On optimal quantization rules for sequential decision problems

XuanLong Nguyen[1], Martin J. Wainwright[1,2] and Michael I. Jordan[1,2]

[1] Department of Electrical Engineering and Computer Science
[2] Department of Statistics
University of California, Berkeley
{xuanlong,wainwrig,jordan}@eecs.berkeley.edu

Abstract— We consider the problem of sequential decentralized detection, a problem that entails the choice of a stopping rule (specifying the sample size), a global decision function (a choice between two competing hypotheses), and a set of quantization rules (the local decisions on the basis of which the global decision is made). The main result of this paper is to resolve an open problem posed by Veeravalli et al. [12] concerning whether optimal local decision functions for the Bayesian formulation of sequential decentralized detection can be found within the class of stationary rules. We provide a negative answer to this question by exploiting an asymptotic approximation to the optimal cost of stationary quantization rules, and the asymmetry of the Kullback-Leibler divergences. In addition, we show that asymptotically optimal quantizers, when restricted to the space of blockwise stationary quantizers, are likelihood-based threshold rules.[1]

I. INTRODUCTION

In this paper, we consider the problem of sequential decentralized detection (see, e.g., [10], [11], [5]). Detection is a classical discrimination or hypothesis-testing problem, in which observations $\{X_1, X_2, \ldots\}$ are assumed to be drawn i.i.d. from the conditional distribution $\mathbb{P}(\,\cdot\,|\,H\,)$, where $H \in \{0, 1\}$ is the unknown state of the environment and in which the goal is to infer H. Placing this problem in a communication-theoretic context, a *decentralized detection problem* is a hypothesis-testing problem in which the decision-maker is not given access to the raw data points X_n, but instead must infer H based only on a set of *quantization rules* or *local decision functions* $\{U_n = \phi_n(X_n)\}$, where n is the time step. Finally, placing the problem in a real-time context, the *sequential decentralized detection problem* considers a stream of data $\{X_1, X_2, \ldots\}$ and a corresponding stream of summary statistics $\{U_1, U_2, \ldots\}$, and

asks the decision-maker to make a decision regarding H in a manner that optimally trades off accuracy and delay. That is, the sequential decentralized detection problem extends the classical formulation of sequential centralized decision-making problems (see, e.g., [8], [4]) to the decentralized setting.

In setting up a general framework of sequential decentralized problems, Veeravalli et al. [12] defined five problems ("case A" through "case E"), distinguished from one another by the amount of information available to the local sensors. In particular, in case E, the local sensors are provided with memory and with feedback from the global decision-maker (also known as the *fusion center*), so that each sensor has available to it the current data X_n, as well as all of the summary statistics from all of the other local sensors. In other words, each local sensor has the same snapshot of past state as the fusion center; this is an instance of a so-called "quasi-classical information structure" [3] for which dynamic programming (DP) characterizations of the optimal decision functions are available. Veeravalli et al. [12] exploit this fact to show that the decentralized case has much in common with the centralized case, in particular that likelihood ratio tests are optimal local decision functions at the sensors and that a variant of a sequential probability ratio test is optimal for the decision-maker.

Unfortunately, however, part of the spirit of the decentralized detection is arguably lost in case E, which requires full feedback. In particular, in applications such as power-constrained sensor networks, we generally do not wish to assume that there are high-bandwidth feedback channels from the decision-maker to the sensors, nor do we wish to assume that the sensors have unbounded memory. Most suited to this perspective—and the focus of this paper—is case A, in which the local decisions are of the simplified form $\phi_n(X_n)$; i.e., neither local memory nor feedback are assumed to be available.

Noting that case A is not amenable to dynamic programming and is presumably intractable, Veeravalli et al. [12] suggest restricting the analysis to the class of *stationary* local decision functions; i.e., local decision functions ϕ_n that are independent of n. They conjecture that stationary decision functions may actually be optimal in the setting of case A (given the intuitive symmetry and high degree of independence of the problem in this case), even though it is not possible to verify this optimality via DP arguments. The truth or falsity of this conjecture has remained open since it was first posed by Veeravalli et al. [12], [11].

In this paper, we resolve this question by showing that stationary decision functions are not in fact optimal for decentralized problems of type A. We do so by providing an asymptotic characterization of stationary decision functions, one which allows us to provide counterexamples to the stationarity conjecture, both in an exact and an asymptotic setting. In the asymptotic setting (i.e., when the cost per sample goes to 0), we provide a proof that in general there is always a range of prior probabilities for which stationary strategies are suboptimal. (We note in passing that the key reason for the suboptimality is easily stated—it arises from the asymmetry of the Kullback-Leibler divergence.)

It is well known [10] that optimal quantizers—*when unrestricted*—can be expressed as threshold rules based on the log likelihood ratio (LLR). Our counterexamples in the stationary conjecture imply that the thresholds need not be stationary (i.e., the threshold may differ from sample to sample). In the remainder of the paper, we addresses a partial converse to this issue: specifically, if we restrict ourselves to stationary (or blockwise stationary) quantizer designs, then there exists an optimal design consisting of LLR-based threshold rules.

II. BACKGROUND

This section provides necessary background on the Bayesian formulation of sequential detection problems, and Wald's approximation of the optimal cost.

Sequential detection problems: Let f^0 and f^1 represent the conditional distributions of X, conditioning on $\{H = 0\}$ and $\{H = 1\}$ respectively. We focus on the Bayesian formulation of the sequential detection problem [8], [11], and let $\pi^1 = \mathbb{P}(H = 1)$ and $\pi^0 = \mathbb{P}(H = 0)$ denote the prior probabilities on the two hypotheses. A sequential decision rule consists of a *stopping time* N (defined with respect to the sigma field $\sigma(X_1, \ldots, X_N)$), and a decision function γ (measurable with respect to $\sigma(X_1, \ldots, X_N)$). The cost function is a weighted sum

of the sample size N and the probability of incorrect decision

$$R(N, \gamma) := \mathbb{E}\big\{cN + \mathbb{I}[\gamma(X_1, \ldots, X_N) \neq H]\big\}, \quad (1)$$

where $c > 0$ is the incremental cost of each sample. The overall goal is to choose the pair (N, γ) so as to minimize the expected loss (1).

It is well known that a stopping rule based on the likelihood ratio is optimal [13], [1]. In particular, given the partial sum $L_n(X_1, \ldots, X_n) := \sum_{i=1}^n \log \frac{f^1(X_i)}{f^0(X_i)}$, the optimal stopping rule takes the form

$$N = \inf\big\{n \geq 1 \,\big|\, L_n \notin (a, b)\big\}, \quad (2)$$

for real numbers $a < b$. Given this stopping rule, the optimal decision function has the form

$$\gamma(L_N) = \begin{cases} 1 & \text{if } L_N \geq b, \\ 0 & \text{if } L_N \leq a. \end{cases} \quad (3)$$

We now develop an exact expression for the optimal cost of the decision rule (3). Consider the two types of error:

$$\begin{aligned} \alpha &= \mathbb{P}_0(\gamma(L_N) \neq H) = \mathbb{P}_0(L_N \geq b) \\ \beta &= \mathbb{P}_1(\gamma(L_N) \neq H) = \mathbb{P}_1(L_N \leq a). \end{aligned}$$

Now define $\mu^1 = \mathbb{E}_1[\log \frac{f^1(X_1)}{f^0(X_1)}] = D(f^1 || f^0)$ and $\mu^0 = -\mathbb{E}_0[\log \frac{f^1(X_1)}{f^0(X_1)}] = D(f^0 || f^1)$. With this notation, the cost $J(a, b)$ of the decision rule based on envelopes a and b can be written as

$$\begin{aligned} J(a, b) &= \mathbb{E}\big\{cN + \mathbb{I}[\gamma(X_1, \ldots, X_N) \neq H]\big\} \\ &= c\pi^1 \frac{\mathbb{E}_1 L_N}{\mu^1} + c\pi^0 \frac{\mathbb{E}_0 L_N}{-\mu^0} + \pi^0 \alpha + \pi^1 \beta, \quad (4) \end{aligned}$$

where the second line follows from Wald's equation.

Wald's approximation: When the data are i.i.d. conditioned on the hypothesis, the optimal cost can be characterized using a dynamic programming approach [1], [8]. Let us consider an approximate expression for the cost, due to Wald (cf. [9]). To begin, the errors α and β are related to a and b by the classical inequalties $\alpha \leq (1 - \beta)/e^b$ and $\beta \leq e^a(1 - \alpha)$. In general, these inequalities need not hold with equality because the likelihood ratio at L_N at the optimal stopping time N might overshoot either a or b (instead of attaining precisely the value a or b at the stopping time). The approximation is based on ignoring this overshoot, and replacing the inequalities by equalities to solve for corresponding values of α and β:

$$\alpha \approx \frac{1 - e^a}{e^b - e^a} \text{ and } \beta \approx \frac{e^{-b} - 1}{e^{-b} - e^{-a}}. \quad (5)$$

Alternatively, we can derive a and b in terms of α and β as:

$$a \approx a(\alpha, \beta) := \log \frac{\beta}{1 - \alpha} \text{ and } b \approx b(\alpha, \beta) := \log \frac{1 - \beta}{\alpha}. \tag{6}$$

The approximate mapping (5) and (6) between (a, b) and (α, β) yields the following approximation

$$\frac{\mathbb{E}_0[L_N]}{-\mu^0} \approx \frac{(1 - \alpha)a + \alpha b}{-\mu^0} \tag{7a}$$

$$\frac{\mathbb{E}_1[L_N]}{\mu^1} \approx \frac{(1 - \beta)b + \beta a}{\mu^1} \tag{7b}$$

Combining this approximation with the approximation (6) for a and b, we obtain *Wald's approximation* for the cost of a sequential test with error α and β:

$$G(\alpha, \beta) := c\pi^0 \frac{D(\alpha || 1 - \beta)}{\mu^0} + c\pi^1 \frac{D(1 - \beta || \alpha)}{\mu^1}$$
$$+ \pi^0 \alpha + \pi^1 \beta, \tag{8}$$

where we define $D(1 - \beta || \alpha) = (1 - \beta) \log \frac{1 - \beta}{\alpha} + \beta \log \frac{\beta}{1 - \alpha}$. In the following section, we will exploit this approximation in our analysis of quantizer design.

III. CHARACTERIZATION OF OPTIMAL QUANTIZERS

Turning now to the decentralized setting, the primary challenge lies in the design of the quantization rules ϕ_n. Any fixed set of quantization rules ϕ_n yields a sequence of compressed data $U_n = \phi_n(X_n)$, to which the classical theory can be applied. In the decentralized setting, we write

$$f_{\phi_n}^i(u) := \mathbb{P}_i(\phi_n(X_n) = u), \qquad i = 0, 1$$

for the distributions of the compressed data, conditioned on the hypothesis. We say that a quantizer design is *stationary* if the rule ϕ_n is independent of n; in this case, we simplify the notation to f_ϕ^1 and f_ϕ^0. In addition, we define the KL divergences $\mu_\phi^1 := D(f_\phi^1 || f_\phi^0)$ and $\mu_\phi^0 := D(f_\phi^0 || f_\phi^1)$. Moreover, let G_ϕ denote the analogue of the function G in equation (8), defined using $\mu_\phi^i, i = 0, 1$. In this section, we describe how—by exploiting Wald's approximation for sequential problems—it is possible to provide an asymptotic characterization of the optimal cost of any stationary quantization rule.

Approximate quantizer design: Given a fixed stationary quantizer ϕ, Wald's approximation (8) suggests the following strategy for solving approximating the cost of sequential detection strategy. For a given set of errors α and β, first assign the values of thresholds $a = a(\alpha, \beta)$ and $b = b(\alpha, \beta)$ using approximation (6). Then use the quantity $G_\phi(\alpha, \beta)$ as an approximation to the true cost $J_\phi(a, b)$. This approximation essentially ignores the overshoot of the likelihood ratio L_N. Analyzing this overshoot to obtain a finer approximation has been a major theme in sequential analysis (cf. [9], [4], [7]). For the purpose of quantizer design, however, as we shall see, the approximation error incurred from ignoring the overshoot is only at most $O(c)$, whereas the choice of quantizer ϕ generally results in a change of the order $\Theta(c \log c^{-1})$.

Asymptotic analysis: The quantization problem involves an output space that consists of a finite set of discrete elements (e.g., $\mathcal{U} = \{0, 1\}$ for 1-bit quantization). Consequently, we assume without loss of generality that $\sup_{u \in \mathcal{U}} \log f_\phi^1(u) / f_\phi^0(u) < M$. The following lemma guarantees that the approximation G_ϕ is asymptotically exact up to a deviation $O(c)$, and provides a basis for characterizing the optimal cost:

Lemma 1. (a) *The error in the approximation* (8) *is bounded as*

$$|J_\phi(a, b) - G_\phi(\alpha, \beta)| \leq c\,M\left(\frac{\pi^0}{\mu_\phi^0} + \frac{\pi^1}{\mu_\phi^1}\right). \tag{9}$$

(b) *Define the optimal cost* $J_\phi^* = \inf_{a,b} J_\phi(a, b)$. *Then as* $c \to 0$, *we have*

$$J_\phi^* = \left(\frac{\pi^0}{\mu_\phi^0} + \frac{\pi^1}{\mu_\phi^1}\right) c \log \frac{1}{c} + O(c). \tag{10}$$

Proof: (a) We begin by bounding the error in the approximation (6). By definition of the stopping time N, we have either (i) $b \leq L_N \leq b + M$ or (ii) $a - M \leq L_N \leq a$. Consider all realizations $u_1, \ldots, u_n$ for which condition (i) holds; for any such sequence, we have

$$e^b P_0(u_1, \ldots, u_n) \leq P_1(u_1, \ldots, u_n)$$
$$e^{(b+M)} P_0(u_1, \ldots, u_n) \geq P_1(u_1, \ldots, u_n).$$

Taking a sum over all such realizations, using the definition of α and β, and performing some algebra yields the inequality $e^b \alpha \leq 1 - \beta \leq e^{b+M} \alpha$, or equivalently $b \leq b(\alpha, \beta) = \log \frac{1-\beta}{\alpha} \leq b + M$. Similar reasoning for case (ii) yields $a - M \leq a(\alpha, \beta) = \log \frac{\beta}{1-\alpha} \leq a$. Now, note that

$$\mathbb{E}_0 L_N = \alpha \mathbb{E}_0[L_N | L_N \geq b] + (1 - \alpha)\mathbb{E}_0[L_N | L_N \leq a].$$

Conditioned on the event $L_N \in [b, b + M]$, we have $|L_N - b(\alpha, \beta)| \leq M$. Similarly, conditioned on the event $L_N \in [a - M, a]$, we have $|L_N - b(\alpha, \beta)| \leq M$. This yields $|\mathbb{E}_0 L_N - (-D(\alpha || 1 - \beta))| \leq M$. Similar reasoning yields $|\mathbb{E}_1 L_N - D(1 - \beta || \alpha)| \leq M$.

3

(b) By part (a), it suffices to establish the asymptotic behavior (10) for the quantity $\widetilde{J}_\phi(a,b) = \inf_{\alpha,\beta} G_\phi(\alpha,\beta)$, where the infimum is taken among pairs of realizable error probabilities (α,β). Moreover, we only need to consider the asymptotic regime $\alpha + \beta \to 0$, since the error probabilities α and β vanish as $c \to 0$. It is simple to see that $D(1-\beta||\alpha) = \log(1/\alpha) + o(1)$, and $D(1-\alpha||\beta) = \log(1/\beta) + o(1)$. Hence, $\inf_{\alpha,\beta} G_\phi(\alpha,\beta)$ can be expressed as

$$\inf_{\alpha,\beta}\left\{\pi^0\alpha + \pi^1\beta + c\pi^0\frac{\log(1/\beta)}{\mu_\phi^0} + c\pi^1\frac{\log(1/\alpha)}{\mu_\phi^1}\right\} + o(c).$$
(11)

This infimum is achieved at $\alpha^* = \frac{c\pi^1}{\mu_\phi^1\pi^0}$ and $\beta^* = \frac{c\pi^0}{\mu_\phi^0\pi^1}$; these error probabilities can be realized (for c small) by using a sufficiently large threshold $b > 0$ and small threshold $a < 0$. Plugging these quantities into equation (11) yields $\inf_{\alpha,\beta} G_\phi(\alpha,\beta) = \left(\frac{\pi^0}{\mu_\phi^0} + \frac{\pi^1}{\mu_\phi^1}\right)c\log\frac{1}{c} + O(c)$ as claimed. □

IV. SUB-OPTIMALITY OF STATIONARY DESIGNS

We now consider the structure of optimal quantizers. It was shown by Tsitsiklis [10] that optimal quantizers ϕ_n take the form of threshold rules based on the likelihood ratio $P_1(X_n)/P_0(X_n)$. Veeravalli et al [12], [11] asked whether these rules can be taken to be stationary; a problem that has remained open. In this section, we resolve this question with a negative answer. First, we provide a counterexample in which the optimal quantizer is not stationary. Next, we show that using a stationary quantizer can be suboptimal even in an asymptotic sense (i.e., as $c \to 0$).

Illustrative counterexample: We begin with a simple but concrete demonstration of the suboptimality of stationary designs. Consider a problem in which $X \in \mathcal{X} = \{1,2,3\}$ and the conditional distributions take the form

$$f^0(x) = \begin{bmatrix} \frac{8}{10} & \frac{1999}{10000} & \frac{1}{10000} \end{bmatrix} \text{ and } f^1(x) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Suppose that the prior probabilities are $\pi^1 = \frac{8}{100}$ and $\pi^0 = \frac{92}{100}$, and that the cost for each sample is $c = \frac{1}{100}$.

If we restrict to binary quantizers (i.e., $\mathcal{U} = \{0,1\}$), then there are only three possible quantizers:

1) Design A: $\phi_A(X_n) = 0 \iff X_n = 1$. As a result, the corresponding distribution for U_n is specified by $f^0_{\phi_A}(u_n) = \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \end{bmatrix}$ and $f^1_{\phi_A}(u) = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$.
2) Design B: $\phi_B(X_n) = 0 \iff X_n \in \{1,2\}$. The corresponding distribution for U_n is given by $f^0_{\phi_B}(u) = \begin{bmatrix} \frac{9999}{10000} & \frac{1}{10000} \end{bmatrix}$ and $f^1_{\phi_B}(u) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$

Method	$J_A(0.08)$	$J_B(0.08)$	$J_C(0.08)$	$J_*(0.08)$
Cost	0.0567	0.0532	0.0800	0.0528

TABLE I. Numerically computed costs for the three stationary designs J_A, J_B and J_C, for the mixed design J_*.

3) Design C: $\phi_C(X_n) = 0 \iff X_n \in \{1,3\}$. The corresponding distribution for U_n is specified by $f^0_{\phi_C} \sim \begin{bmatrix} \frac{8001}{10000} & \frac{1999}{10000} \end{bmatrix}$ and $f^1_{\phi_C}(u) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$.

Now consider the three stationary strategies, each of which uses only one fixed design A, B or C. For any given stationary quantization rule ϕ, we have a classical centralized sequential problem, for which the optimal cost (achieved by a SPRT) can be computed using a dynamic-programming procedure [13], [1]. Accordingly, for each stationary strategy, we compute the optimal cost function J for 10^6 points on the p-axis by performing 300 updates of Bellman's equation [2]. In all cases, the difference in cost between the 299th and 300th updates is less than 10^{-6}. Let J_A, J_B and J_C denote the optimal cost function for sequential tests using all A's, all B's, and all C's, respectively. When evaluated at $\pi^1 = 0.08$, these computations yield the numerical values shown in Table I. Finally, we consider a non-stationary rule obtained by applying design A for only the first sample, and applying design B for the remaining samples. Again using Bellman's equation, we find that the cost for this design is $J_* = 0.0528$, which is better than any of the stationary strategies.

Asymptotic suboptimality of stationary designs: We now prove that there is always a range of prior probabilities for which stationary quantizer designs are suboptimal. Our result stems from the following observation: the form of the approximation (8) dictates that in order to achieve small cost we need to choose a quantizer ϕ for which the KL divergences $D(f^1_\phi||f^0_\phi)$ and $D(f^0_\phi||f^1_\phi)$ are both as large as possible. However, due to the asymmetry of the KL divergence, these maxima are not necessarily achieved by a single quantizer ϕ. In such settings, it can be possible to construct a non-stationary quantizer with better cost than a stationary design. Recall the notation, for any fixed quantizer, $\mu^0_\phi := D(f^0_\phi||f^1_\phi)$ and $\mu^1_\phi := D(f^1_\phi||f^0_\phi)$.

Proposition 2. *Let ϕ_1 and ϕ_2 be any two quantizers. If the following inequalities hold*

$$\mu^0_{\phi_1} < \mu^0_{\phi_2} \text{ and } \mu^1_{\phi_1} > \mu^1_{\phi_2} \quad (12)$$

4

then there exists a non-empty interval $(A, B) \subseteq (0, +\infty)$ such that as $c \to 0$,

$$J^*_{\phi_1} \leq J^*_{\phi_1,\phi_2} \leq J^*_{\phi_2} \quad if \quad \frac{\pi^0}{\pi^1} \leq A$$

$$J^*_{\phi_1,\phi_2} < \min\{J^*_{\phi_1}, J^*_{\phi_2}\} - \Theta(c \log c^{-1}) \quad if \quad \frac{\pi^0}{\pi^1} \in (A, B)$$

$$J^*_{\phi_1} \geq J^*_{\phi_1,\phi_2} \geq J^*_{\phi_2} \quad if \quad \frac{\pi^0}{\pi^1} \geq B,$$

where $J^*_{\phi_i}$ denotes the optimal cost of the stationary design based on the quantizer ϕ_i, and $J^*_{\phi_1,\phi_2}$ denotes the optimal cost of a sequential test that alternates between using ϕ_1 and ϕ_2 on odd and even samples respectively.

Proof: According to Lemma 1, we have

$$J^*_{\phi_i} = \left(\frac{\pi^0}{\mu^0_{\phi_i}} + \frac{\pi^1}{\mu^1_{\phi_i}} \right) c \log c^{-1} + O(c), \quad i = 0, 1 \quad (13)$$

Now consider the sequential test that applies quantizers ϕ_1 and ϕ_2 alternately to odd and even samples. Furthermore, this test considers two samples at a time. Let $f^0_{\phi_1\phi_2}$ and $f^1_{\phi_1\phi_2}$ denote the induced conditional probability distributions, jointly on the odd-even pairs of quantized variables. From the additivity of the KL divergence and assumption (12), there holds:

$$D(f^0_{\phi_1\phi_2}||f^1_{\phi_1\phi_2}) = \mu^0_{\phi_1} + \mu^0_{\phi_2} > 2\mu^0_{\phi_1} \quad (14a)$$
$$D(f^1_{\phi_1\phi_2}||f^0_{\phi_1\phi_2}) = \mu^1_{\phi_1} + \mu^1_{\phi_2} < 2\mu^1_{\phi_1}. \quad (14b)$$

Clearly, the cost of the proposed sequential test is an upper bound for $J^*_{\phi_1,\phi_2}$. Furthermore, the gap between this upper bound and the true optimal cost is no more than $O(c)$. Hence, as in the proof of Lemma 1, as $c \to 0$, the optimal cost $J^*_{\phi_1,\phi_2}$ can be written as

$$\left(\frac{2\pi^0}{\mu^0_{\phi_1} + \mu^0_{\phi_2}} + \frac{2\pi^1}{\mu^1_{\phi_1} + \mu^1_{\phi_2}} \right) c \log c^{-1} + O(c). \quad (15)$$

From equations (13) and (15), simple calculations yield the claim with

$$A = \frac{\mu^0_{\phi_1}(\mu^1_{\phi_1} - \mu^1_{\phi_2})(\mu^0_{\phi_1} + \mu^0_{\phi_2})}{\mu^1_{\phi_1}(\mu^1_{\phi_1} + \mu^1_{\phi_2})(\mu^0_{\phi_2} - \mu^0_{\phi_1})}$$

$$< B = \frac{\mu^0_{\phi_2}(\mu^1_{\phi_1} - \mu^1_{\phi_2})(\mu^0_{\phi_1} + \mu^0_{\phi_2})}{\mu^1_{\phi_2}(\mu^1_{\phi_1} + \mu^1_{\phi_2})(\mu^0_{\phi_2} - \mu^0_{\phi_1})}.$$

$\square$

Remarks: Let us return to the example provided in the previous subsection. Note that the two quantizers ϕ_A and ϕ_B satisfy assumption (12), since $D(f^0_{\phi_B}||f^1_{\phi_B}) = 0.4045 < D(f^0_{\phi_A}||f^1_{\phi_A}) = 0.45$ and $D(f^1_{\phi_B}||f^0_{\phi_B}) = 2.4337 > D(f^1_{\phi_A}||f^0_{\phi_A}) = 0.5108$. As a result, there exist priors for which a sequential test using stationary quantizer design (either ϕ_A, ϕ_B or ϕ_C

for all samples) is not optimal.

Blockwise stationary designs: Restricting to the class of stationary designs, despite its possible suboptimality, is computationally desirable. In this section, we consider designs that are *blockwise stationary* in the sense that there exists some some natural number T such that $\phi_{T+1} = \phi_1, \phi_{T+2} = \phi_2$, and so on. For each T, let C_T denote the class of all blockwise stationary designs with period T. Thus, as T increases, we have a hierarchy of increasingly rich quantizer classes that will yield progessively better approximations to the optimal solution. It is well known [10] that optimal quantizers *when unrestricted* are necessarily likelihood-based threshold rules. Our preceding results imply that the thresholds need not be stationary (i.e., the threshold may differ from sample to sample). The following theorem addresses a partial converse to this issue:

Theorem 3. *Restricting to the class of blockwise stationary and deterministic decision rules, there exists an asymptotically optimal quantizer $\phi = (\phi_1, \ldots, \phi_T)$ such that ϕ_n $(n = 1, \ldots, T)$ are likelihood ratio rules.*

The key insight underlying this theorem is a quasiconcavity result of J^*_ϕ (modulo the $O(c)$ term) with respect to the vector of probabilities $f^i_\phi(u)$ $(i = 0, 1; u \in \mathcal{U})$ induced by the quantizer ϕ. A proof and further detail can be found in a technical report [6].

REFERENCES

[1] K. J. Arrow, D. Blackwell, and M. A. Girshick. Bayes and minimax solutions of sequential decision problems. *Econometrica*, 17(3/4):213–244, 1949.

[2] D.P. Bertsekas. *Dynamic programming and stochastic control*, volume 1. Athena Scientific, Belmont, MA, 1995.

[3] Y. C. Ho. Team decision problems and information structures. *Proceedings of the IEEE*, 68:644–654, 1980.

[4] T. L. Lai. Sequential analysis: Some classical problems and new challenges (with discussion). *Statist. Sinica*, 11:303–408, 2001.

[5] Y. Mei. *Asymptotically optimal methods for sequential change-point detection*. PhD thesis, California Institute of Technology, 2003.

[6] X. Nguyen, M. J. Wainwright, and M. I. Jordan. On optimal quantization rules in some sequential decision problems. Technical report, Dept of Statistics, UC Berkeley, May 2006.

[7] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, NY, 1994.

[8] A. N. Shiryayev. *Optimal stopping rules*. Springer-Verlag, 1978.

[9] D. Siegmund. *Sequential analysis*. Springer-Verlag, 1985.

[10] J. N. Tsitsiklis. On threshold rules in decentralized detection. In *Proc. 25th IEEE Conf. Decision Control*, pages 232–236, 1986.

[11] V. V. Veeravalli. Sequential decision fusion: theory and applications. *Journal of the Franklin Institute*, 336:301–322, 1999.

[12] V. V. Veeravalli, T. Basar, and H. V. Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Trans. Info. Theory*, 39(2):433–442, 1993.

[13] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Annals of Statistics*, 19:326–339, 1948.