Optimal Stationary Binary Quantizer for Decentralized Quickest Change Detection in Hidden Markov Models

Cheng-Der Fuh
Graduate Institute of Statistics
National Central University
and Institute of Statistical Science
Academia Sinica
Taiwan, R.O.C.
Email: stcheng@stat.sinica.edu.tw

Yajun Mei
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA, U.S.A.
Email: ymei@isye.gatech.edu

Abstract—The decentralized quickest change detection problem is studied in sensor networks, where a set of sensors receive observations from a hidden Markov model $X$ and send sensor messages to a central processor, called the fusion center, which makes a final decision when observations are stopped. It is assumed that the parameter $\theta$ in the hidden Markov model for $X$ changes from $\theta_0$ to $\theta_1$ at some unknown time. The problem is to determine the policies at the sensor and fusion center levels to jointly optimize the detection delay subject to the average run length (ARL) to false alarm constraint. In this article, a CUSUM-type fusion rule with stationary binary sensor messages is studied and a simple method for choosing the optimal local sensor thresholds is introduced. Further research is also given.

Keywords: Asymptotic optimality, CUSUM, multi-sensor, hidden Markov models, quantization, sensor networks, sequential detection.

I. INTRODUCTION

Sensor networks have many important applications including environment monitoring, target detection and tracking, and security and surveillance systems [1], [14]. There are various possible configurations for sensor networks. Figure 1 illustrates the general setting of a widely used configuration [12]. In such a sensor network, at time $n$, each of a set of $K$ sensors receives an observation $Y_{k,n}$, and then sends a sensor message $U_{k,n}$ to a central processor, called the fusion center, which makes a final decision when observations are stopped. If necessary, the fusion center can also send feedback to the local sensors.

Since communication costs are usually dominant in sensor networks, especially for wireless sensor networks, it is crucial to reduce communication costs of sensor networks. A standard mathematical formulation to accomplish this task is to require that the sensor messages belong to a finite alphabet (possibly binary). While this approach exhibits information loss as compared to the centralized setting of sending all raw data to the fusion center, it can often provide a formulation that is better suited for practical applications, due to the reduction in channel bandwidth and a decrease in the computing complexity at the fusion center.

In sensor networks, the information is accumulated over time at the sensor and fusion center levels, and thus there are many important dynamical decision problems in sensor networks. One such problem is the decentralized quickest change detection problem, in which it is assumed that the distributions of sensor observations changes abruptly at some unknown time. The goal is to detect the change as soon as possible over all policies at the sensor and fusion center levels, under a restriction on the frequency of false alarms.

Early work in decentralized quickest change detection problems considers the scenario when the sensor observations are independent from sensor to sensor and over time, see, for example, [2], [6], [9], [10], [13] and the references therein. In certain applications of interest, however, the independence...
assumption may be inappropriate, and it is more reasonable to model the underlying physical process as a Markov chain and corresponding sensor observations as hidden Markov chains. One example occurs if one is interested in modeling different operating modes of a target [4].

In this article, we consider decentralized quickest change detection problems in hidden Markov models (HMM), where all sensor observations depend on an underlying Markov chain whose parameter may change at some unknown time. Motivated by the exact and asymptotic optimality results of Page's CUSUM procedure [8] in different contexts (Moustakides [7], Fuh [3] and Mei [6]), here we focus on examining a CUSUM-type fusion rule with stationary binary sensor messages, due to its simplicity and the potential of asymptotic optimality.

In Section II, the decentralized quickest change detection problems in HMM is stated precisely, as are the performance criteria of detection schemes. In Section III, some preliminaries involving the centralized quickest change detection in HMM are presented, particularly the Kullback-Leibler information number in HMM. In Sections IV and V, we focus on the stationary sensor message functions, and propose asymptotically optimal fusion center rule and a simple procedure for choosing the optimal thresholds at local sensors based upon an asymptotic performance measure. Section VI considers a specific example, and Section VII lists some further questions to be studied in the near future. Section VIII includes some concluding remarks.

II. PROBLEM FORMULATION

Let \( X = \{X_n, n \geq 0\} \) be a discrete time Markov chain on a finite state space \( D = \{1, \cdots, d\} \), with transition probability matrix

\[
P_\theta = \begin{bmatrix}
p_\theta(1, 1) & \cdots & p_\theta(1, d) \\
\vdots & \ddots & \vdots \\
p_\theta(d, 1) & \cdots & p_\theta(d, d)
\end{bmatrix},
\]

and stationary distribution

\[
\pi_\theta = (\pi_\theta(1), \cdots, \pi_\theta(d))^t,
\]

where \( t \) denotes transpose. Suppose there are \( K \) sensors in a system. At time \( n \), each sensor \( S_k \) takes a sensor observation \( Y_{k, n} \) which can be modeled as

\[
Y_{k, n} = h_k(X_n) + \varepsilon_{k, n},
\]

where \( h_k(\cdot) \) is a real-valued (non-constant) function on the state space of the Markov chain \( X \) and the sensor noises \( \varepsilon_{k, n} \)'s are \( N(0, \sigma_k^2) \) random variables that are independent over time and from sensor to sensor. Here \( h_k(\cdot) \) or \( \sigma_k^2 \) may or may not depend on \( \theta \), but we will suppress the index \( \theta \) to simplify notation. Moreover, at time \( n \), each sensor \( S_k \) quantizes the raw sensor observation \( Y_{k, n} \) and sends the quantized data as a sensor message \( U_{k, n} \) to the fusion center, due to data compression and limitations of channel bandwidth. For simplicity, here we assume

\[
U_{k, n} = \phi_{k, n}(Y_{k, n}) = \begin{cases}
1, & \text{if } Y_{k, n} \geq \lambda_k \\
0, & \text{if } Y_{k, n} < \lambda_k
\end{cases},
\]

where the thresholds \( \lambda_k \)'s are constants chosen by the statisticians. The fusion center then uses the stream of messages \( U_{k, n} \)'s from the sensors as inputs to make a final decision.

In quickest change detection problems in HMM, it is assumed that at some unknown (possibly \( \infty \)) time \( \nu \), the parameter \( \theta \) in the hidden Markov model for \( X \) changes from \( \theta_0 \) to \( \theta_1 \), which of course induces a change in the distributions of sensor observations \( Y_{k, n} \)'s. The goal is to detect the true change as soon as possible over all possible stationary thresholds \( \lambda_k \)'s at the sensors and over all possible decision rules at the fusion center.

Denote by \( P_\nu \) and \( E_\nu \), the probability measure and expectation when the change occurs at time \( \nu \), and denote the same by \( P_\infty \) and \( E_\infty \) when there is no change. A detection scheme at the fusion center is defined as a stopping time \( \tau \) with respect to \( \{U_n\}_{n \geq 1} \). The interpretation of \( \tau \) is that, when \( \tau = n \), the fusion center raises an alarm at time \( n \) to declare that a change occurred somewhere in the first \( n \) time steps. The performance of \( \tau \) is evaluated by two criteria: the long and short Average Run Lengths (ARL). The long ARL is defined by \( E_\infty(\tau) \), which is also often called the ARL to false alarm. The short ARL can be defined by the worst case detection delay in Lorden [5]:

\[
E_1(\tau) = \sup_{\nu \geq 1} \left( \text{ess sup} E_\nu [(\tau - \nu + 1)^+ | \mathcal{F}_{n-1}] \right),
\]

where \( \mathcal{F}_{n-1} = \mathcal{F}(U_1, \cdots, U_{n-1}) \) denotes all sensor messages available at the fusion center before time \( n \), and \( U_i = (U_{1, i}, \cdots, U_{K, i}) \) denotes the sensor messages at time \( i \). Note that \( E_1(\tau) \) can be replaced by the so-called average detection delay \( \sup_{\nu \geq 1} E_\nu (\tau - \nu | \tau \geq \nu) \) in our results, as they are asymptotically equivalent.

A simple but useful mathematical formulation of our problem is to develop a protocol to determine the thresholds \( \lambda_k \)'s at the sensors and to find a stopping time \( \tau \) at the fusion center that minimizes \( E_1(\tau) \) subject to the constraint

\[
E_\infty(\tau) \geq \gamma,
\]

where \( \gamma \) is a given, fixed lower bound.

It is worth pointing out that here we only consider a special form of stationary binary quantizer defined in (3). A natural question one may ask is whether a better decentralized detection scheme can be found by considering non-stationary quantizers via the feedback from the fusion center, or any other forms of stationary quantizers that are (deterministic or random) measurable function from the range of \( Y_{k, n} \) to binary alphabet, say, \( \{0, 1\} \). For independent sensor observations, Tsitsiklis [11], Veeravalli [13] and Mei [6] provided a definite negative answer based on both asymptotic and numerical analysis: asymptotically optimal detection schemes (with either stationary or non-stationary binary quantizers) actually can be found in a subclass when the sensor quantizers are defined in (3). Given the difficulty of decentralized quickest change problems in HMM, it is logical to first consider the quantizer of the form in (3), since the results will shed light on general cases.
III. Preliminaries Involving the Centralized Setting and Notation

The detection scheme presented in this article is motivated by an asymptotic study in Fuh [3] of the classical centralized quickest change detection problems in HMM. For simplicity, in the following we consider the centralized setting of the sensor network with a single sensor K = 1, since the basic arguments are clearer for the single sensor setting and are readily extendable to the multiple sensor case.

For θ ∈ Θ, define \( f(y; \theta, X_n) \) as the probability density function of \( Y_n \) given \( X_n \), with respect to the Lebesgue measure. Let \( Y_1, \ldots, Y_n \) be the random variables from the hidden Markov model (2) with an unknown parameter θ, and denote \( p_n(Y_1, \ldots, Y_n; \theta) \) by the corresponding joint density function. Define

\[
S_n := \frac{p_n(Y_1, \ldots, Y_n; \theta_1)}{p_n(Y_1, \ldots, Y_n; \theta_0)} := \frac{\sum_{\sum_{x_{0}, \ldots, x_{n-1}}=1} \pi_0(x_0) \prod_{i=1}^{n-1} \pi_{\theta_1}(x_{i-1}, x_i) f(Y_i; \theta_1 | x_i)}{\sum_{\sum_{x_{0}, \ldots, x_{n-1}}=1} \pi_0(x_0) \prod_{i=1}^{n-1} \pi_{\theta_0}(x_{i-1}, x_i) f(Y_i; \theta_0 | x_i)}
\]

for fixed \( \theta_0, \theta_1 \in \Theta \). Under some general regularity conditions, Fuh [3] showed that for HMM models, the CUSUM procedure defined by

\[
T(c) := \inf \{ n : \max_{1 \leq s \leq n} \{ \log S_n - \log S_l \} \geq c \},
\]

is asymptotically optimal in the sense of asymptotically minimizing \( E_0(\tau(c)) \) among all detection schemes \( \tau(c) \) satisfying \( E_\infty(\tau(c)) \geq E_\infty(T(c)) \), where \( E_\infty(c) \rightarrow \infty \) as \( c \rightarrow \infty \).

To derive the asymptotic properties of the CUSUM procedure in HMM, it is useful to consider the alternative expression of \( S_n \). For this purpose, given a column vector \( u = (u_1, \ldots, u_d)^t \in R^d \), define the \( L_1 \)-norm of \( u \) as \( \|u\| = \sum_{i=1}^{d} |u_i| \). The likelihood ratio (5) then can be represented as

\[
S_n = \frac{p_n(Y_1, \ldots, Y_n; \theta_1)}{p_n(Y_1, \ldots, Y_n; \theta_0)} \cdot \frac{M_{\theta_1} \cdots M_{\theta_1} \pi_{\theta_1}}{M_{\theta_0} \cdots M_{\theta_0} \pi_{\theta_0}},
\]

where

\[
M_{\theta}(\theta) = \begin{bmatrix}
p_\theta(1, 1) f(Y_1; \theta | 1) & \cdots & p_\theta(d, 1) f(Y_1; \theta | 1) \\
\vdots & \ddots & \vdots \\
p_\theta(1, d) f(Y_1; \theta | d) & \cdots & p_\theta(d, d) f(Y_1; \theta | d)
\end{bmatrix}
\]

for \( l = 1, \ldots, n \), and \( \pi_{\theta} = (\pi_{\theta}(1), \ldots, \pi_{\theta}(d))^t \) is the pre-specified stationary distribution defined in (1).

Note that each component \( p_\theta(x, x') f(Y_1; \theta | x') \) in \( M_{\theta}(\theta) \) represents \( X_{l-1} = x \) and \( X_l = x' \), and \( Y_1 \) is a random variable with probability density \( f(y; \theta, x') \), for \( l = 1, \ldots, n \), therefore the \( M_{\theta}(\theta) \) are random matrices. By definition (2), \( \{X_n, Y_n, n \geq 0\} \) is a Markov chain on a state space \( D \times R \), and this implies that \( \{M_{\theta}(\theta), l = 1, \ldots, n\} \) is a sequence of Markov random matrices (See Section 2 of Fuh [3] for a formal definition). Hence, \( S_n \) is the ratio of the \( L_1 \)-norm of the products of Markov random matrices via representation (7)-(8). Note that \( \pi \) is fixed in (7).

By definition (2), \( \{(X_n, Y_n), n \geq 0\} \) is a Markov chain on a state space \( D \times R \), with \( \sigma \)-algebra \( A \), which is irreducible with respect to a maximal irreducibility measure on \( (D \times R, A) \) and is aperiodic. Denote \( Z_n := (X_n, Y_n) \) and \( D' := D \times R \). Define \( G_l(d,R) \) as the set of invertible \( d \times d \) matrices with real entries. For given \( l = 1, \ldots, n \), and \( \theta = \theta_0 \) or \( \theta_1 \), let \( M_{\theta}(\theta) \) be the random matrix from \( D' \times D' \) to \( G_l(d,R) \), as defined in (8). For convenience of notation, we still denote \( \theta = (\theta_0, \theta_1) \) and let

\[
T_n(\theta) = M_{\theta}(\theta) \cdots M_{\theta}(\theta) = (T_n(\theta_0), T_n(\theta_1))
\]

Then the system \( \{(Z_n, T_n(\theta)), n \geq 0\} \) is called a product of Markov random matrices on \( D' \times G_l(d,R) \times G_l(d,R) \). Denote \( P_\infty^\theta \) as the probability distribution of \( \{(Z_n, T_n(\theta)), n \geq 0\} \) with \( Z_0 = z \), and \( E_\infty^\theta \) as the expectation under \( P_\infty^\theta \).

Let \( u \in R^d \) be a d-dimensional vector, \( \pi := u/\|u\| \) the normalization of \( u \) (\( \|u\| \neq 0 \)), and denote \( P(R^d) \times P(R^d) \) as the projection space of \( R^d \) which contains all elements \( \pi \). For given \( \pi \in P(R^d) \) and \( M \in G_l(d,R) \), denote \( M \cdot \pi = M \cdot u \) and \( T_1(\theta)u = (T_1(\theta_0)u, T_1(\theta_1)u) \), for \( l = 0, \ldots, n \). Let

\[
W_0^\theta = (Z_0, T_0(\theta_0)u), W_1^\theta = (Z_1, T_1(\theta_0)u), \ldots, W_n^\theta = (Z_n, T_n(\theta_0)u).
\]

Then, \( \{W_n^\theta, n \geq 0\} \) is a Markov chain on the state space \( D' \times P(R^d) \times P(R^d) \) with the transition kernel

\[
P_\theta^\theta((z, \pi), A \times B) := E_\infty^\theta(I_{A \times B}(Z_1, M_1(\theta_1)u))
\]

for all \( z \in D' \), \( \pi := (\pi, \pi) \in P(R^d) \times P(R^d) \), \( A \in A \), and \( B \in B(P(R^d) \times P(R^d)) \), the Borel \( \sigma \)-algebra of \( P(R^d) \times P(R^d) \). For simplicity, we let \( P_\theta^\theta((z, \pi)) := P_\theta^\theta(\cdot, \cdot) \) and denote \( E_\infty^\theta((z, \pi)) \) as the expectation under \( P_\theta^\theta((z, \pi)) \). Since the Markov chain \( \{(X_n, Y_n), n \geq 0\} \) has transition probability density and the random matrix \( M_{\theta}(\theta) \) is driven by \( \{X_n, Y_n, n \geq 0\} \), it implies that the induced transition probability \( P_\theta^\theta(\cdot, \cdot) \) has a density with respect to \( \mu \). Denote it as \( P \) for simplicity. By definition (2), the Markov chain \( \{X_n, Y_n, n \geq 0\} \) satisfies Condition A of Fuh [3]. Therefore, the Markov chain \( \{W_n^\theta, n \geq 0\} \) has an invariant probability measure \( m_0^\theta \) on \( D' \times P(R^d) \times P(R^d) \), cf. Fuh [3].

Now, for \( z_0, z_1 \in D' \), \( \pi = (\pi, \pi) = (u(\theta_0), u(\theta_1)) \in P(R^d) \times P(R^d) \) and \( M = M_{z_0, z_1} = M_{\theta} = (M_{\theta_0}, M_{\theta_1}) \in G_l(d,R) \times G_l(d,R) \), let \( \sigma : (D' \times P(R^d) \times P(R^d)) \times (D' \times P(R^d) \times P(R^d)) \rightarrow R \) be \( \sigma(z_0, z_1, \pi, \bar{u}(\theta_0)) = \log \|M_{\theta}(\theta_0)u(\theta_0)\|/\|u(\theta_0)\| \). For \( \pi_{\theta_0}, \pi_{\theta_1} \in P(R^d) \),

\[
\log S_n = \log \|M_{\theta_1}(\theta_1)\cdots M_{\theta_1}(\theta_1)\pi_{\theta_1}\|/\|\pi_{\theta_1}\| + \cdots
\]

...
Furthermore, assume the parameter \( \theta \) and aperiodic on a finite state space \( \{W_n^\theta, n \geq 0\} \). It is evident.

For given \( P_{\theta_1} \) and \( P_{\theta_0} \), define the Kullback-Leibler information number as (4.2) of Fuh [3]

\[
K(P_{\theta_1}, P_{\theta_0}) = E_{P_{\theta_0}} \left( \log \frac{\|M_1(\theta_1)\pi_{\theta_1}\|}{\|M_1(\theta_0)\pi_{\theta_0}\|} \right),
\]

(10)

where \( P_{\theta_0} \) (\( P_{\theta_1} \)) denotes the probability of the Markov chain \( \{W_n^{\theta_0}, n \geq 0\} \) (\( \{W_n^{\theta_1}, n \geq 0\} \)), and \( E_{P_{\theta_0}} \) (\( E_{P_{\theta_1}} \)) refers to the expectation under \( P_{\theta_0} \) (\( P_{\theta_1} \)). Moreover, Fuh (2003) showed that the CUSUM procedure \( T(c) \) in (6) satisfies

\[
E_1(T(c)) = (1 + o(1)) \log E_\infty(T(c)) / K(P_{\theta_1}, P_{\theta_0}),
\]
as \( c \to \infty \).

Using the above notation, as in the classical centralized quickest change detection problems in HMM, we assume the following conditions hold throughout this article.

(A1) The sensor noises \( \varepsilon_{k,n} \)'s in the hidden Markov model (2) are independent over time and sensor.

(A2) For each \( \theta \in \Theta \), the Markov chain \( X = \{X_n, n \geq 0\} \), defined in (2), is ergodic (positive recurrent, irreducible and aperiodic) on a finite state space \( D = \{1, \ldots, d\} \).

Furthermore, assume the parameter \( \theta \) is identifiable in the sense that if, for some \( \theta, \theta' \in \Theta \), \( P_{\theta}^{(n)} = P_{\theta'}^{(n)} \) for all \( n \), then \( \theta = \theta' \).

(A3) For each \( 1 \leq k \leq K \), assume

\[
0 < K(P_{k,\theta_1}, P_{k,\theta_0}) < \infty,
\]

and

\[
E_{P_{\theta_0}} \left( \log \frac{\|M_1(\theta_1)\pi_{\theta_1}\|}{\|M_1(\theta_0)\pi_{\theta_0}\|} \right)^2 < \infty.
\]

Remark: Condition (A1) is a standard condition in sensor networks. The ergodicity condition in (A2) for Markov chains is quite general and covers several interesting examples. Condition (A3) is a standard moment condition for the Kullback-Leibler information number.

IV. FUSION RULES WITH GIVEN STATIONARY SENSOR QUANTIZERS

In this section, we are interested in determining the asymptotically optimal fusion center schemes when the stationary binary quantizers in (3) are given.

In this case, the fusion center faces the classical change-point detection problems in hidden Markov models based on the quantized binary vector \( U_n = (U_{1,n}, \ldots, U_{K,n}) \). Denote the stationary thresholds by \( \lambda = (\lambda_1, \ldots, \lambda_K) \). It is evident that conditional on \( X_n = x_n \), by (2),

\[
P(U_{k,n} = 1|x_n) = P(Y_{k,n} \geq \lambda_k|x_n)
= P(\varepsilon_{k,n} \geq \lambda_k - h(x_n)|x_n) = \Phi \left( \frac{h_k(x_n) - \lambda_k}{\sigma_k} \right),
\]

and

\[
P(U_{k,n} = 0|x_n) = 1 - P(U_{k,n} = 1|x_n) = \Phi \left( \frac{\lambda_k - h_k(x_n)}{\sigma_k} \right).
\]

Or equivalently, for \( u_{k,n} = 0 \) or 1,

\[
P(U_{k,n} = u_{k,n}|x_n) = \Phi \left( (2u_{k,n} - 1) \frac{h_k(x_n) - \lambda_k}{\sigma_k} \right).
\]

Thus we have

\[
P(U_n = u_n|x_n) = \prod_{k=1}^{K} \Phi \left( (2u_{k,n} - 1) \frac{h_k(x_n) - \lambda_k}{\sigma_k} \right).
\]

Hence the conditional distribution of \( U_n \) depends only on \( x_n \). By [3], based on \( U_n \), an asymptotic optimal rule at the fusion center is the following CUSUM scheme.

Based on the vector observations \( U_n = (U_{1,n}, \ldots, U_{K,n}) \) from the hidden Markov model (2) with an unknown parameter \( \theta \), denote

\[
S_n^u := \frac{p_n(U_{1,n}, \ldots, U_{K,n}; \theta_1)}{p_n(U_{1,n}, \ldots, U_{K,n}; \theta_0)}
\]

for fixed \( \theta_0, \theta_1 \in \Theta \). For HMM models, the calculation of likelihood ratios \( S_n^u \) can be very complicated, and a more efficient approach is to use (9) to calculate it.

The fusion center then uses the CUSUM procedure for hidden Markov models with log-likelihood ratio boundary \( c_\gamma \) to detect whether or not a change has occurred, i.e., the stopping time \( N(c_\gamma) \) is given by

\[
N(c_\gamma) := \inf \{ n : \max_{1 \leq k \leq n} \left( \log S_k^u - \log S_k^l \right) \geq c_\gamma \},
\]

(11)

where \( c_\gamma \) is chosen such that \( E_\infty N(c_\gamma) = \gamma \).

V. CHOOSING THE OPTIMAL LOCAL SENSOR THRESHOLDS

For any given local stationary thresholds at the sensor level, \( N(c_\gamma) \) in (11) is the asymptotically optimal rule at the fusion center level. A natural question is how to choose good local stationary thresholds. A simple approach for selecting good local thresholds is based on the asymptotic properties of \( N(c_\gamma) : \gamma \to \infty \),

\[
E_1(\tau) = (1 + o(1)) \frac{\log \gamma}{K^u(\lambda)},
\]

where the \( K^u(\lambda) \) is the Kullback-Leibler information number induced on \( U_n \) when the stationary thresholds are \( \lambda = (\lambda_1, \ldots, \lambda_K) \). Thus, the Kullback-Leibler information number (or relative entropy) \( K^u(\lambda) \) plays an essential role in our setting, the optimal sensors thresholds can be defined by selecting \( \lambda_k \) which maximize the Kullback-Leibler information number \( K^u(\lambda) \).

Specifically, denote by \( P_{\theta_1} \) the “induced” probability measure on \( U \) when the underlying probability measure on the raw data \( Y \)’s is \( P_{\theta_1} \). Then, using the notations in Section III, we have

\[
K^u(\lambda) = E_{P_{\theta_1}} \left( \log \frac{\|M_{1,\lambda}(\theta_1)\pi_{\theta_1}\|}{\|M_{1,\lambda}(\theta_0)\pi_{\theta_0}\|} \right),
\]

(12)
where $M_{1,\lambda}(\theta) = (m_{i,j}^{u})_{d \times d}^{1}$ is a $d \times d$ matrix with entries
\[
m_{i,j}^{u} = p_{\theta}(i,j)f(U_{1}; \theta | x_{u} = j) = p_{\theta}(i,j)\prod_{k=1}^{K} \Phi \left( (2U_{k,1} - 1) \frac{h_{k}(j) - \lambda_{k}}{\sigma_{k}} \right).
\]

Then the optimal values of the thresholds $\lambda = (\lambda_{1}, \cdots, \lambda_{K})$ can be defined by selecting $\lambda = (\lambda_{1}, \cdots, \lambda_{K})$ which jointly maximizes $K^{u}(\lambda)$.

Unfortunately, this global maximization seems to be very complicated, since it is a nonlinear optimization problem over a $K$-dimensional space $\lambda = (\lambda_{1}, \cdots, \lambda_{K})$ and the expectation in $K^{u}(\lambda)$ is defined over $2^{K}$ binary vector space $U$.

A simpler (possibly suboptimal) method for selecting the local thresholds is to optimize each sensor individually in the sense of maximizing the local Kullback-Leibler information number termwise. That is, the threshold $\lambda_{k}$ can be chosen by maximizing the local Kullback-Leibler information number
\[
K_{\text{local}, k}(\lambda) = \mathbb{E}_{p_{\theta}^{(k)}} \left( \log \left\| \frac{M_{1,\lambda_{k}}^{k, u}(\theta_{1})\pi_{\theta_{1}}}{\lambda_{k}} \right\| \right),
\]

where $M_{1,\lambda_{k}}^{k, u}(\theta) = (m_{i,j}^{k, u})_{d \times d}$ is a $d \times d$ matrix with entries
\[
m_{i,j}^{k, u} = p_{\theta}(i,j)\Phi \left( (2U_{k,1} - 1) \frac{h_{k}(j) - \lambda_{k}}{\sigma_{k}} \right).
\]

Note that $K_{\text{local}, k}(\lambda)$ in (13) can be thought of as the Kullback-Leibler information number when the sensor $S_{k}$ is the only sensor in a system. The local maximization is conducted for each sensor, and is optimized over $(-\infty, \infty)$, and the expectation $K^{u}_{\text{local}, k}(\lambda)$ is over the binary random variable $U_{k} \in \{0, 1\}$.

Compared to the global maximization method, the local maximization approach is much easier to implement (though suboptimal), as it considers $K$ optimization problems over an $1$-dimensional space instead of one complicated optimization problem over a $K$-dimensional space. It is interesting to note that for independent sensor observations, the global maximization is equivalent to the local maximization, see, for example, Crow and Schwartz (1996). However, they seem to be different in HMM in many situations, partly due to the fact that the global Kullback-Leibler information number in (12) is not necessary the sum of all local Kullback-Leibler information numbers in (13).

So far we simply apply the results in the classical centralized quickest change detection problems in HMM to the decentralized quickest change detection problem in Section II without checking the conditions of these results. It turns out that the following additional assumption is needed in our asymptotic optimality theory for decentralized quickest change detection:

(A4) Assume there exists some $\lambda$ such that
\[
0 < K^{u}(\lambda) < \infty,
\]

and
\[
\mathbb{E}_{p_{\theta}^{(k)}} \left( \log \left\| \frac{M_{1,\lambda_{k}}^{k, u}(\theta_{1})\pi_{\theta_{1}}}{\lambda_{k}} \right\| \right) ^{2} < \infty.
\]

This condition is necessary to avoid a very subtle situation where the change is detectable by the raw sensor observations, but is not by any (binary) quantized observations. Note that this situation rarely occurs for independent observations (and never occurs for many widely used distributions such as normal distributions), but it is possible under HMM models (see our example below).

By Fuh (2003), it is straightforward to prove the following theorem which establishes the asymptotic optimality properties of the proposed decentralized detection scheme $N(c_{i})$ defined in (11) when the threshold $\lambda_{i}$’s are chosen from the global maximization method:

\textbf{Theorem 1}: Under the conditions (A1)-(A3) and (A4), if the local sensor thresholds are chosen by the global maximization method, then the corresponding procedure $N(c_{i})$ defined in (11) asymptotically minimizes the detection delay $\mathbb{E}_{1}(N(c_{i}))$ among any decentralized detection schemes satisfying the constraint (4) and with stationary local sensor thresholds.

\section{VI. EXAMPLE}

To illustrate our ideas, let us focus on a simple specific example on a two-state Markov chain $X_{n} \in \{1, 2\}$. Let $\theta = (\alpha, \beta)$ and assume the transition probability matrix of $\{X_{n}\}$ has the form $P_{\theta} = \left( \begin{array}{cc} 1 - \alpha & \alpha \\ \beta & 1 - \alpha \beta \end{array} \right)$, where $0 < \alpha, \alpha \beta < 1$. Thus the stationary distribution $\pi_{\theta} = \left( \frac{\beta}{\beta + 1}, \frac{1}{\beta + 1} \right)^{T}$. Suppose there are $K$ identical sensors in the network. At each sensor, assume the actual noiseless value levels for sensor observations are $h(X_{n}) \in \{-1, 1\}$ such that $h(1) = -1$ and $h(2) = 1$, and the sensor noises $\varepsilon_{k,n}$’s are i.i.d. $N(0, 1)$.

Suppose we are interested in detecting a change in $\theta_{0} = (\alpha_{0}, \beta_{0})$ to $\theta_{1} = (\alpha_{1}, \beta_{1})$, and now we apply our proposed methods to derive the optimal decentralized detection scheme (among those with stationary sensor thresholds). Note that for stationary sensor thresholds, the asymptotically optimal fusion center scheme is defined in (11), and thus it suffices to find the optimal threshold values $\lambda$’s at local sensors and the corresponding value of the global Kullback-Leibler information number $K^{u}(\lambda)$ defined in (12).

In this specific example, for a given threshold $\lambda = (\lambda_{1}, \cdots, \lambda_{K})$, define $U_{1,\lambda} = I(Y_{1} \geq \lambda_{1}), \cdots,$ and $U_{K,\lambda} = I(Y_{K} \geq \lambda_{K})$, and denote by $f_{1}(\cdot)$ and $f_{2}(\cdot)$ the probability mass functions of $(U_{1,\lambda}, \cdots, U_{K,\lambda})$ conditional on $X_{1} = 1$ and 2, respectively, where we suppress the subscript $\lambda$ in $f_{1}(\cdot)$ and $f_{2}(\cdot)$ to simplify the notation. Now define two new random variables
\[
F = f_{1}(U_{1,\lambda}, \cdots, U_{K,\lambda}),
\]
\[
G = f_{2}(U_{1,\lambda}, \cdots, U_{K,\lambda})
\]

Then the corresponding random matrix
\[
M_{1,\lambda}(\theta) = \left( \begin{array}{c} (1 - \alpha)F \\ \alpha G \end{array} \right)
\]

and thus
\[
\|M_{1,\lambda}(\theta)\| = \frac{\beta}{\beta + 1} F + \frac{1}{\beta + 1} G = \frac{\beta F + G}{\beta + 1}.
\]
Moreover, under \(P_\theta\) (i.e., when \(\theta\) is the true parameter), the marginal probability mass function \((U_{1\lambda}, \cdots, U_{K\lambda})\) is
\[
f_1(U_{1\lambda}, \cdots, U_{K\lambda}) = f_1(U_{1\lambda})P_\theta(X = 1) + f_2(U_{1\lambda}, \cdots, U_{K\lambda})P_\theta(X = 2),
\]
which is simply \(\frac{\beta F + G}{\beta + 1}\) by the notation in (14) and the fact that \((P_\theta(X = 1), P_\theta(X = 2)) = \pi_\theta = (\frac{\beta}{\pi + 1}, \frac{\lambda}{\pi + 1})\) is the stationary distribution. Thus, in this specific example, the global Kullback-Leibler information \(K^u(\lambda)\) in (12), i.e., the Kullback-Leibler information on \((U_{1\lambda}, \cdots, U_{K\lambda})\), can be rewritten as
\[
K^u(\lambda) = \sum \left( \frac{\beta_1 F + G}{\beta_1 + 1} \right) \log \frac{\beta_1 F + G}{\beta_1 + 1}, \tag{15}
\]
where the summation is over all possible \(2^K\) binary vector space \((U_{1\lambda}, \cdots, U_{K\lambda})\).

It is worth mentioning an interesting fact in this example: the Kullback-Leibler information \(K^u(\lambda)\) in (15) does not depend on \(\alpha\), although it does depend on \(\beta\), as the random variables \(F\) and \(G\) only depend on local thresholds \(\lambda\)'s. In other words, when detecting a change in \(\theta = (\alpha, \beta)\), if \(\beta\) is a constant and \(\alpha\) is the parameter of interest, then the global Kullback-Leibler information number \(K^u(\lambda)\) is 0 for any thresholds \(\lambda\)'s, no matter how many sensors we have.

That is, when \(\beta\) is constant, for any quantized thresholds \(\lambda\)'s, the quantized observations \(U_{K\lambda}\)'s bear no information about the change in \(\alpha\), although the raw sensor observations \(Y\)'s do. This is equivalent to saying that when \(\beta\) is a constant, a change of \(\alpha\) is undetectable by using any (binary) quantized sensor observations, although such a change is detectable by using the raw sensor observations \(Y\)'s.

For this reason, in the following we assume the parameter \(\alpha \in (0, 1)\) is a constant and we are interested in monitoring a change in \(\beta\) from \(\beta_1\) to another value \(\beta_2\). Here we assume \(\alpha = 0.2\), but the actual choice of \(\alpha\) can be flexible and does not impact our results below, other than the range of suitable \(\beta\), due to the requirement that \(0 < \alpha \beta < 1\) in the definition of the transition probability matrix.

A. Local Maximization Approach

Since all sensors are homogeneous and the local maximization approach is equivalent to the global maximization approach in a single sensor system, the local Kullback-Leibler information at each sensor can be calculated as in (15) with the system with \(K = 1\) sensor. The remaining part is to derive \(F\) and \(G\) in (14) when \(K = 1\).

For a given threshold \(\lambda\), \(U_{1\lambda} = I(Y \geq \lambda)\), the respective probability mass functions of \(U_{1\lambda}\) conditional on \(X = 1\) and 2 are
\[
f_1(U_{1\lambda}) = \begin{cases} P(U_{1\lambda} = 1|x = 1) = \Phi(-1 - \lambda), & \text{if } U_{1\lambda} = 1; \\ P(U_{1\lambda} = 0|x = 1) = \Phi(1 + \lambda), & \text{if } U_{1\lambda} = 0; \end{cases}
\]
and
\[
f_2(U_{1\lambda}) = \begin{cases} P(U_{1\lambda} = 1|x = 2) = \Phi(1 - \lambda), & \text{if } U_{1\lambda} = 1; \\ P(U_{1\lambda} = 0|x = 2) = \Phi(-1 + \lambda), & \text{if } U_{1\lambda} = 0. \end{cases}
\]
Hence, under the local maximization approach, the corresponding random variables \(F\) and \(G\) in (14) are
\[
F = f_1(U_{1\lambda}) = \begin{cases} \Phi(-1 - \lambda), & \text{if } U_{1\lambda} = 1; \\ \Phi(1 + \lambda), & \text{if } U_{1\lambda} = 0; \end{cases}
\]
and
\[
G = f_2(U_{1\lambda}) = \begin{cases} \Phi(1 - \lambda), & \text{if } U_{1\lambda} = 1; \\ \Phi(-1 + \lambda), & \text{if } U_{1\lambda} = 0. \end{cases}
\]
To further simplify the computation, if we define
\[
Q_\beta(\lambda) = \beta \beta \beta + 1 \Phi(-1 - \lambda) + \beta + 1 \Phi(1 - \lambda),
\]
which is the marginal probability \(P_\theta(U_{1\lambda} = 1)\), then we have
\[
\beta F + G \beta \beta \beta + 1 = \begin{cases} Q_\beta(\lambda), & \text{if } U_{1\lambda} = 1; \\ 1 - Q_\beta(\lambda), & \text{if } U_{1\lambda} = 0. \end{cases}
\]
Thus, the local Kullback-Leibler information at each sensor can be rewritten as
\[
K_{local}(\lambda) = Q_\beta(\lambda) \log \frac{Q_\beta(\lambda)}{Q_\beta(\lambda)} + (1 - Q_\beta(\lambda)) \log \frac{1 - Q_\beta(\lambda)}{1 - Q_\beta(\lambda)}. \tag{16}
\]
The local maximization is then to find \(\lambda\) which maximizes \(K_{local}(\lambda)\) in (16).

For illustration purpose, Fig. 2 plots local Kullback-Leibler information number \(K_{local}(\lambda)\) as a function of \(\lambda\) when the parameter \(\beta\) changes from \(\beta_1 = 1\) to \(\beta_1 = 2\). The graph suggests that there is a unique \(\lambda\) to maximize \(K_{local}(\lambda)\).
Furthermore, denote by \( \lambda_{\text{max}}(\beta_1) \) the optimal values of \( \lambda \) in the problem when the parameter \( \beta \) changes from \( \beta_0 = 1 \) to \( \beta_1 \). Denote by \( K_{\text{max}}(\beta_1) \) the corresponding local Kullback-Leibler information number. Then it is easy to see that \( \lambda_{\text{max}}(\beta) = -\lambda_{\text{max}}(1/\beta) \) and \( K_{\text{max}}(\beta) = K_{\text{max}}(1/\beta) \). Some values of \( \lambda_{\text{max}} \) and \( K_{\text{max}} \) are reported in the following table:

<table>
<thead>
<tr>
<th>( \beta_1 )</th>
<th>0.25</th>
<th>0.5</th>
<th>0.8</th>
<th>1.2</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_{\text{max}} )</td>
<td>-0.1517</td>
<td>-0.0792</td>
<td>-0.0258</td>
<td>0.0211</td>
<td>0.0468</td>
<td>0.0792</td>
</tr>
<tr>
<td>( K_{\text{max}} )</td>
<td>0.0874</td>
<td>0.0262</td>
<td>0.0029</td>
<td>0.0019</td>
<td>0.0094</td>
<td>0.0262</td>
</tr>
</tbody>
</table>

It is interesting to note that the noiseless sensor value are \( \pm 1 \) (symmetric), an intuitive choice of \( \lambda \) may simply be 0. However, based on our approach, the optimal threshold \( \lambda \) is slightly different from 0. Similar results have already been reported for quickest change detection problems with independent sensor observations, see Crow and Schwartz [2] and Mei [6].

B. Global Maximization Approach with \( K = 2 \)

Now let us investigate the global maximization approach in a network with 2 sensors. For a given threshold \( \lambda = (\lambda_1, \lambda_2) \), define \( U_{1\lambda} = I(Y_1 \geq \lambda_1) \) and \( U_{2\lambda} = I(Y_2 \geq \lambda_2) \).

As in the previous subsection, it suffices to calculate \( F \) and \( G \) in (14) explicitly in this case. Note that conditional on \( X_1 \), \( U_{1\lambda} \) and \( U_{2\lambda} \) are independent. Then

\[
F = f_1(U_{1\lambda})f_1(U_{2\lambda}) \quad \text{and} \quad G = f_2(U_{1\lambda})f_2(U_{2\lambda}),
\]

where \( f_1(\cdot) \) and \( f_2(\cdot) \) are also used to denote the marginal mass probability functions of \( U \)'s conditional on \( X_1 = 1 \) or 2, respectively (we suppress the subscript \( k \) in \( f_k(\cdot) \) and \( f_k(\cdot) \) to simplify the notation).

Note that the global Kullback-Leibler information \( K^{u}(\lambda) \) in (15) is defined through \( \frac{\beta F + G}{\beta + 1} \), the marginal distribution of \( (U_{1\lambda}, U_{2\lambda}) \) under \( P_\theta \), which can be expressed explicitly as follows. For \( i = 1,2 \), define

\[
A_i = P(U_i = 1|x_n = 1) = \Phi(-1 - \lambda_i),
\]

\[
B_i = P(U_i = 0|x_n = 1) = \Phi(1 + \lambda_i) - 1 - A_i,
\]

\[
C_i = P(U_i = 1|x_n = 2) = \Phi(-1 - \lambda_i),
\]

\[
D_i = P(U_i = 0|x_n = 2) = \Phi(-1 + \lambda_i) - 1 - C_i.
\]

and the unconditional probability mass function of \( (U_{1\lambda}, U_{2\lambda}) \) under \( P_\theta \) is

\[
Q_{11} = P_{\theta}(U_1 = 1, U_2 = 1) = \frac{\beta}{\beta + 1}A_1 A_2 + \frac{1}{\beta + 1} C_1 C_2,
\]

\[
Q_{22} = P_{\theta}(U_1 = 1, U_2 = 0) = \frac{\beta}{\beta + 1}A_1 B_2 + \frac{1}{\beta + 1} C_1 D_2,
\]

\[
Q_{32} = P_{\theta}(U_1 = 0, U_2 = 1) = \frac{\beta}{\beta + 1}B_1 A_2 + \frac{1}{\beta + 1} D_1 C_2,
\]

\[
Q_{42} = P_{\theta}(U_1 = 0, U_2 = 0) = \frac{\beta}{\beta + 1}B_1 B_2 + \frac{1}{\beta + 1} D_1 D_2.
\]

Then we have

\[
\frac{\beta F + G}{\beta + 1} = \begin{cases} 
Q_{11}, & \text{if } U_1 = 1 \text{ and } U_2 = 1; \\
Q_{22}, & \text{if } U_1 = 1 \text{ and } U_2 = 0; \\
Q_{32}, & \text{if } U_1 = 0 \text{ and } U_2 = 1; \\
Q_{42}, & \text{if } U_1 = 0 \text{ and } U_2 = 0.
\end{cases}
\]

For the parameter \( \theta = \theta_0 \) or \( \theta_1 \), define \( Q_{i}^{(0)} \) and \( Q_{i}^{(1)} \) similarly as above. Then the global Kullback-Leibler information on \( (U_1, U_2) \) can be rewritten as

\[
K^{u}(\lambda_1, \lambda_2) = \sum_{i=1}^{4} Q_{i}^{(1)} \log \frac{Q_{i}^{(1)}}{Q_{i}^{(0)}}. \tag{17}
\]

Fig. 3 plots the global Kullback-Leibler information number \( K^{u}(\lambda_1, \lambda_2) \) as a function of both local thresholds \( (\lambda_1, \lambda_2) \) in the problem of monitoring a change in \( \beta \) from \( \beta_0 = 1 \) to another value \( \beta_1 = 2 \). To better illustrate the structure of the global Kullback-Leibler information number, Fig. 4 also plots the global Kullback-Leibler information number \( K^{u}(\lambda_1, \lambda_2) \) when \( \lambda_1 = \lambda_2 \). Numerical calculations show that for the global maximization methods, the optimal choices values of the \( \lambda \)'s are \( \lambda_1 = \lambda_2 = -0.170 \), and the corresponding maximum global Kullback-Leibler information number is \( K_{\text{max}} = 0.03597 \).

Recall that the optimal values of \( \lambda \)'s under the local maximization method is \( \lambda_1 = \lambda_2 = 0.0792 \). By (17), it is easy to show that the corresponding global Kullback-Leibler information number is \( K = 0.03588 \), which is slightly smaller than the optimal value from the global maximization method. This is the price we have to pay to simplify the computation.

VII. DISCUSSIONS

Our results have raised the following interesting questions:

1. **Arbitrary stationary sensors**: Here we only consider a special form of stationary binary quantizers that compares the current sensor observations to a constant threshold. As we mentioned earlier, for independent sensor observations,
stationary quantizers. Whether this is still true for HMM models, at least among all or non-stationary) sensor quantizers. It is interesting to see decentralized detection schemes among all possible (stationary or non-stationary) sensor quantizers. It is interesting to see whether this is still true for HMM models, at least among all stationary quantizers.

2. Non-stationary quantizers: Here we focus on decentralized detection schemes with stationary binary sensors, and develop efficient schemes within the class of stationary quantizers. A natural question one may ask is whether or not the “best” schemes with stationary quantizers is still asymptotically optimal if the sensor thresholds $\lambda_k$’s, or more generally, sensor quantizers, are allowed to be adapted to past sensor messages (i.e., the fusion center can change the system parameters every time step). While reference [6] surprisingly showed that this is true for independent observations, it is doubtful for HMM models. The corresponding asymptotic analysis in HMM models will be presented elsewhere.

3. Block length effects: In practice, sensors may be designed to send messages every $M$ time steps to further reduce the communication cost. What is the best choice for the blocksize $M$ in the sense of balancing the tradeoff between the desire to quickly detect the true change and the need to minimize the communication cost? It is useful to extend the idea of Crow and Schwartz [2] for independent observations to HMM models.

4. More Complicated Models. Here we only consider a simple hidden Markov model (2) with a finite space Markov chain. In many applications we may deal with a continuous Markov chain, or more complicated hidden Markov models. It seems to be straightforward to generalize our results to these more complicated cases, but probably we need to pay special attentions to condition (A4).

VIII. Conclusions

In this article, we examined the decentralized quickest change detection problem in sensor networks. In the detection scheme we considered, each sensor simply compared its current sensor observation to a stationary threshold, and the fusion center used a CUSUM-type scheme based on binary vectors. We also introduced the global and local maximization methods for choosing the “optimal” stationary local thresholds at the sensor level.

While the classical quickest change detection problems have been studied for several decades, the decentralized quickest change detection, particularly in HMM models, is still in its infancy. Clearly, there is a lot of research ahead to improve the theory, algorithms and methodology. Hopefully this article can stimulate further research.

REFERENCES


Fig. 4. Global maximization on two sensors with identical thresholds: plot of local K-L information $K^*(\lambda)$ as a function of $\lambda_1 = \lambda_2 = \lambda$. Top: when $\lambda \in [-5, 5]$ and bottom: when $\lambda \in [-0.5, 0.5]$. 

This type of simple sensor quantizers, or more generally, the Monotone likelihood ratio Quantizer (MLRQ) proposed in Tsitsiklis [11], actually leads to asymptotically optimal decentralized detection schemes among all possible (stationary or non-stationary) sensor quantizers. It is interesting to see whether this is still true for HMM models, at least among all stationary quantizers.