ (see footnote associated with (A2) and (A3) or [16, p. 122]).

Finally, $A'e_X \geq_r A^{2'}e_X$ implies $A^{2'}e_X \geq_r A^{3'}e_X$ by Theorem 9(1) under (A2). This implies $A'e_X \geq_r A^{2'}e_X \geq_r A^{3'}e_X$ by the transitive property of the MLR dominance. Therefore, $A^{D_u{}'}e_X \geq_r A^{D_{u+1}{}'}e_X$ since $D_u < D_{u+1}$ by assumption.

For $X = 2$, recall $\geq_r$ and $\geq_s$ coincide. ■

i) For $X = 2$, from (28), $\Pi_{\mathcal{M}}$ is a convex polytope. A 1-D polytope is an interval. Also we established above that $e_2 \in \Pi_{\mathcal{M}}$. Therefore, $\Pi_{\mathcal{M}}$ in terms of $\pi(1)$ is an interval of the form $[0, \pi_{\mathcal{M}}]$.

Proof: With the above preparation, we present the proof of Theorem 2.

The first claim follows from the general result that the stopping set $\mathcal{S}$ for a POMDP is always a convex subset of $\Pi(X)$—see Theorem 4. Of course, a 1-D convex set is an interval and since $e_1 \in \mathcal{S}$ [Assumption A1(ii)], it follows that the interval $\mathcal{S} = (\pi_1^*, 1]$.

In light of the first claim, the optimal strategy is of the form

$$\mu^*(\pi) = \begin{cases} 0, & \pi \in \mathcal{S} \\ \text{argmin}_{u\in\{1,2,\ldots,L\}} Q(\pi, u), & \pi \in \Pi(X) - \mathcal{S} \end{cases}.$$

So to prove the second claim, we focus on belief states in the interval $\Pi_{\mathcal{M}} - \mathcal{S}$ and consider actions $u \in \{1, 2, \ldots, L\}$. To prove that $\mu^*(\pi)$ is increasing in $\pi \in \Pi_{\mathcal{M}} - \mathcal{S}$, from Theorem 11, we need to prove that $Q(\pi, u)$ is submodular, i.e.,

$$Q(\pi, u) - Q(\pi, \bar{u}) - Q(\bar{\pi}, u) + Q(\bar{\pi}, \bar{u}) \leq 0, \quad u > \bar{u},\ \pi \geq_s \bar{\pi}.$$

The proof is similar to [1]. Recall $\geq_s$ and $\geq_r$ coincide for $X = 2$. From (18), the left hand side of the above expression is

$$\begin{aligned} &C(\pi, u) - C(\pi, \bar{u}) - C(\bar{\pi}, \bar{u}) + C(\bar{\pi}, \bar{u}) \\ &+\sum_y V(T(\pi,y,\bar{u}))[\sigma(\pi,y,u) - \sigma(\pi,y,\bar{u}) - \sigma(\bar{\pi},y,u) + \sigma(\bar{\pi},y,\bar{u})] \\ &+\sum_y [V(T(\pi, y, \bar{u})) - V(T(\bar{\pi}, y, \bar{u}))]\,\sigma(\bar{\pi}, y, \bar{u}) \\ &+\sum_y [V(T(\pi, y, \bar{u})) - V(T(\pi, y, u))]\,\sigma(\pi, y, u) \\ &+\sum_y [V(T(\bar{\pi}, y, u)) - V(T(\pi, y, \bar{u}))]\,\sigma(\bar{\pi}, y, u). \end{aligned} \tag{47}$$

Since the cost is submodular by (A4), the first line of (47) is negative. Since $V(\pi)$ is MLR increasing from Theorem 12 [under (A1), (A2), (A3)] and $T(\pi, y, u)$ is MLR increasing in y [under (A3)] from Theorem 9(4), it follows that $V(T(\pi, y, u))$ is MLR increasing in y. Therefore, since $\sigma(\pi, \cdot, u)$ is submodular from Theorem 9(3) under (A2), (A3), the second line of (47) is negative.

It only remains to prove that the third, fourth and fifth lines of (47) are negative. From statements (1) and (6) of Theorem 9, it follows under (A2) that for $\pi, \bar{\pi} \in \Pi_{\mathcal{M}}$,

$$\begin{aligned} &T(\pi, y, \bar{u}) \geq_s T(\pi, y, u) \geq_s T(\bar{\pi}, y, u), \\ &T(\pi, y, \bar{u}) \geq_s T(\bar{\pi}, y, \bar{u}) \geq_s T(\bar{\pi}, y, u), \quad u > \bar{u}, \pi \geq_s \bar{\pi}. \end{aligned}$$

Assume that $T(\pi, y, u) \geq_s T(\bar{\pi}, y, \bar{u})$ (for the reverse ordering an identical result holds). For $X = 2$, $\Pi(X)$ is a 1-D simplex that can be represented by $\pi(2) \in [0, 1]$. So below, we represent π, $T(\pi, y, u)$, etc. by their second elements. Then, using concavity of $V(\cdot)$, we can express the last two summations in (47) as follows:

$$\begin{aligned} &V(T(\pi, y, \bar{u})) - V(T(\pi, y, u)) \\ &\leq [T(\pi, y, \bar{u}) - T(\pi, y, u)]\,\frac{V(T(\pi, y, \bar{u})) - V(T(\bar{\pi}, y, \bar{u}))}{T(\pi, y, \bar{u}) - T(\bar{\pi}, y, \bar{u})} \\ &V(T(\bar{\pi}, y, u)) - V(T(\pi, y, \bar{u})) \\ &\leq \frac{T(\bar{\pi}, y, u) - T(\pi, y, \bar{u})}{T(\pi, y, \bar{u}) - T(\bar{\pi}, y, \bar{u})}\,[V(T(\pi, y, \bar{u})) - V(T(\bar{\pi}, y, \bar{u}))] \\ &+ V(T(\bar{\pi}, y, \bar{u})) - V(T(\pi, y, \bar{u})). \end{aligned}$$

Using these expressions, the summation of the last three lines of (47) are upper bounded by

$$\begin{aligned} &\sum_y \Big[V(T(\pi, y, \bar{u})) - V(T(\bar{\pi}, y, \bar{u}))\Big] \\ &\quad\Bigg[\sigma(\bar{\pi}, y, \bar{u}) + \frac{T(\pi, y, \bar{u}) - T(\pi, y, u)}{T(\pi, y, \bar{u}) - T(\bar{\pi}, y, \bar{u})}\sigma(\pi, y, u) \\ &\quad + \left(\frac{T(\bar{\pi}, y, u) - T(\pi, y, \bar{u})}{T(\pi, y, \bar{u}) - T(\bar{\pi}, y, \bar{u})} - 1\right)\sigma(\bar{\pi}, y, u)\Bigg]. \end{aligned} \tag{48}$$

Since $V(\pi)$ is MLR increasing (see Theorem 12) and $T(\pi, y, \bar{u}) \geq_r T(\bar{\pi}, y, \bar{u})$ (using the fact that $\pi \geq_r \bar{\pi}$ and Statement 1 of Theorem 9), clearly $V(T(\pi, y, \bar{u})) - V(T(\bar{\pi}, y, \bar{u})) \geq 0$. The term in square brackets in (48) can be expressed as (see [1])

$$\frac{f_{2y} f_{1y} (\pi - \bar{\pi})\left(A^{D_u}|_{22} - A^{D_u}|_{12} - A^{D_{\bar{u}}}|_{22} + A^{D_{\bar{u}}}|_{12}\right)}{\sigma(\pi, y, \bar{u})[T(\pi, y, \bar{u}) - T(\bar{\pi}, y, \bar{u})]}. \tag{49}$$

From Theorem 9, Part 3(a), under (A2) $A^{D_u}|_{22} - A^{D_u}|_{12} - A^{D_{\bar{u}}}|_{22} + A^{D_{\bar{u}}}|_{12}$ for $\bar{u} < u$ is negative, and so (49) is negative. Hence, (47) is negative, thereby concluding the proof.

C) *Proof of Theorem 3:*

Statement 1: Consider Bellman's equation (18) and define $\underline{V}(\pi) = V(\pi) + \phi'\pi$. It is easily checked that $\underline{V}(\pi)$ satisfies Bellman's equation with costs $C(\pi, u)$ replaced by $\underline{C}(\pi, u)$ defined in (30). Also since $\phi'\pi$ is functionally independent of the minimization variable u, the argument of the minimum of (18), which is the optimal strategy $\mu^*(\pi)$, is unchanged. In terms of the value iteration algorithm (46), it requires initialization with $\underline{V}_0(\pi) = \phi'\pi$.

Statement 2: Recall for $X = 2$, MLR $\geq_r$ and first-order stochastic dominance $\geq_s$ coincide. Since the aim is to transform the delay cost to yield an MLR increasing submodular transformed cost, for notational convenience, assume the measurement cost $m(e_i, u) = 0$. Choose $\phi = -\alpha C(\pi, L)$, where the scalar α will be specified below. From its definition in (30), straightforward computations yield that the transformed cost is

$$\underline{C}(\pi, 0) = e_2'\pi - \alpha d e_1'(I + A + \cdots A^{D_L - 1})'\pi$$
$$\underline{C}(\pi, u) = d e_1' \left((1-\alpha)I + \alpha A^{D_L}\right)' (I + A + \cdots + A^{D_u - 1})'\pi.$$

So clearly for $\alpha \geq 0$, $\underline{C}(e_1, 0) \leq \underline{C}(e_2, 0)$, and so $\underline{C}(\pi, 0)$ is $\geq_r$ increasing.

We now give conditions for $\underline{C}(\pi, u)$, for $u \in \{1, 2, \ldots, L\}$ to be MLR increasing in $\pi \in \Pi(X)$. By (A2), $(I + A + \cdots + A^{D_u - 1})'\pi$ is MLR increasing in $\pi \in \Pi(X)$. So for $\underline{C}(\pi, u)$ to be MLR increasing in π, it suffices to choose α so that the elements of $\left((1-\alpha)I + \alpha A^{D_L}\right) e_1$ are increasing. The elements of $\left((1-\alpha)I + \alpha A^{D_L}\right) e_1$ are decreasing if $\alpha \geq 1/(1 - A^{D_L}|_{11} + A^{D_L}|_{21})$. This is sufficient for $\underline{C}(\pi, u)$ to be MLR increasing in π for $u \in \{1, 2, \ldots, L\}$.

Next for the transformed cost $\underline{C}(\pi, u)$ to be submodular for $u \in \{1, 2, \ldots, L\}$, we require $\underline{C}(\pi, u+1) - \underline{C}(\pi, u)$ to be MLR decreasing in π. Straightforward computations yield for $u \in \{1, 2, \ldots, L\}$,

$$\underline{C}(\pi, u+1) - \underline{C}(\pi, u) = d e_1' \left((1-\alpha)I + \alpha A^{D_L}\right)' (A^{D_u} + A^{D_u + 1} + \cdots + A^{D_{u+1}})'\pi.$$

So for $\underline{C}(\pi, u)$ to be submodular, it suffices to choose α so that the elements of $\left((1-\alpha)I + \alpha A^{D_L}\right) e_1$ are decreasing, i.e., $\alpha \leq 1/(1 - A^{D_L}|_{11} + A^{D_L}|_{21})$. Therefore, choosing $\alpha = 1/(1 - A^{D_L}|_{11} + A^{D_L}|_{21})$ is sufficient for the transformed cost $\underline{C}(\pi, u)$ to be both MLR increasing for $u \in \{0, 1, \ldots, L\}$ and submodular for $u \in \{1, 2, \ldots, L\}$ on $[\Pi_{\mathcal{M}} - \mathcal{S}, \geq_r]$.

D) Proof of Theorem 4:

Statement 1: The proof of convexity of the stopping set $\mathcal{S}$ follows from arguments in [22]. We repeat this for completeness here. Pick any two belief states $\pi_1, \pi_2 \in \mathcal{S}$. To demonstrate convexity of $\mathcal{S}$, we need to show for any $\lambda \in [0, 1]$, $\lambda\pi_1 + (1 - \lambda)\pi_2 \in \mathcal{S}$. Since $V(\pi)$ is concave (by Theorem 12 above), it follows from (18) that

$$\begin{aligned} &V(\lambda\pi_1 + (1-\lambda)\pi_2) \\ &\geq \lambda V(\pi_1) + (1-\lambda)V(\pi_2) \\ &= \lambda Q(\pi_1, 0) + (1-\lambda)Q(\pi_2, 0) \ (\text{since } \pi_1, \pi_2 \in \mathcal{S}) \\ &= Q(\lambda\pi_1 + (1-\lambda)\pi_2, 0) \ (\text{since } Q(\pi, 0) \text{ is linear in } \pi) \\ &\geq V(\lambda\pi_1 + (1-\lambda)\pi_2) \ (\text{since } V(\pi) \text{ is the value function}). \end{aligned} \tag{50}$$

Thus, all the inequalities above are equalities, and $\lambda\pi_1 + (1 - \lambda)\pi_2 \in \mathcal{S}$.

Statement 2: Since the costs $C(\pi, u)$ are nonnegative, so is $V(\pi)$ in (18). So from (18), $C(\pi, 0) \leq C(\pi, u) \Longrightarrow Q(\pi, 0) \leq Q(\pi, u) \Longrightarrow \pi \in \mathcal{S}$. Therefore, $\underline{\mathcal{S}} \subset \mathcal{S}$.

Statement 3: Since $D_u < D_{u+1}$, then assuming (A2) and using Statement 6 of Theorem 9, it follows that $T(\pi, y, u + 1) \leq_r T(\pi, y, u)$, $\pi \in \Pi_{\mathcal{M}}$. By Theorem 12, $V(\pi)$ is MLR increasing in $\pi \in \Pi(X)$. Therefore, $V(T(\pi, y, u + 1)) \leq V(T(\pi, y, u))$ for $\pi \in \Pi_{\mathcal{M}}$. So for $\pi \in \Pi_{\mathcal{M}}$,

$$\sum_y V(T(\pi,y,u+1)\sigma(\pi,y,u+1) \leq \sum_y V(T(\pi,y,u)\sigma(\pi,y,u+1).$$

Since under (A3) $T(\pi, y, u)$ is MLR increasing in y (Statement 4 of Theorem 9) and $V(\pi)$ is MLR increasing in π, clearly $V(T(\pi, y, u))$ is increasing in y. Also from Statement 6 of Theorem 9, under (A2), $\sigma(\pi, \cdot, u + 1) \leq_s \sigma(\pi, \cdot, u)$, $\pi \in \Pi_{\mathcal{M}}$. So for $\pi \in \Pi_{\mathcal{M}}$,

$$\sum_y V(T(\pi, y, u)\, \sigma(\pi, y, u + 1) \leq \sum_y V(T(\pi, y, u)\, \sigma(\pi, y, u).$$

Therefore, $\sum_y V(T(\pi, y, u + 1)\sigma(\pi, y, u + 1) \leq \sum_y V(T(\pi, y, u)\sigma(\pi, y, u)$ for $\pi \in \Pi_{\mathcal{M}}$, which is equivalent to $Q(\pi, u + 1) - Q(\pi, u) \leq C(\pi, u + 1) - C(\pi, u)$. Then, [23, Lemma 2.2] implies that the minimizers of $Q(\pi, u + 1) - Q(\pi, u)$ are larger than that of $C(\pi, u + 1) - C(\pi, u)$. That is, $\mu^*(\pi) \geq \underline{\mu}(\pi)$ for $\pi \in \Pi_{\mathcal{M}}$.

Statement 4: By (A4), $C(\pi, u)$ is submodular on the poset $[\Pi(X), \geq_r]$. So using Theorem 11 it follows that $\underline{\mu}(\pi)$ is MLR increasing.

E) Proof of Theorem 5: Statements 1 and 2 follow directly from Theorem 4.

Statement 3: For $X = 2$, $\bar{C}(\pi, u) = \bar{C}_u'\pi$, where $\bar{C}_u = C_u + (I - A^{D_u})\bar{\phi} = m_u + (I + A + \cdots + A^{D_u - 1})e_1 + (I - A^{D_u})\bar{\phi}$. Clearly, the elements of $(I + A + \cdots + A^{D_u - 1})e_1$ are decreasing. So for $(\overline{\text{A1}})$ to hold, it suffices for $m_u + (I - A^{D_u})\bar{\phi}$ to be decreasing. Since $I - A^{D_u} = I - A + A(I - A) + \cdots + A^{D_u - 1}(I - A)$, and A satisfies (A2), it suffices to show that $m_u + (I - A)\bar{\phi}$ is decreasing. This is equivalent to $\bar{\phi}_1 - \bar{\phi}_2 \geq \frac{m(e_2,u) - m(e_1,u)}{A_{12} + A_{21}}$. For $\bar{C}_0 = e_2 + \bar{\phi}$ to satisfy $(\overline{\text{A1}})$, $\bar{\phi}_1 - \bar{\phi}_2 \geq 1$.

Proof of Lemma 2: As in Statement 3 above, for C_0 to satisfy $(\overline{\text{A1}})$ requires $\bar{\phi}_i$ to be decreasing with $\bar{\phi}_1 \geq \bar{\phi}_2 + 1$. For $\bar{C}_u$ to satisfy $(\overline{\text{A1}})$, it suffices to give conditions for $(I - A)\bar{\phi}$ to be decreasing elementwise. The difference between elements i and $i + 1$ is given by

$$\Delta_i \stackrel{\triangle}{=} \bar{\phi}_i - \bar{\phi}_{i+1} - \sum_{j=1}^{X} A_{ij}\bar{\phi}_j + \sum_{j=1}^{X} A_{i+1,j}\bar{\phi}_j.$$

The idea below is to lower bound each Δ_i by $\underline{\Delta}_i$ and show that $\underline{\Delta}_1 \leq \underline{\Delta}_i$ for all $i > 1$. This implies that if $\underline{\Delta}_1 \geq 0$, then so are $\underline{\Delta}_i$ and therefore $\Delta_i \geq 0$, implying that $(I - A)\bar{\phi}$ is decreasing. Assume $\bar{\phi}_i \geq 0$. Then, since A satisfies (A2) and $\bar{\phi}$ is decreasing, it follows that $\sum_{j=1}^{X} A_{i+1,j}\bar{\phi}_j \geq \sum_{j=1}^{X} A_{X,j}\bar{\phi}_j$. So clearly

$$\Delta_i \geq \underline{\Delta}_i \stackrel{\triangle}{=} \bar{\phi}_i - \bar{\phi}_{i+1} - \sum_{j=1}^{X} A_{ij}\bar{\phi}_j + \sum_{j=1}^{X} A_{X,j}\bar{\phi}_j.$$

Then, since A satisfies (A2) and $\bar{\phi}$ is decreasing, clearly, $-\sum_{j=1}^{X} A_{ij}\bar{\phi}_j \leq -\sum_{j=1}^{X} A_{i+1,j}\bar{\phi}_j$. Next construct $\bar{\phi}_i$ to be integer concave so that $\bar{\phi}_i - \bar{\phi}_{i+1} \leq \bar{\phi}_{i+1} - \bar{\phi}_{i+2}$. Then clearly,

$\underline{\Delta}_1 \leq \underline{\Delta}_2 \cdots \leq \underline{\Delta}_{X-1}$. The condition for $\underline{\Delta}_1 \geq 0$ is precisely that given in the lemma.

F) Proof of Theorem 6:

Part 1: We first prove that dominance of transition matrices $A \succeq \bar{A}$ [with respect to (36)] results in dominance of optimal costs, i.e., $V(\pi; A) \geq V(\pi; \bar{A})$. The proof is by induction. For $n = 0$, $V_n(\pi; A) \geq V_n(\pi; \bar{A}) = 0$ by the initialization of the value iteration algorithm (46).

Next, to prove the inductive step, assume that $V_n(\pi; A) \geq V_n(\pi; \bar{A})$ for $\pi \in \Pi(X)$. By Theorem 12(ii), under (A1), (A2), (A3), $V_n(\pi; A)$ and $V_n(\pi; \bar{A})$ are MLR increasing in $\pi \in \Pi(X)$. From Statement 7(a) of Theorem 9, it follows that $T(\pi, y, u; A) \geq_r T(\pi, y, u; \bar{A})$. This implies

$$V_n(T(\pi, y, u; A); A) \geq V_n(T(\pi, y, u; \bar{A}); A), \quad A \succeq \bar{A}.$$

Since $V_n(\pi; A) \geq V_n(\pi; \bar{A}) \ \forall \pi \in \Pi(X)$ by the induction assumption, clearly $V_n(T(\pi, y, u, \bar{A}); A) \geq V_n(T(\pi, y, u, \bar{A}); \bar{A})$. Therefore

$$\begin{aligned} V_n(T(\pi, y, u; A); A) &\geq V_n(T(\pi, y, u; \bar{A}); A) \\ &\geq V_n(T(\pi, y, u, \bar{A}); \bar{A}), \quad A \succeq \bar{A}. \end{aligned}$$

Under (A2) and (A3), Statement 4 of Theorem 9 states that $T(\pi, y, u; A); A)$ is MLR increasing in y. Therefore, $V_n(T(\pi, y, u; A); A)$ is increasing in y. Also from Statement 2 of Theorem 9, $\sigma(\pi, \cdot, u; A) \geq_s \sigma((\pi, \cdot, u; \bar{A})$ for $A \succeq \bar{A}$. Therefore,

$$\begin{aligned} &\sum_y V_n(T(\pi, y, u; A); A)\sigma(\pi, \cdot, u; A) \\ &\qquad \geq \sum_y V_n(T(\pi, y, u; \bar{A}); \bar{A})\sigma(\pi, \cdot, u; \bar{A}). \end{aligned} \quad (51)$$

Next, we claim that under (A1) and (A2), $A \succeq \bar{A}$ implies that $C(\pi, u; A) \geq C(\pi, u, ; \bar{A})$. This follows since $c(e_i, u)$ defined in (16) has increasing components by (A1) and $(A^l)'\pi \geq_r (\bar{A}^l)'\pi$ (Statement 5(b), Theorem 9). Therefore, $c_u'(A^l)'\pi \geq c_u'(\bar{A}^l)'\pi$ implying that $C(\pi, u; A) \geq C(\pi, u, \bar{A})$. This together with (51) implies

$$\begin{aligned} &C(\pi, u; A) + \sum_y V_n(T(\pi, y, u; A); A)\sigma(\pi, \cdot, u; A) \\ &\quad \geq C(\pi, u, \bar{A}) + \sum_y V_n(T(\pi, y, u; \bar{A}); \bar{A})\sigma(\pi, \cdot, u; \bar{A}). \end{aligned}$$

Minimizing both sides with respect to action u yields $V_{n+1}(\pi; A) \geq V_{n+1}(\pi; \bar{A})$ and concludes the induction argument.

Part 2: Next we show that dominance of observation distributions $f \succeq_B \bar{f}$ (with respect to the order (37)) results in dominance of the optimal costs, namely $V(\pi; f) \geq V(\pi, \bar{f})$. Let $T(\pi, y, u)$ and $\bar{T}(\pi, y, u)$ denote the Bayesian filter update (10) with observation f and $\bar{f}$, respectively, and let $\sigma(\pi, y, u)$ and $\bar{\sigma}(\pi, y, u)$ denote the corresponding normalization measures.

Then, for $a \in \mathbb{Y}$ and $\sigma(\pi, a, u) = \sum_{y \in \mathbb{Y}} \bar{\sigma}(\pi, y, u)P(a|y)$,

$$T(\pi, a, u) = \sum_{y \in \mathbb{Y}} \bar{T}(\pi, y, u)\frac{\bar{\sigma}(\pi, y, u)}{\sigma(\pi, a, u)}P(a|y).$$

Therefore, $\frac{\bar{\sigma}(\pi, y, u)}{\sigma(\pi, y, a)}P(a|y)$ is a probability measure wrt y. Since from Theorem 12, $V_n(\cdot)$ is concave for $\pi \in \Pi(X)$, using Jensen's inequality, it follows that

$$\begin{aligned} V_n(T(\pi, a, u); \bar{f}) &= V_n\left(\sum_{y \in \mathbb{Y}} \bar{T}(\pi, y, u)\frac{\bar{\sigma}(\pi, y, u)}{\sigma(\pi, a, u)}P(a|y); \bar{f}\right) \\ &\geq \sum_{y \in \mathbb{Y}} V_n(\bar{T}(\pi, y, u); \bar{f})\frac{\bar{\sigma}(\pi, y, u)}{\sigma(\pi, a, u)}P(a|y). \end{aligned}$$

This implies

$$\begin{aligned} &\sum_a V_n(T(\pi, a, u); \bar{f})\sigma(\pi, a, u) \\ &\qquad \geq \sum_y V_n(\bar{T}(\pi, y, u); \bar{f})\bar{\sigma}(\pi, y, u). \end{aligned} \quad (52)$$

With the above inequality, the proof of the theorem follows by mathematical induction using the value iteration algorithm (46). Assume $V_n(\pi; f) \geq V_n(\pi; \bar{f})$ for $\pi \in \Pi(X)$. Then

$$\begin{aligned} &C(\pi, u) + \sum_a V_n(T(\pi, a, u); f)\sigma(\pi, a, u) \\ &\geq C(\pi, u) + \sum_a V_n(T(\pi, a, u); \bar{f})\sigma(\pi, a, u) \\ &\geq C(\pi, u) + \sum_y V_n(\bar{T}(\pi, y, u); \bar{f})\bar{\sigma}(\pi, y, u) \end{aligned}$$

where the second inequality follows from (52). Thus, $V_{n+1}(\pi; f) \geq V_{n+1}(\pi; \bar{f})$. This completes the induction step. Since value iteration algorithm (46) converges uniformly, $V(\pi; f) \geq V(\pi; \bar{f})$, thus proving the theorem.

G) Proof of Theorem 7: Recall the aim is to prove (39) and (40). Most of our efforts below are to prove (39). We will prove that (39) holds for any strategy μ. That is, for any strategy μ:

$$\sup_{\pi \in \Pi(X)} |J_\mu(\pi; \theta) - J_\mu(\pi; \bar{\theta})| \leq K\|\theta - \bar{\theta}\|. \quad (53)$$

We start this appendix with the proof of (40) since it follows straightforwardly from (53). To prove (40), note that trivially

$$\begin{aligned} J_{\mu^*(\bar{\theta})}(\pi, \theta) &\leq J_{\mu^*(\bar{\theta})}(\pi, \bar{\theta}) + \sup_\pi |J_{\mu^*(\bar{\theta})}(\pi, \theta) - J_{\mu^*(\bar{\theta})}(\pi, \bar{\theta})| \\ J_{\mu^*(\theta)}(\pi, \bar{\theta}) &\leq J_{\mu^*(\theta)}(\pi, \theta) + \sup_\pi |J_{\mu^*(\theta)}(\pi, \theta) - J_{\mu^*(\theta)}(\pi, \bar{\theta})|. \end{aligned}$$

Also by definition $J_{\mu^*(\bar{\theta})}(\pi, \bar{\theta}) \leq J_{\mu^*(\theta)}(\pi, \bar{\theta})$ since $\mu^*(\bar{\theta})$ is the optimal strategy for model $\bar{\theta}$. Therefore,

$$\begin{aligned} J_{\mu^*(\bar{\theta})}(\pi, \theta) &\leq J_{\mu^*(\theta)}(\pi, \theta) + \sup_\pi |J_{\mu^*(\bar{\theta})}(\pi, \theta) - J_{\mu^*(\bar{\theta})}(\pi, \bar{\theta})| \\ &\quad + \sup_\pi |J_{\mu^*(\theta)}(\pi, \theta) - J_{\mu^*(\theta)}(\pi, \bar{\theta})| \\ &\leq J_{\mu^*(\theta)}(\pi, \theta) + 2\sup_\mu \sup_\pi |J_{\mu(\bar{\theta})}(\pi, \theta) - J_{\mu(\bar{\theta})}(\pi, \bar{\theta})|. \end{aligned}$$

Then, from (53), clearly (40) follows.

Proof of (39): We now present the proof of (53) and thus (39). Define the set of belief states $\bar{\mathcal{S}} = \cap_u\{\pi : (C_u - C_0)'\pi \geq$

$0\}$. Clearly, $\bar{\mathcal{S}} \subseteq \mathcal{S}$. Let us characterize the set of observations such that the Bayesian filter update $T(\pi, y, u; \theta)$ lies in $\bar{\mathcal{S}}$ for any action u. Accordingly, define

$$\mathcal{R}_{\pi;\theta} = \{y : (C_{\bar{u}} - C_0)' T(\pi, y, u; \theta) \geq 0, \tag{54}$$

for $u, \bar{u} \in \{1, 2, \ldots, L\}\}$, $y^*_{\pi;\theta} = \inf\{y : y \in \mathcal{R}^c_{\pi;\theta}\}$. Here, $\mathcal{R}^c_{\pi;\theta}$ denotes the complement of set $\mathcal{R}_{\pi;\theta}$.

Lemma 5: Under (A2),(A3), and (A4), the following hold for $\mathcal{R}_{\pi;\theta}$ and $y^*_{\pi;\theta}$ defined in (54):

i) $\mathcal{R}^c_{\pi;\theta} = \{y : y \geq y^*_{\pi;\theta}\}$.

ii) $\pi \geq_r \bar{\pi} \implies \mathcal{R}_{\pi;\theta} \subset \mathcal{R}_{\bar{\pi};\theta}$.

iii) $\pi \geq_r \bar{\pi} \implies y^*_{\pi;\theta} < y^*_{\bar{\pi};\theta}$.

Proof: The first assertion says that the set of observations for continuing is the set $\{y : y \geq y^*_{\pi;\theta}\}$. By (A4), $C_{\bar{u}} - C_0$ has decreasing elements. Since $T(\pi, y, u; \theta)$ is MLR increasing in y, clearly $(C_{\bar{u}} - C_0)' T(\pi, y, u; \theta)$ is decreasing in y. Therefore, there exists a $y^*_{\pi;\theta}$ such that $y \geq y^*_{\pi;\theta}$ implies $T(\pi, y, u; \theta) \in \mathcal{R}^c_{\pi;\theta}$. This proves the first statement. By (A4), $C_{\bar{u}} - C_0$ has decreasing elements. By (A2), (A3), $T(\pi, y, u; \theta)$ is MLR increasing in π. Therefore, $(C_{\bar{u}} - C_0)' T(\pi, y, u; \theta) \geq (C_{\bar{u}} - C_0)' T(e_X, y, u; \theta)$ which implies $\mathcal{R}_{\pi;\theta} \subset \mathcal{R}_{\bar{\pi};\theta}$. Statement (i) states that $T(\pi, y, u; \theta)$ is MLR increasing in y; statement (ii) states that $\mathcal{R}_{\pi;\theta} \subset \mathcal{R}_{\bar{\pi};\theta}$. Combining these yields $y^*_{\pi;\theta} \leq y^*_{\bar{\pi};\theta}$. ■

From (16), the total cost incurred by applying strategy $\mu(\pi)$ to model θ satisfies at time n

$$\begin{aligned} &J^{(n)}_\mu(\pi; \theta) \\ &= C'_{\mu(\pi)}\pi + \sum_{y \in \mathbb{Y}} J^{(n-1)}_\mu(T(\pi, y, \mu(\pi); \theta)\sigma(\pi, y, \mu(\pi); \theta) \\ &= C'_{\mu(\pi)}\pi + \sum_{y \in \mathcal{R}^c_{\pi;\theta}} J^{(n-1)}_\mu(T(\pi, y, \mu(\pi); \theta)\sigma(\pi, y, \mu(\pi); \theta) \end{aligned}$$

since for $y \in \mathcal{R}_{\pi;\theta}$, $T(\pi, y, \mu(\pi); \theta) \in \bar{\mathcal{S}}$ and so $V(T(\pi, y, \mu(\pi); \theta)) = 0$.

Therefore, the absolute difference in total costs for models $\theta, \bar{\theta}$ satisfies

$$\begin{aligned} &|J^{(n)}_\mu(\pi; \theta) - J^{(n)}_\mu(\pi; \bar{\theta})| \\ &\leq \sum_{y \in \mathcal{R}^c_{\pi;\theta} \cup \mathcal{R}^c_{\pi;\bar{\theta}}} \sigma(\pi, y, \mu(\pi); \theta) \\ &\quad \left| J^{(n-1)}_\mu(T(\pi, y, \mu(\pi); \theta)) - J^{(n-1)}_\mu(T(\pi, y, \mu(\pi); \bar{\theta})) \right| \\ &\quad + \sum_{y \in \mathcal{R}^c_{\pi;\theta} \cup \mathcal{R}^c_{\pi;\bar{\theta}}} J^{(n-1)}_\mu(T(\pi, y, \mu(\pi); \bar{\theta})) \\ &\qquad \left| \sigma(\pi, y, \mu(\pi); \theta) - \sigma(\pi, y, \mu(\pi); \bar{\theta}) \right| \\ &\leq \sup_{\pi \in \Pi(X)} |J^{(n-1)}_\mu(\pi; \theta) - J^{(n-1)}_\mu(\pi; \bar{\theta})| \sum_{y \in \mathcal{R}^c_{\pi;\theta} \cup \mathcal{R}^c_{\pi;\bar{\theta}}} \sigma(\pi, y, \mu(\pi); \theta) \\ &\quad + \sup_{\pi \in \Pi(X)} J^{(n-1)}_\mu(\pi; \bar{\theta}) \sum_{y \in \mathbb{Y}} \left| \sigma(\pi, y, \mu(\pi); \theta) - \sigma(\pi, y, \mu(\pi); \bar{\theta}) \right|. \end{aligned} \tag{55}$$

We will upper bound the various terms on the RHS of (55). Statement (i) of Lemma 5 yields $\mathcal{R}^c_{\pi;\theta} \cup \mathcal{R}^c_{\pi;\bar{\theta}} = \{y \geq y^*_{\pi;\theta,\bar{\theta}}\}$ where $y^*_{\pi;\theta,\bar{\theta}} = \min(y^*_{\pi;\theta}, y^*_{\pi;\bar{\theta}})$. Next Statement (iii) of Lemma 5 yields $y^*_{e_X;\theta,\bar{\theta}} \leq y^*_{\pi;\theta,\bar{\theta}}$. Therefore,

$$\begin{aligned} &\sup_{\pi \in \Pi(X)} |J^{(n-1)}_\mu(\pi; \theta) - J^{(n-1)}_\mu(\pi; \bar{\theta})| \sum_{y \in \mathcal{R}^c_{\pi;\theta} \cup \mathcal{R}^c_{\pi;\bar{\theta}}} \sigma(\pi, y, \mu(\pi); \theta) \\ &\leq \sup_{\pi \in \Pi(X)} |J^{(n-1)}_\mu(\pi; \theta) - J^{(n-1)}_\mu(\pi; \bar{\theta})| \max_u \sum_{y \geq y^*_{e_X;\theta,\bar{\theta}}} \sigma(\pi, y, u; \theta) \\ &\leq \sup_{\pi \in \Pi(X)} |J^{(n-1)}_\mu(\pi; \theta) - J^{(n-1)}_\mu(\pi; \bar{\theta})| \max_u \sum_{y \geq y^*_{e_X;\theta,\bar{\theta}}} \sigma(e_X, y, u; \theta) \end{aligned}$$

where the last line follows since $e_X \geq_s \pi$, and so Statement 2 of Theorem 9 implies $\sigma(\pi, \cdot, u; \theta) \leq_s \sigma(e_X, \cdot, u; \theta)$. Also evaluating $\sigma(\pi, y, \mu(\pi); \theta) = \mathbf{1}'_X f_y (A')^{\mu(\pi)} \pi$ defined in (10) yields

$$\begin{aligned} &\sum_{y \in \mathbb{Y}} \left| \sigma(\pi, y, \mu(\pi); \theta) - \sigma(\pi, y, \mu(\pi); \bar{\theta}) \right| \\ &\qquad \leq \max_u \sum_y \sum_i \sum_j \left| f_{jy} A^u|_{ij} - \bar{f}_{jy} \bar{A}^u|_{ij} \right| \pi(i) \\ &\qquad \leq \max_u \max_i \sum_y \sum_j \left| f_{jy} A^u|_{ij} - \bar{f}_{jy} \bar{A}^u|_{ij} \right|. \end{aligned} \tag{56}$$

Finally, $\sup_{\pi \in \Pi(X)} J^{(n-1)}_\mu(\pi; \bar{\theta}) \leq \max_{i \in \mathbb{X}} C(e_i, 0)$. Using these bounds in (55) yields

$$\begin{aligned} &\sup_{\pi \in \Pi(X)} |J^{(n)}_\mu(\pi; \theta) - J^{(n)}_\mu(\pi; \bar{\theta})| \\ &\qquad \leq \rho_{\theta,\bar{\theta}} \sup_{\pi \in \Pi(X)} |J^{(n-1)}_\mu(\pi; \theta) - J^{(n-1)}_\mu(\pi; \bar{\theta})| \\ &\qquad\qquad + \max_{i \in \mathbb{X}} C(e_i, 0) \|\theta - \bar{\theta}\| \end{aligned} \tag{57}$$

where $\rho_{\theta,\bar{\theta}} = \max_u \sum_{y \geq y^*_{e_X;\theta,\bar{\theta}}} \sigma(e_X, y, u; \theta)$ and $\|\theta - \bar{\theta}\|$ is given by (56). Since $\max_u \sum_{y \in \mathbb{Y}} \sigma(e_X, y, u; \theta) = 1$, then (A5) implies $\rho_{\theta,\bar{\theta}} = \max_u \sum_{y \geq y^*_{e_X;\theta,\bar{\theta}}} \sigma(e_X, y, u; \theta) < 1$. Then, starting with $J^{(0)}_\mu(\pi; \theta) = J^{(0)}_\mu(\pi; \bar{\theta}) = 0$, unraveling (57) yields (53). In particular, choosing $\mu = \mu^*(\theta)$ yields (39).

Proof of (42): When θ and $\bar{\theta}$ have identical transition matrices, then (56) becomes

$$\max_u \max_i \sum_j A^u|_{ij} \sum_y |f_{jy} - \bar{f}_{jy}|.$$

From Pinsker's inequality [10], the total variation norm is bounded by Kullback–Leibler distance D defined in (42) as

$$\sum_y |f_{jy} - \bar{f}_{jy}| \leq \sqrt{2\, D(f_j \| \bar{f}_j)}.$$

H) Proof of Theorem 9: We quote the following result from [14], which adapted to our notation reads

Theorem 13 (see [14, Lemma 8.2, p. 382]): Suppose ϕ_i is increasing for $i \in X$ and nonnegative. Then, for arbitrary vectors $p, q \in \mathbb{R}^X$, $\phi' p \geq \phi' q$ iff $\sum_{j \geq \bar{j}} p_j \geq \sum_{j \geq \bar{j}} q_j$ for all $\bar{j} \in \mathbb{X}$ ■

The above theorem is similar to Statement (ii) of Theorem 8 with some important differences. Unlike Theorem 8, p and q

need not be probability measures. On the other hand, Theorem 8 does not require ϕ to be nonnegative.

Proof of Theorem 9: Statements 1, 2, and 4 of the theorem are proved in [23].

Statement 3(a): Proved in Lemma 4 in Appendix B for e_X. The proof for e_1 is similar since matrix $\begin{bmatrix} e_1' \\ A_{1,1:X} \end{bmatrix} A = \begin{bmatrix} e_1' A \\ e_1' A^2 \end{bmatrix}$ is TP2.

Statement 3(b): For $X = 2$, Theorem 9(3a) implies that $\sum_{j \geq q} A^{D_{u+1}}|_{ij} - A^{D_u}|_{ij}$ is decreasing in i. Then, for $\pi \geq_s \bar{\pi}$,

$$\sum_{j \geq q} \sum_{i \in \mathbb{X}} \left(A^{D_{u+1}}|_{ij} - A^{D_u}|_{ij}\right) \pi(i) \leq \sum_{j \geq q} \sum_{i \in \mathbb{X}} \left(A^{D_{u+1}}|_{ij} - A^{D_u}|_{ij}\right) \bar{\pi}(i).$$

Also (A3) implies that $\phi(j) \stackrel{\triangle}{=} \sum_{y \geq \bar{y}} f_{jy}$ is increasing in j. Then, applying Theorem 13 yields

$$\sum_{j} \sum_{y \geq \bar{y}} f_{jy} \sum_{i} \left(A^{D_{u+1}}|_{ij} - A^{D_u}|_{ij}\right) \pi(i) \leq \sum_{j} \sum_{y \geq \bar{y}} f_{jy} \sum_{i} \left(A^{D_{u+1}}|_{ij} - A^{D_u}|_{ij}\right) \bar{\pi}(i).$$

Statement 5(a): The proof is as follows: By definition, $A'\pi \geq_r \bar{A}'\pi$ is equivalent to

$$\sum_{i \in \mathbb{X}} \sum_{m \in \mathbb{X}} \left(A_{ij} \bar{A}_{m,j+1} - \bar{A}_{ij} A_{m,j+1}\right) \pi_i \pi_m \leq 0.$$

Thus, (36) is a sufficient condition for $A'\pi \geq_r \bar{A}'\pi$.

Statement 5(b): Since $A \succeq \bar{A}$ implies $A'\pi \geq_r \bar{A}'\pi$, it follows from (A2) that $A'A'\pi \geq_r A'\bar{A}'\pi$. Also Statement 5(a) implies $A'\bar{A}'\pi \geq_r \bar{A}'\bar{A}'\pi$. Since the MLR order is transitive, these inequalities imply $A'A'\pi \geq_r \bar{A}'\bar{A}'\pi$. Continuing similarly, it follows that for any positive integer l, $(A^l)'\pi \geq_r (\bar{A}^l)'\pi$.

Statement 6: Recall from (28) and (33) that for $\pi \in \Pi_{\mathcal{M}}$, $A'\pi \geq_r A^{2'}\pi$. Then, applying Statement 5(b) yields that for $\pi \in \Pi_{\mathcal{M}}$, $T(\pi, y, u) \geq_r T(\pi, y.u + 1)$. The dominance of $\sigma(\pi, \cdot, u)$ follows from the proof of Statement 7(b) below.

Statement 7(a): This follows trivially since Bayes' rule preserves MLR dominance. That is, $\pi \geq_r \bar{\pi}$ implies $\frac{f_y \pi}{\mathbf{1}_X' f_y \pi} \geq_r \frac{f_y \bar{\pi}}{\mathbf{1}_X' f_y \bar{\pi}}$. Since by Statement 4(a), $A \succeq \bar{A}$ implies $A'\pi \geq_r \bar{A}'\pi$, applying the Bayes rule preservation of MLR dominance proves the result.

Statement 7(b): Since $A \succeq \bar{A}$ implies $A'\pi \geq_r \bar{A}'\pi$, it follows that $A'\pi \geq_s \bar{A}'\pi$. Next, (A3) implies that $\sum_{y \geq q} f_{iy}$ is increasing in i. Therefore, $\sum_{i \in \mathbb{X}} \sum_{y \geq q} f_{iy}[A'\pi](i) \geq \sum_{i \in \mathbb{X}} \sum_{y \geq q} f_{iy}[\bar{A}'\pi](i)$.

I) Proof of Theorem 10: The proofs of Statements 1, 2, and 4 are identical to the proof of Theorem 4 in Appendix D.

Statement 3: For notational convenience, define the belief state update

$$T(\pi, y, u, u+1) = \frac{f_{yu} A^{D_{u+1}} \pi}{\sigma(\pi, y, u, u+1)}$$

where $\sigma(\pi, y, u, u+1) = \mathbf{1}_X' f_{yu} A^{D_{u+1}} \pi$. Using an identical proof to Theorem 6 that yields (52), since $f_u \succeq_B f_{u+1}$, it follows that

$$\sum_{y} V(T(\pi, y, u+1)) \sigma(\pi, y, u+1) \leq \sum_{y} V(T(\pi, y, u, u+1)) \sigma(\pi, y, u, u+1). \quad (58)$$

Next, using (A2) and Statement 6 of Theorem 9, it follows that for $\pi \in \Pi_{\mathcal{M}}$,

$$T(\pi, y, u, u+1) \leq_r T(\pi, y, u) \text{ and } \sigma(\pi, \cdot, u, u+1) \leq_s \sigma(\pi, \cdot, u).$$

Also from Theorem 12(ii), under (A1), (A2), and (A3), $V(\pi)$ is MLR increasing in $\pi \in \Pi(X)$. So

$$\begin{aligned} &\sum_{y} V(T(\pi, y, u, u+1)) \sigma(\pi, y, u, u+1) \\ &\leq \sum_{y} V(T(\pi, y, u, u+1)) \sigma(\pi, y, u) \\ &\leq \sum_{y} V(T(\pi, y, u)) \sigma(\pi, y, u), \quad \pi \in \Pi_{\mathcal{M}}. \end{aligned} \quad (59)$$

From (58) and (59), it follows that for $\pi \in \Pi_{\mathcal{M}}$

$$\sum_{y} V(T(\pi, y, u+1)) \sigma(\pi, y, u+1) \leq \sum_{y} V(T(\pi, y, u)) \sigma(\pi, y, u)$$

which is equivalent to $Q(\pi, u+1) - Q(\pi, u) \leq C(\pi, u+1) - C(\pi, u)$, $\pi \in \Pi_{\mathcal{M}}$. Then, [23, Lemma 2.2] implies that the minimizers of $Q(\pi, u+1) - Q(\pi, u)$ are larger than that of $C(\pi, u+1) - C(\pi, u)$. That is, $\mu^*(\pi) \geq \underline{\mu}(\pi)$ for $\pi \in \Pi_{\mathcal{M}}$.

Acknowledgment

This paper is dedicated to the memory of my father E. V. Krishnamurthy who passed away in October 2012.

"Even death is not to be feared by one who has lived wisely."

References

[1] S. C. Albright, "Structural results for partially observed Markov decision processes," *Oper. Res.*, vol. 27, no. 5, pp. 1041–1053, Sep./Oct. 1979.

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

[3] S. Athey, "Monotone comparative statics under uncertainty," *Quart. J. Econ.*, vol. 117, no. 1, pp. 187–223, 2002.

[4] T. Banerjee and V. Veeravalli, "Data-efficient quickest change detection with on-off observation control," *Seq. Anal.*, vol. 31, pp. 40–77, 2012.

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

[6] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House, 1999.

[7] D. Blackwell, "Discounted dynamic programming," *Anna. Math. Statist.*, vol. 36, pp. 226–235, 1965.

[8] J. Boger, P. Poupart, and J. Hoey, "A decision-theoretic approach to task assistance for persons with dementia," in *Proc. Int. Joint Conf. Artif. Intell.*, 2005, pp. 1293–1299.

[9] A. R. Cassandra, "Exact and approximate algorithms for partially observed Markov decision process," Ph.D. dissertation, Brown Univ., Providence, RI, USA, 1998.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 2006.

[11] S. Dayanik and C. Goulding, "Sequential detection and identification of an unobservable change in the distribution of a Markov-modulated random sequence," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3323–3345, Jul. 2009.

[12] G. Fellouris and G. Moustakides, "Decentralized sequential change detection," ArXiv Oct. 2012.

[13] O. Hernández-Lerma and J. Bernard Laserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. New York, NY, USA: Springer-Verlag, 1996.

[14] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research*. New York, NY, USA: McGraw-Hill, 1984, vol. 2.

[15] S. Karlin and Y. Rinott, "Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions," *J. Multivariate Anal.*, vol. 10, no. 4, pp. 467–498, Dec. 1980.

[16] M. Kijima, *Markov Processes for Stochastic Modelling*. London, U.K.: Chapman & Hall, 1997.

[17] V. Krishnamurthy, "Algorithms for optimal scheduling and management of hidden Markov model sensors," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1382–1397, Jun. 2002.

[18] 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.

[19] V. Krishnamurthy, "Quickest detection POMDPs with social learning: Interaction of local and global decision makers," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug. 2012.

[20] V. Krishnamurthy, R. Bitmead, M. Gevers, and E. Miehling, "Sequential detection with mutual information stopping cost: Application in GMTI radar," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 700–714, Feb. 2012.

[21] V. Krishnamurthy and D. Djonin, "Structured threshold policies for dynamic sensor scheduling–A partially observed Markov decision process approach," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4938–4957, Oct. 2007.

[22] 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.

[23] W. S. Lovejoy, "Some monotonicity results for partially observed Markov decision processes," *Oper. Res.*, vol. 35, no. 5, pp. 736–743, Sep./Oct. 1987.

[24] A. Muller and D. Stoyan, *Comparison Methods for Stochastic Models and Risk*. New York, NY, USA: Wiley, 2002.

[25] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and their Applications*. New York, NY, USA: Marcel Dekker, 1989.

[26] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, no. 3, pp. 441–450, 1987.

[27] M. Pollack, L. Brown, and D. Colbry, "Autominder: An intelligent cognitive orthotic system for people with memory impairment," *Robot. Autonom. Syst.*, vol. 44, pp. 273–282, 2003.

[28] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[29] U. Rieder, "Structural results for partially observed control models," *Methods Models Oper. Res.*, vol. 35, no. 6, pp. 473–490, 1991.

[30] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probabil. Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[31] A. N. Shiryayev, *Optimal Stopping Rules*. New York, NY, USA: Springer-Verlag, 1978.

[32] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Probabil. Appl.*, vol. 49, no. 3, pp. 458–497, 2005.

[33] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ, USA: Princeton Univ. Press, 1998.

[34] W. Whitt, "Multivariate monotone likelihood ratio and uniform conditional stochastic order," *J. Appl. Probabil.*, vol. 19, pp. 695–701, 1982.

[35] Y. Yilmaz, G. Moustakides, and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4509–4524, Sep. 2012.

Vikram Krishnamurthy (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 game theory, stochastic control in sensor networks, and modeling of biological ion channels and biosensors. Dr. Krishnamurthy served as Editor in Chief of IEEE Journal Selected Topics in Signal Processing and as Distinguished lecturer for the IEEE signal processing society. He has also served as associate editor for several journals including IEEE Transactions Automatic Control and IEEE Transactions on Signal Processing.