Do Optimal Entropy-Constrained Quantizers Have a Finite or Infinite Number of Codewords?

András György, *Associate Member, IEEE*,
Tamás Linder, *Senior Member, IEEE*,
Philip A. Chou, *Senior Member, IEEE*, and Bradley J. Betts

***Abstract*—An entropy-constrained quantizer Q is optimal if it minimizes the expected distortion $D(Q)$ subject to a constraint on the output entropy $H(Q)$. In this correspondence, we use the Lagrangian formulation to show the existence and study the structure of optimal entropy-constrained quantizers that achieve a point on the lower convex hull of the operational distortion-rate function $D_h(R) = \inf_Q\{D(Q) : H(Q) \le R\}$. In general, an optimal entropy-constrained quantizer may have a countably infinite number of codewords. Our main results show that if the tail of the source distribution is sufficiently light (resp., heavy) with respect to the distortion measure, the Lagrangian-optimal entropy-constrained quantizer has a finite (resp., infinite) number of codewords. In particular, for the squared error distortion measure, if the tail of the source distribution is lighter than the tail of a Gaussian distribution, then the Lagrangian-optimal quantizer has only a finite number of codewords, while if the tail is heavier than that of the Gaussian, the Lagrangian-optimal quantizer has an infinite number of codewords.**

***Index Terms*—Difference distortion measures, entropy coding, infinite-level quantizers, Lagrangian performance, optimal quantization.**

I. Introduction

In the design of locally optimal entropy-constrained vector quantizers (ECVQs) from training data [1], it has been repeatedly observed that the number of codewords in a locally optimal ECVQ is bounded by a number that depends on the source and the target entropy. That is, the number of codewords does not increase even if the ECVQ design algorithm is initialized with a greater number of codewords, or if a greater number of training vectors is made available. In some sense, there is a natural number of codewords for a given source at a given rate.

The above observation suggests that optimal entropy-constrained quantizers may not necessarily have an infinite number of codewords. Of course, one anticipates this for sources with bounded support. The question is, do optimal entropy-constrained quantizers always have an infinite number of codewords when the source has an unbounded region of support?

In this correspondence, we answer this question for a large class of optimal entropy-constrained quantizers. To be precise, given $\lambda > 0$ we consider optimal ECVQs Q^* that minimize the Lagrangian performance $J(\lambda, Q) = D(Q) + \lambda H(Q)$, where $D(Q)$ and $H(Q)$ are the distortion and the entropy of Q, respectively. If Q^* is such a Lagrangian-optimal quantizer, it is also an optimal entropy-constrained quantizer whose distortion $D(Q^*)$ and output entropy $H(Q^*)$ achieve a point on the lower convex hull of the operational distortion-rate function

$$D_h(R) = \inf_Q\{D(Q) : H(Q) \le R\}.$$

Apart from their practical significance in quantizer design [1], Lagrangian-optimal quantizers studied in this correspondence are also of theoretical interest. The Lagrangian formulation of entropy-constrained quantization serves as a useful tool in the rigorous treatment of the high-rate theory of entropy-constrained quantization [2], [3] and it has important connections with the theory of fixed-slope lossy source coding [4], [5].

Our first result, Theorem 1, shows that under some mild conditions on the distortion measure, for any $\lambda > 0$ there always exists a quantizer minimizing $J(\lambda, Q)$. We then show in Theorem 2 that if the tail of the source distribution is sufficiently light (with respect to the distortion measure), then such a Lagrangian-optimal entropy-constrained quantizer has only a finite number of codewords. The converse result, Theorem 3, shows that for source distributions with slightly heavier tail, a Lagrangian-optimal entropy-constrained quantizer must have an infinite number of codewords.

In particular, for the squared error distortion measure these results imply that the Gaussian distribution is a breakpoint. If the tail of the source distribution is lighter than the tail of a Gaussian distribution, then the Lagrangian-optimal entropy-constrained quantizer has only a finite number of codewords, while for distributions with tail heavier than the Gaussian, the Lagrangian-optimal quantizer must have an infinite number of codewords. For the Gaussian distribution there exists a critical value of the quantizer rate such that for rates less than this critical value, the Lagrangian-optimal quantizer has a finite number of codewords, and for rates higher than the critical value, the Lagrangian-optimal quantizer has infinitely many codewords.

II. Preliminaries

A vector quantizer Q can be described by the following mappings and sets: an *encoder* $\alpha : \mathbb{R}^k \to \mathcal{I}$, where $\mathcal{I}$ is a countable index set, an associated measurable partition $\mathcal{S} = \{S_i;\ i \in \mathcal{I}\}$ of $\mathbb{R}^k$ such that $\alpha(x) = i$ if $x \in S_i$, a *decoder* $\beta : \mathcal{I} \to \mathbb{R}^k$, and an associated reproduction *codebook* $\mathcal{C} = \{\beta(i);\ i \in \mathcal{I}\}$. The overall quantizer $Q : \mathbb{R}^k \to \mathcal{C}$ is

$$Q(x) = \beta(\alpha(x)).$$

Without loss of generality, we assume that the *codewords* (or codevectors) $\beta(i),\ i \in \mathcal{I}$, are all distinct. If $\mathcal{I}$ is finite with N elements, we take $\mathcal{I} = \{1, \ldots, N\}$ and call Q an N-level quantizer. Otherwise, $\mathcal{I}$ is taken to be the set of all positive integers and Q is called an infinite-level quantizer. To define a quantizer Q, we will sometimes write $Q \equiv (\alpha, \beta)$. Note that Q is also uniquely defined by the partition $\mathcal{S}$ and codebook $\mathcal{C}$ via the rule

$$Q(x) = \beta(i) \quad \text{if and only if} \quad x \in S_i.$$

We suppose a nonnegative measurable *distortion measure* $d : \mathbb{R}^k \times \mathbb{R}^k \to [0, +\infty)$. For an $\mathbb{R}^k$-valued random vector X with distribution μ, the distortion of Q is measured by the expectation

$$\begin{aligned} D(Q) &\triangleq E\{d(X, \beta(\alpha(X)))\} \\ &= E\{d(X, Q(X))\} \\ &= \int_{\mathbb{R}^k} d(x, Q(x))\, d\mu(x). \end{aligned}$$

Manuscript received June 20, 2002; revised May 23, 2003. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Cambridge, MA, August 1998 and at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

A. György is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada, on leave from the Computer and Automation Research Institute of the Hungarian Academy of Sciences, Lágymányosi u. 11, Budapest, Hungary, H-1111 (e-mail: gyorgy@mast.queensu.ca).

T. Linder is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (email: linder@mast.queensu.ca).

P. A. Chou is with the Microsoft Corporation, Redmond, WA 98052 USA (email: pachou@microsoft.com).

B. J. Betts is with the NASA Ames Research Center, Moffet Field, CA 94035, USA (email: bbetts@email.arc.nasa.gov).

Communicated by R. Zamir, Associate Editor for Source Coding.
Digital Object Identifier 10.1109/TIT.2003.819340

0018-9448/03$17.00 © 2003 IEEE

The entropy-constrained *rate* of Q is the entropy of its output $Q(X)$

$$H(Q) \triangleq H(Q(X)) = H(\alpha(X)) = -\sum_{i\in\mathcal{I}} \Pr\{X \in S_i\} \log \Pr\{X \in S_i\}$$

where $\log$ denotes base 2 logarithm. A vector quantizer Q whose rate is measured by $H(Q)$ is called an *entropy-constrained vector quantizer* (ECVQ).

Unless otherwise stated, we always assume that the partition cell probabilities $\Pr\{X \in S_i\} = \mu(S_i)$, $i \in \mathcal{I}$, are all positive. One can always redefine Q on a set of probability zero (by possibly reducing the number of cells) to satisfy this requirement.

For any $R \geq 0$ let $D_h(R)$ denote the lowest possible distortion of any quantizer with output entropy not greater than R. This function, which we call the *operational distortion-rate function*, is formally defined by

$$D_h(R) \triangleq \inf_Q \{D(Q) : H(Q) \leq R\}$$

where the infimum is taken over all finite or infinite-level vector quantizers whose entropy is less than or equal to R. If there is no Q with finite distortion and entropy $H(Q) \leq R$, then we formally define $D_h(R) = +\infty$. Any Q that achieves $D_h(R)$ in the sense that $H(Q) \leq R$ and $D(Q) = D_h(R)$ is called an *optimal* ECVQ.

The Lagrangian formulation of entropy-constrained quantization defines for each value of a parameter $\lambda > 0$ the *Lagrangian performance* of a quantizer Q by

$$J(\lambda, Q) \triangleq D(Q) + \lambda H(Q).$$

The optimum Lagrangian performance is given by

$$J^*(\lambda) \triangleq \inf_Q J(\lambda, Q) = \inf_Q \{D(Q) + \lambda H(Q)\} \tag{1}$$

where the infimum is taken over all finite or infinite-level quantizers Q.

Any quantizer Q that achieves the infimum in (1) is called a *Lagrangian-optimal* quantizer. It is easy to see that if Q is Lagrangian-optimal for some $\lambda > 0$, then it is also an optimal ECVQ for its rate, i.e., if $J(\lambda, Q) = J^*(\lambda)$, then $D(Q) = D_h(H(Q))$. Moreover, if Q is Lagrangian-optimal, then $(H(Q), D(Q))$ is a point on the lower convex hull[1] of $D_h(R)$, and $-\lambda$ is the slope of a line that supports the lower convex hull and passes through this point.

Conversely, if Q is an optimal ECVQ such that $(H(Q), D(Q))$ is a point on the lower convex hull of $D_h(R)$, then there exists a $\lambda > 0$ such that $J(\lambda, Q) = J^*(\lambda)$, i.e., Q is Lagrangian-optimal. Therefore, the class of Lagrangian-optimal quantizers can be characterized as the class of optimal ECVQs that achieve the operational distortion-rate function $D_h(R)$ at rates where $D_h(R)$ coincides with its lower convex hull. Note that since $D_h(R)$ is not necessarily convex (see, e.g., [7]), in general, not all points of $D_h(R)$ are achievable by a Lagrangian-optimal quantizer.

III. Results

It is well known [1] that a Lagrangian-optimal quantizer must have an encoder that maps an input x to its "nearest" codeword, where the distance to the codeword is penalized by λ times the negative log probability of the partition cell associated with the codeword. This "generalized nearest neighbor" condition forms the basis of the iterative ECVQ design algorithm in [1]. The condition is formalized in the following lemma which is crucial in our development.

[1]The lower convex hull of $D_h(R)$ is the largest convex function $\widehat{D}_h(R)$ such that $\widehat{D}_h(R) \leq D_h(R)$ for all $R \geq 0$; see, e.g., [6].

Lemma 1: Let $Q \equiv (\alpha, \beta)$ be an arbitrary quantizer with partition cell probabilities $p_i = \mu(S_i) = \Pr\{\alpha(X) = i\}$, codewords $c_i = \beta(i)$, $i \in \mathcal{I}$, and finite Lagrangian performance $J(\lambda, Q) < +\infty$ for some $\lambda > 0$. Let the encoder α' be defined for all $x \in \mathbb{R}^k$ by

$$\alpha'(x) = \arg\min_{i\in\mathcal{I}} \Big(d(x, c_i) - \lambda \log p_i \Big) \tag{2}$$

(ties are broken arbitrarily), and set $Q' \equiv (\alpha', \beta)$. Then

$$J(\lambda, Q') \leq J(\lambda, Q)$$

where equality holds only if

$$d(x, \beta(\alpha(x))) - \lambda \log p_{\alpha(x)} = \min_{i\in\mathcal{I}} \Big(d(x, c_i) - \lambda \log p_i \Big) \tag{3}$$

for μ-almost all x.

The lemma implies that if $J(\lambda, Q) = J^*(\lambda)$, then a Lagrangian-optimal ECVQ must use the generalized nearest neighbor encoding rule (2) with probability 1. For the sake of completeness we give the proof of the lemma below.

Proof of Lemma 1: First note that the minimum in (2) exists for all x even if Q is an infinite-level quantizer, and so α' is well defined if a particular rule for breaking ties is set. Indeed, since $\lim_{i\to\infty}(-\log p_i) = \infty$ for infinite-level quantizers, we have

$$d(x, c_1) - \lambda \log p_1 < d(x, c_i) - \lambda \log p_i$$

for all i large enough, and hence for any $x \in \mathbb{R}^k$ the minimum

$$\min_{i\in\mathcal{I}} (d(x, c_i) - \lambda \log p_i)$$

is achieved by some $i \in \mathcal{I}$. Therefore,

$$d(x, \beta(\alpha'(x))) - \lambda \log p_{\alpha'(x)} = \min_{i\in\mathcal{I}} (d(x, c_i) - \lambda \log p_i).$$

Hence, defining $p'_i = \Pr\{\alpha'(X) = i\}$, we can write

$$\begin{aligned} J(\lambda, Q) &= E\{d(X, Q(X)) - \lambda \log p_{\alpha(X)}\} \qquad (4)\\ &= \sum_{j\in\mathcal{I}} \int_{S_j} \Big(d(x, c_j) - \lambda \log p_j \Big) \mu(dx) \\ &\geq \sum_{j\in\mathcal{I}} \int_{S_j} \min_{i\in\mathcal{I}} \Big(d(x, c_i) - \lambda \log p_i \Big) \mu(dx) \\ &= E\{d(x, \beta(\alpha'(X))) - \lambda \log p_{\alpha'(X)}\} \\ &= E\{d(x, \beta(\alpha'(X))) - \lambda \log p'_{\alpha'(X)}\} + \lambda E\left\{ \log \frac{p'_{\alpha'(X)}}{p_{\alpha'(X)}} \right\} \\ &= J(\lambda, Q') + \lambda \sum_{i\in\mathcal{I}} p'_i \log \frac{p'_i}{p_i} \qquad (5) \end{aligned}$$

from which the lemma follows since

$$\sum_{i\in\mathcal{I}} p'_i \log \frac{p'_i}{p_i} \geq 0$$

by the divergence inequality [8].

It is easy to see that the first inequality becomes an equality if and only if (3) holds for μ-almost all x, so a necessary condition for $J(\lambda, Q) = J(\lambda, Q')$ is that (3) holds for μ-almost all x. $\square$

Our first result shows the existence of Lagrangian-optimal quantizers for any $\lambda > 0$ under mild conditions on the distortion measure. Here and throughout the correspondence, $\|x\|$ denotes the usual Euclidean norm of $x \in \mathbb{R}^k$.

Theorem 1: Assume that for any $x \in \mathbb{R}^k$ the nonnegative distortion measure $d(x,y)$ is a lower semicontinuous function of y such that for any $y' \in \mathbb{R}^k$, $d(x,y') \le \liminf_{\|y\|\to\infty} d(x,y)$. Then for any $\lambda > 0$ there is a Lagrangian-optimal quantizer, i.e., there exists Q such that

$$D(Q) + \lambda H(Q) = J^*(\lambda).$$

The proof of Theorem 1 is deferred to the Appendix . The basic idea is to consider a sequence of quantizers with Lagrangian performance converging to the optimum. It is shown that there exists a subsequence of these quantizers whose codewords and cell probabilities converge, respectively, to a set of codewords and corresponding probabilities, which then can by used to define a "limit" quantizer via the generalized nearest neighbor rule (2). This limit quantizer is then shown to be optimal.

It is worth noting that $J^*(\lambda)$ is finite for all $\lambda > 0$ if there exists $y \in \mathbb{R}^k$ such that $E\{d(X,y)\} < +\infty$. In particular, for the squared error distortion measure $d(x,y) = \|x-y\|^2$ a sufficient (but not necessary) condition for the finiteness of $J^*(\lambda)$ is that $E\{\|X\|^2\} < +\infty$.

The conditions of the theorem are clearly satisfied if $d(x,y)$ is a difference distortion measure $d(x,y) = \rho(\|x-y\|)$, where $\rho(t), t \ge 0$ is a nonnegative, monotone increasing, and continuous function. Next, we consider such distortion measures and show that if the tail of the distribution of X is sufficiently light, then the Lagrangian-optimal quantizer has only a finite number of codewords. In the theorem $f(t) = o(g(t))$ means $\lim_{t\to+\infty} f(t)/g(t) = 0$.

Theorem 2: Assume a difference distortion measure $d(x,y) = \rho(\|x-y\|)$, where $\rho : [0,+\infty) \to [0,+\infty)$ is monotone increasing and continuous. For some $\lambda > 0$ let Q be a Lagrangian-optimal quantizer achieving $J^*(\lambda) < +\infty$. If for some $\epsilon > 0$

$$\Pr\{\|X\| \ge t\} = o\left(2^{-\rho((1+\epsilon)t)/\lambda}\right)$$

then Q has a finite number of codewords.

Proof: Let $\{c_i; i \in \mathcal{I}\}$ and $\{S_i; i \in \mathcal{I}\}$ be the codebook and partition of Q. To exclude pathological cases, we assume that the cell probabilities $p_i = \mu(S_i) = \Pr\{X \in S_i\}$ are positive for all $i \in \mathcal{I}$. (Any countable collection of cells with probability zero can be merged with a cell of positive probability without affecting the quantizer's performance.)

First we "regularize" the partition cells. For each i, define $\bar{S}_i$ by

$$\bar{S}_i = \{x : d(x,c_i) - \lambda \log p_i \le d(x,c_j) - \lambda \log p_j \text{ for all } j \in \mathcal{I}\}.$$

By Lemma 1, $\bar{S}_i$ contains μ-almost all x's in S_i, and hence $p_i \le \mu(\bar{S}_i)$. (In particular, $\bar{S}_i$ is not empty.) Since $d(x,y)$ is continuous, $\bar{S}_i$ is closed. Now for any $x \in \bar{S}_i$

$$d(x,c_i) - \lambda \log p_i \le d(x,c_1) - \lambda \log p_1.$$

In particular, for an $x_i \in \bar{S}_i$ closest (in Euclidean distance) to the origin

$$d(x_i,c_i) - \lambda \log p_i \le d(x_i,c_1) - \lambda \log p_1.$$

But $d(x_i,c_i) \ge 0$ and $p_i \le \mu(\bar{S}_i) \le \Pr\{\|X\| \ge \|x_i\|\}$, so that

$$-\lambda \log \Pr\{\|X\| \ge \|x_i\|\} \le d(x_i,c_1) - \lambda \log p_1$$

or, equivalently

$$\Pr\{\|X\| \ge \|x_i\|\} \ge p_1 2^{-d(x_i,c_1)/\lambda}.$$

Now by the triangle inequality and the monotonicity of ρ

$$d(x_i,c_1) = \rho(\|x_i - c_1\|) \le \rho(\|x_i\| + \|c_1\|).$$

Suppose $\sup_{i\in\mathcal{I}} \|x_i\| = +\infty$. Then we can pick $\|x_i\|$ sufficiently large so that the above bound gives $d(x_i,c_1) \le \rho((1+\epsilon)\|x_i\|)$, which, in turn, implies

$$\Pr\{\|X\| \ge \|x_i\|\} \ge p_1 2^{-\rho((1+\epsilon)\|x_i\|)/\lambda}.$$

On the other hand, if $P\{\|X\| \ge t\} = o\left(2^{-\rho((1+\epsilon)t)/\lambda}\right)$, then for $\|x_i\|$ sufficiently large we must have

$$P\{\|X\| \ge \|x_i\|\} < p_1 2^{-\rho((1+\epsilon)\|x_i\|)/\lambda}$$

a contradiction. Consequently, there must exist a finite $T > 0$ such that $\|x_i\| \le T$ for all $i \in \mathcal{I}$. Thus, to show that Q is a finite-level quantizer we only need to show that there can be only a finite number of partition cells with $\|x_i\| \le T$. Suppose, to the contrary, that $\|x_i\| \le T$ for all $i \in \mathcal{I}$ and $\mathcal{I}$ is countably infinite. Then we must have for all $i = 1, 2, \ldots$ that

$$\begin{aligned} d(x_i,c_i) - \lambda \log p_i &\le d(x_i,c_1) - \lambda \log p_1 \\ &\le \rho(T + \|c_1\|) - \lambda \log p_1 \end{aligned}$$

which is a contradiction since $\lim_{i\to\infty}(-\log p_i) = +\infty$. Hence, $\mathcal{I}$ must be finite, which proves the theorem. □

Note that if $\rho(t)$ converges to a finite limit as $t \to +\infty$, then $\lim_{t\to+\infty} 2^{-\rho((1+\epsilon)t)/\lambda} > 0$, and so the tail condition of the theorem is satisfied for *any* source distribution. Thus, for such a bounded distortion measure, the Lagrangian-optimal ECVQ always has a finite number of codewords.

The preceding proof also shows that regardless of the tails of the source distribution, a Lagrangian-optimal ECVQ is *locally finite* in the sense that the number of partition cells that intersect any bounded subset of $\mathbb{R}^k$ is finite. To be more precise, we can claim that all Lagrangian-optimal ECVQs that satisfy the generalized nearest neighbor condition of Lemma 1 for all x are locally finite. Indeed, for such quantizers, $S_i \subset \bar{S}_i$ for all $i \in \mathcal{I}$, and so the last part of the proof shows that any ball $\{x : \|x\| \le T\}$ can intersect only a finite number of cells S_i.

The next result is a converse to Theorem 2 for convex difference distortion measures. In the theorem, $f(t) = \Omega(g(t))$ means that there is a constant $c > 0$ such that $f(t) \ge c g(t)$ for all sufficiently large t.

Theorem 3: Assume a difference distortion measure $d(x,y) = \rho(\|x-y\|)$, where $\rho : [0,+\infty) \to [0,+\infty)$ is strictly increasing and convex. For some $\lambda > 0$, let Q be a Lagrangian-optimal quantizer achieving $J^*(\lambda) < +\infty$. If for some $0 < \epsilon < 1$

$$\Pr\{\|X\| > t\} = \Omega\left(2^{-\rho((1-\epsilon)t)/\lambda}\right)$$

then Q has infinitely many codewords.

Proof: The basic idea of the proof is simple: Suppose Q with N codewords minimizes $J(\lambda, Q) = D(Q) + \lambda H(Q)$. We create a new quantizer Q' with $N + 1$ codewords by splitting a cell of Q into two new cells. Splitting a cell reduces distortion, but increases entropy. The tail condition implies that if N is finite, then an appropriate split gives $D(Q) - D(Q') > \lambda(H(Q') - H(Q))$. Thus, $J(\lambda, Q') < J(\lambda, Q)$, so Q cannot be optimal.

To give a formal proof, we assume without loss of generality that $\rho(0) = 0$ (adding a constant to the distortion measure does not affect quantizer optimality).

Given $y \in \mathbb{R}^k$ and $0 < \theta < \pi/2$, let $C(y, \theta)$ denote the circular cone with half-angle θ and vertex at the origin defined by

$$C(y, \theta) = \{x : \langle x, y\rangle \geq \|x\| \, \|y\| \cos\theta\}$$

where $\langle x, y\rangle$ denotes the usual inner product in $\mathbb{R}^k$. Clearly, given any $0 < \theta < \pi/2$, there exists a finite collection of $M = M(\theta)$ vectors $\{y_1, \ldots, y_M\}$ such that $\{C(y_1, \theta), \ldots, C(y_M, \theta)\}$ cover $\mathbb{R}^k$, i.e.,

$$\mathbb{R}^k = \bigcup_{j=1}^{M} C(y_j, \theta).$$

Let Q be an N-level quantizer with codebook $\{c_1, \ldots, c_N\}$ and partition $\{S_1, \ldots, S_N\}$ such that $J(\lambda, Q) = J^*(\lambda)$. Since the sets $S_i \cap C(y_j, \theta)$ cover $\mathbb{R}^k$, the union bound gives

$$\Pr\{\|X\| > t\} \leq \sum_{i=1}^{N}\sum_{j=1}^{M} \Pr\{\|X\| > t, X \in S_i, X \in C(y_j, \theta)\}.$$

Since

$$\limsup_{t\to+\infty} \frac{\Pr\{\|X\| > t\}}{2^{-\rho((1-\epsilon)t)/\lambda}} > 0$$

by the tail condition, there exist i and j (which depend on θ) such that

$$\limsup_{t\to+\infty} \frac{\Pr\{\|X\| > t, X \in S_i, X \in C(y_j, \theta)\}}{2^{-\rho((1-\epsilon)t)/\lambda}} > 0. \tag{6}$$

Now define

$$S \triangleq \{x : \|x\| > t, x \in S_i, x \in C(y_j, \theta)\}$$

(the dependence of S on θ and t is suppressed in the notation). In the Appendix , we prove that if $0 < \delta < 1$ is fixed, and $\theta > 0$ is sufficiently small, then we can choose $c \in \mathbb{R}^k$ (which depends on θ and t just as S does) such that for all sufficiently large t and all $x \in S$

$$d(x, c_i) - d(x, c) \geq \rho(t(1-\delta)). \tag{7}$$

Fix $K > 0$ and choose t_K such that $\rho(t_K) \geq K$ (this is always possible since $\lim_{t\to+\infty}\rho(t) = +\infty$). We have $\rho(a) - \rho(b) \geq \rho(a - b)$ for all $a > b \geq 0$ since ρ is convex and $\rho(0) = 0$, and, hence, for all sufficiently large t

$$\begin{aligned} d(x, c_i) - d(x, c) - K &\geq \rho(t(1-\delta)) - \rho(t_K) && (8)\\ &\geq \rho(t(1-\delta) - t_K)\\ &= \rho\left(t\left(1 - \delta - \frac{t_K}{t}\right)\right). && (9)\end{aligned}$$

Therefore, if $K > 0$ and $0 < \delta < 1$ are fixed, then there exists $\theta > 0$ such that for all sufficiently large t and for all $x \in S$

$$d(x, c_i) - d(x, c) - K \geq \rho(t(1-\delta)). \tag{10}$$

The asymptotic relation (6) and an argument similar to (8) and (9) imply that if we choose δ such that $0 < \delta < \epsilon$, then there exists t arbitrarily large such that

$$\rho(t(1-\delta)) \geq -\lambda \log \mu(S).$$

For such t and all $x \in S$, (10) gives

$$d(x, c_i) - d(x, c) + \lambda \log \mu(S) \geq K. \tag{11}$$

Now let Q' be the $(N+1)$-level quantizer with codebook $\{c_1, \ldots, c_N, c\}$ and partition $\{S'_1, \ldots, S'_{N+1}\}$, where $S'_j = S_j$ for $j = 1, \ldots, N$, $j \neq i$, $S'_i = S_i \setminus S$, and $S'_{N+1} = S$. Since Q and Q' have $N - 1$ common partition cells and codewords, from (11) there exists arbitrarily large t such that

$$\begin{aligned} &J(\lambda, Q) - J(\lambda, Q')\\ &\quad= \int_{S_i}\Big(d(x, c_i) - \lambda\log\mu(S_i)\Big)\mu(dx)\\ &\qquad - \int_{S_i\setminus S}\Big(d(x, c_i) - \lambda\log\mu(S_i\setminus S)\Big)\mu(dx)\\ &\qquad - \int_{S}\Big(d(x, c) - \lambda\log\mu(S)\Big)\mu(dx)\\ &\quad= \int_{S}\Big(d(x, c_i) - d(x, c) + \lambda\log\mu(S)\Big)\mu(dx)\\ &\qquad - \lambda\mu(S_i)\log\mu(S_i) + \lambda\mu(S_i\setminus S)\log\mu(S_i\setminus S)\\ &\quad\geq \mu(S)K - \lambda\mu(S_i)\log\mu(S_i) + \lambda\mu(S_i\setminus S)\log\mu(S_i\setminus S)\\ &\quad= \mu(S)K - \mu(S)\lambda\log\mu(S_i)\\ &\qquad - \lambda\mu(S_i\setminus S)\log\left(\frac{\mu(S_i\setminus S) + \mu(S)}{\mu(S_i\setminus S)}\right)\\ &\quad\geq \mu(S)\left[K - \lambda\frac{\mu(S_i\setminus S)}{\mu(S)}\log\left(1 + \frac{\mu(S)}{\mu(S_i\setminus S)}\right)\right]\end{aligned}$$

where the last equality holds since $S \subset S_i$. Note that

$$\lim_{t\to+\infty} \mu(S)/\mu(S_i\setminus S) = 0$$

since $\lim_{t\to+\infty}\mu(S) = 0$. Since

$$\lim_{u\to 0}(1/u)\log(1+u) = \log e$$

if we choose $K > \lambda\log e$, then there exists a large t such that the last expression is positive. Then $J(\lambda, Q) > J(\lambda, Q')$, which contradicts the optimality of Q. □

Note that the conditions on $d(x, y)$ in Theorems 2 and 3 are satisfied for the rth power distortion measures $d(x, y) = \|x - y\|^r$ if $r \geq 1$. In particular, both theorems hold for the squared error distortion measure. In this case, we obtain that the Gaussian distribution is a breakpoint: For distributions with tail lighter than the tail of a Gaussian distribution (including distributions with bounded support), the optimal entropy-constrained quantizer must have only a finite number of codewords, and for distributions with tail heavier than that of the Gaussian, the optimal entropy-constrained quantizer has an infinite number of codewords.

The Gaussian case itself is of particular interest. For a Gaussian source, the results show that there is a critical value $\lambda^* > 0$ (and a corresponding critical rate $R^* > 0$) such that the Lagrangian-optimal quantizer Q has a finite number of codewords if $\lambda > \lambda^*$ (i.e., $H(Q) < R^*$), and it has an infinite number of codewords if $\lambda < \lambda^*$ (i.e., $H(Q) > R^*$).

Corollary 1: Let $d(x,y) = \|x-y\|^2$ and assume that X is Gaussian with covariance matrix K having largest eigenvalue $\gamma > 0$. Then for any $\lambda > 2\gamma \ln 2$, the Lagrangian-optimal ECVQ has a finite number of codewords, and for $\lambda < 2\gamma \ln 2$ the Lagrangian-optimal ECVQ has an infinite number of codewords.

The condition $\gamma > 0$ means that at least one component of X has nonzero variance. If X has independent Gaussian components with common variance $\sigma^2 > 0$, then $\gamma = \sigma^2$ in the theorem.

Proof: Since K is symmetric and nonnegative definite, there is an orthogonal matrix U that diagonalizes it: $UKU^t = \mathrm{diag}(\gamma_1, \ldots, \gamma_k)$ where the γ_i, $i = 1, \ldots, k$ are the (nonnegative) eigenvalues corresponding to the k orthogonal eigenvectors of K. Then $Y = UX$ has independent Gaussian components $Y_1, \ldots, Y_k$ with variance $\mathrm{Var}(Y_i) = \gamma_i$ for all i (some of which may be zero), so $Y_i = \sqrt{\gamma_i} Z_i$, where $Z = (Z_1, \ldots, Z_k)^t$ has independent Gaussian components with common unit variance. Note that we can also assume without loss of generality that the X_i (and so the Y_i and the Z_i) have zero mean. Since U is orthogonal, $\|Y\| = \|UX\| = \|X\|$. Setting $\gamma \triangleq \max(\gamma_1, \ldots, \gamma_k)$, we have for all $t > 0$

$$\begin{aligned}\Pr\{\|X\| > t\} &= \Pr\{\|Y\| > t\} \\ &= \Pr\left\{\sum_{i=1}^k \gamma_i Z_i^2 > t^2\right\} \\ &\le \Pr\left\{\sum_{i=1}^k \gamma Z_i^2 > t^2\right\} \\ &= \Pr\{\sqrt{\gamma}\|Z\| > t\}.\end{aligned}$$

But $\|Z\|$ has the chi distribution with k degrees of freedom with asymptotic tail probability given by

$$\lim_{t\to+\infty} \frac{\Pr\{\|Z\| > t\}}{t^{k-2}e^{-t^2/2}} = a_k \tag{12}$$

where a_k is a positive constant (see, e.g., [9]). Thus,

$$\limsup_{t\to+\infty} \frac{1}{t^2} \log \Pr\{\|X\| > t\} \le -\frac{1}{2\gamma \ln 2}$$

and, hence, if $\lambda > 2\gamma \ln 2$, then there exists an $\epsilon > 0$ such that

$$\Pr\{\|X\| > t\} = o\left(2^{-(1+\epsilon)^2 t^2/\lambda}\right).$$

Then, by Theorem 2, Q has only a finite number of codewords.

On the other hand, let j be an index such that $\gamma_j = \gamma$. Then

$$\begin{aligned}\Pr\{\|X\| > t\} &= \Pr\left\{\sum_{i=1}^k \gamma_i Z_i^2 > t^2\right\} \\ &\ge \Pr\{\sqrt{\gamma_j}|Z_j| > t\} = \Pr\{\sqrt{\gamma}|Z_j| > t\}.\end{aligned}$$

Using (12) with $k = 1$, we obtain

$$\liminf_{t\to+\infty} \frac{1}{t^2} \log \Pr\{\|X\| > t\} \ge -\frac{1}{2\gamma \ln 2}.$$

If $\lambda < 2\gamma \ln 2$, then

$$\Pr\{\|X\| > t\} = \Omega\left(2^{-(1-\epsilon)^2 t^2/\lambda}\right)$$

for some $1 > \epsilon > 0$, and Q must have infinitely many codewords by Theorem 3. □

APPENDIX

Proof of Theorem 1: Assume $J^*(\lambda)$ is finite; otherwise the statement is trivial. Let $\{Q_n \equiv (\alpha_n, \beta_n)\}_{n=1}^\infty$ be a sequence of quantizers such that $\lim_{n\to\infty} J(\lambda, Q_n) = J^*(\lambda)$. Assume, without loss of generality, the common index set $\mathcal{I} = \{1, 2, \ldots\}$ for all Q_n and denote the partition cell probabilities of Q_n by $\{p_1^{(n)}, p_2^{(n)}, \ldots\}$ and the corresponding codewords by $\{c_1^{(n)}, c_2^{(n)}, \ldots\}$ (hence, $p_i^{(n)} = \Pr\{\alpha_n(X) = i\}$ and $c_i^{(n)} = \beta_n(i)$). The assumption of the common index set implies that some of the cells $S_i^{(n)} = \{x : \alpha_n(x) = i\}$ may be empty with the corresponding $p_i^{(n)}$ being zero.

The following lemma is proved in [10].

Lemma 2: For $R > 0$, define the set of probability vectors C_R by the equation at the bottom of the page. Then C_R is compact under pointwise convergence.

Without loss of generality, we assume that for each Q_n the partition cells and codewords are indexed so that $p_i^{(n)} \ge p_{i+1}^{(n)}$ for all $i \ge 1$. Since $\lim_{n\to\infty} J(\lambda, Q_n) = J^*(\lambda)$, for all n large enough, we have

$$H(Q_n) \le J(\lambda, Q_n)/\lambda \le (J^*(\lambda) + 1)/\lambda.$$

Thus, if we set $R = (J^*(\lambda) + 1)/\lambda$, then for all n large enough

$$(p_1^{(n)}, p_2^{(n)}, \ldots) \in C_R.$$

Let $\overline{\mathbb{R}^k} = \mathbb{R}^k \cup \{\infty\}$ be the usual one-point compactification of $\mathbb{R}^k$ (see, e.g., [11]). Then by Lemma 2 and Cantor's diagonal method, we can pick a subsequence of $\{Q_n\}$, also denoted by $\{Q_n\}$ for convenience, such that for some $\bar{c}_1, \bar{c}_2, \ldots \in \overline{\mathbb{R}^k}$ and a probability vector $(p_1, p_2, \ldots)$ we have $\lim_{n\to\infty} c_i^{(n)} = \bar{c}_i$ and $\lim_{n\to\infty} p_i^{(n)} = p_i$ for all $i \ge 1$.

Now for all $i \in \mathcal{I}$, let $c_i = \bar{c}_i$ if $\bar{c}_i \in \mathbb{R}^k$, and choose $c_i \in \mathbb{R}^k$ in an arbitrary manner if $\bar{c}_i = \infty$. Define Q to be the quantizer with codewords $\{c_1, c_2, \ldots\}$ and encoder α given by

$$\alpha(x) = \arg\min_{i\in\mathcal{I}} \left(d(x, c_i) - \lambda \log p_i\right)$$

(ties are broken arbitrarily). Here we use the convention that $-\log p_i = +\infty$ if $p_i = 0$, so that α (and hence Q) is well defined.

In the remainder of the proof, we show that Q is a Lagrangian-optimal quantizer. First observe that the conditions on $d(x,y)$ imply that for any $i \in \mathcal{I}$ and $x \in \mathbb{R}^k$, $\liminf_{n\to\infty} d(x, c_i^{(n)}) \ge d(x, c_i)$. Hence, we obtain

$$\liminf_{n\to\infty} \left(d(x, c_i^{(n)}) - \lambda \log p_i^{(n)}\right) \ge d(x, c_i) - \lambda \log p_i$$

$$C_R = \left\{(p_1, p_2, \ldots) : \, p_i \ge 0 \text{ for all } i, p_1 \ge p_2 \ge \cdots, \, \sum_{i=1}^\infty p_i = 1, \, -\sum_{i=1}^\infty p_i \log p_i \le R\right\}.$$

which implies

$$J(\lambda, Q) \le E\Big\{\min_{i\in\mathcal{I}}\Big(d(X,c_i) - \lambda\log p_i\Big)\Big\}$$
$$\le E\Big\{\min_{i\in\mathcal{I}}\liminf_{n\to\infty}\Big(d(X,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)\Big\} \quad (13)$$

where the first inequality follows from the generalized nearest neighbor condition (see (4) and (5) in the proof of Lemma 1).

Let $i^*(x,n) \in \mathcal{I}$ denote an index such that

$$\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big) = d(x,c_{i^*(x,n)}^{(n)}) - \lambda\log p_{i^*(x,n)}^{(n)}\Big)$$

(recall from the proof of Lemma 1 that the minimum exists) and let n_j, $j = 1,2,\dots,$ be an increasing sequence of positive integers such that

$$\liminf_{n\to\infty}\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)$$
$$= \lim_{j\to\infty}\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n_j)}) - \lambda\log p_i^{(n_j)}\Big)$$

and the limit $i^*(x) \triangleq \lim_{j\to\infty} i^*(x,n_j)$ exists, where $i^*(x) = +\infty$ is allowed. Since the $p_i^{(n)}$, $i = 1,2,\dots,$ are decreasing, we have $p_i^{(n)} \le 1/i$ for all i and n, so if $i^*(x) = +\infty$, then

$$\begin{aligned}
&\lim_{j\to\infty}\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n_j)}) - \lambda\log p_i^{(n_j)}\Big)\\
&\quad= \lim_{j\to\infty}\Big(d(x,c_{i^*(x,n_j)}^{(n_j)}) - \lambda\log p_{i^*(x,n_j)}^{(n_j)}\Big)\\
&\quad\ge \lim_{j\to\infty}\lambda\log i^*(x,n_j)\\
&\quad= +\infty.
\end{aligned}$$

This implies

$$\min_{i\in\mathcal{I}}\liminf_{n\to\infty}\Big(d(x,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)$$
$$= \liminf_{n\to\infty}\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big) \quad (14)$$

(with both sides being equal to $+\infty$) since the right-hand side is always less than or equal to the left-hand side. On the other hand, if $i^*(x)$ is finite, then $i^*(x,n_j) = i^*(x)$ for all sufficiently large j, so for such j

$$\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n_j)}) - \lambda\log p_i^{(n_j)}\Big) = \min_{i\le i^*(x)}\Big(d(x,c_i^{(n_j)}) - \lambda\log p_i^{(n_j)}\Big)$$

and we obtain

$$\begin{aligned}
&\min_{i\in\mathcal{I}}\liminf_{n\to\infty}\Big(d(x,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)\Big\}\\
&\quad\le \min_{i\le i^*(x)}\liminf_{j\to\infty}\Big(d(x,c_i^{(n_j)}) - \lambda\log p_i^{(n_j)}\Big)\\
&\quad= \liminf_{j\to\infty}\min_{i\le i^*(x)}\Big(d(x,c_i^{(n_j)}) - \lambda\log p_i^{(n_j)}\Big)\\
&\quad= \liminf_{n\to\infty}\min_{i\in\mathcal{I}}\Big(d(x,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big).
\end{aligned} \quad (15)$$

Thus, (14) and (15) yield

$$E\Big\{\min_{i\in\mathcal{I}}\liminf_{n\to\infty}\Big(d(X,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)\Big\}$$
$$= E\Big\{\liminf_{n\to\infty}\min_{i\in\mathcal{I}}\Big(d(X,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)\Big\}.$$

Combining this with (13) shows that Q is a Lagrangian-optimal quantizer

$$\begin{aligned}
J(\lambda,Q) &\le E\Big\{\liminf_{n\to\infty}\min_{i\in\mathcal{I}}\Big(d(X,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)\Big\}\\
&\le \liminf_{n\to\infty} E\Big\{\min_{i\in\mathcal{I}}\Big(d(X,c_i^{(n)}) - \lambda\log p_i^{(n)}\Big)\Big\}\\
&\le \liminf_{n\to\infty} J(\lambda,Q_n)\\
&= J^*(\lambda)
\end{aligned}$$

where the second inequality follows from Fatou's lemma [11], and the third from the generalized nearest neighbor condition (see (4) and (5)). □

Proof of Inequality (7): Without loss of generality, we can assume that $y_j = (1,0,\dots,0)$. Let $(c_{i1},\dots,c_{ik})$ denote the components of c_i and define $c \in \mathbb{R}^k$ by

$$c = (t\cos\theta, c_{i2},\dots,c_{ik}).$$

For any $x = (x_1,\dots,x_k)$, we have $\|x - c_i\| = \sqrt{(x_1 - c_{i1})^2 + A}$ and $\|x - c\| = \sqrt{(x_1 - t\cos\theta)^2 + A}$, where

$$A = \sum_{l=2}^{k}(x_l - c_{il})^2.$$

Observe that if $x = (x_1,\dots,x_k) \in S$, then $x \in C(y_j,\theta)$ and $\|x\| > t$, implying $x_1 > t\cos\theta$. Also, if t is large enough, then $t\cos\theta > |c_{i1}|$. Hence, for all sufficiently large t and for all $x \in S$

$$\begin{aligned}
&d(x,c_i) - d(x,c)\\
&\quad= v\rho(\|x - c_i\|) - \rho(\|x - c\|)\\
&\quad= \rho\Big(\sqrt{(x_1 - c_{i1})^2 + A}\Big) - \rho\Big(\sqrt{(x_1 - t\cos\theta)^2 + A}\Big)\\
&\quad\ge \rho\Big(\sqrt{(x_1 - c_{i1})^2 + A} - \sqrt{(x_1 - t\cos\theta)^2 + A}\Big)
\end{aligned} \quad (16)$$

where the inequality holds since $\rho(a) - \rho(b) \ge \rho(a - b)$ for all $a > b \ge 0$ by the convexity of ρ and the assumption $\rho(0) = 0$. Also, $x \in S \subset C(y_j,\theta)$ implies $\sum_{l=2}^{k} x_l^2 \le x_1^2\tan^2\theta$. Therefore,

$$A \le 2\sum_{l=2}^{k} x_l^2 + 2\sum_{l=2}^{k} c_{il}^2 \le 2x_1^2\tan^2\theta + B$$

where $B = 2\sum_{l=2}^{k} c_{il}^2$ is a nonnegative constant. Since $\sqrt{a^2+u} - \sqrt{b^2+u}$ is a monotone decreasing function of $u > 0$ for any fixed $a > b \ge 0$, and ρ is monotone increasing, we can continue (16) as

$$\begin{aligned}
&\rho\Big(\sqrt{(x_1 - c_{i1})^2 + A} - \sqrt{(x_1 - t\cos\theta)^2 + A}\Big)\\
&\ge \rho\Bigg(\sqrt{(x_1 - c_{i1})^2 + 2x_1^2\tan^2\theta + B}\\
&\qquad - \sqrt{(x_1 - t\cos\theta)^2 + 2x_1^2\tan^2\theta + B}\Bigg)\\
&\ge \rho\Bigg(\sqrt{(x_1 - |c_{i1}|)^2 + 2x_1^2\tan^2\theta + B}\\
&\qquad - \sqrt{(x_1 - t\cos\theta)^2 + 2x_1^2\tan^2\theta + B}\Bigg) \quad (17)\\
&\ge \rho\Bigg(\sqrt{(t\cos\theta - |c_{i1}|)^2 + 2t^2\cos^2\theta\tan^2\theta + B}\\
&\qquad - \sqrt{2t^2\cos^2\theta\tan^2\theta + B}\Bigg)\\
&= \rho\Bigg(t\sqrt{\Big(\cos\theta - \frac{|c_{i1}|}{t}\Big)^2 + 2\sin^2\theta + \frac{B}{t^2}}\\
&\qquad - t\sqrt{2\sin^2\theta + \frac{B}{t^2}}\Bigg). \quad (18)
\end{aligned}$$

Here, the third inequality holds since the argument of ρ in (17) is a monotone increasing function of x_1 for $x_1 \ge 0$, as can be checked by differentiating with respect to x_1.

Given $0 < \delta < 1$, we can choose a small $\theta > 0$ such that for all sufficiently large $t > 0$

$$\sqrt{\Big(\cos\theta - \frac{|c_{i1}|}{t}\Big)^2 + 2\sin^2\theta + \frac{B}{t^2}} - \sqrt{2\sin^2\theta + \frac{B}{t^2}} \ge 1 - \delta.$$

Then (16) and (18) yield

$$d(x, c_i) - d(x, c) \geq \rho(t(1 - \delta))$$

as desired. □

REFERENCES

[1] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 31–42, Jan. 1989.

[2] R. M. Gray, T. Linder, and J. Li, "A Lagrangian formulation of Zador's entropy-constrained quantization theorem," *IEEE Trans. Inform. Theory*, vol. 48, pp. 695–707, Mar. 2002.

[3] R. M. Gray and T. Linder, "Mismatch in high rate entropy constrained vector quantization," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1204–1217, May 2003.

[4] E.-H. Yang, Z. Zhang, and T. Berger, "Fixed slope universal lossy data compression," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1465–1476, Sept. 1997.

[5] N. Merhav and I. Kontoyiannis, "Source coding exponents for zero-delay coding with finite memory," *IEEE Trans. Inform. Theory*, vol. 49, pp. 609–625, Mar. 2003.

[6] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[7] A. György and T. Linder, "Optimal entropy-constrained scalar quantization of a uniform source," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2704–2711, Nov. 2000.

[8] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York: Wiley, 1994, vol. 1.

[10] A. György and T. Linder, "On the structure of optimal entropy-constrained scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 48, pp. 416–427, Feb. 2002.

[11] R. M. Dudley, *Real Analysis and Probability*. New York: Chapman & Hall, 1989.

A Lower Bound for the Detection/Isolation Delay in a Class of Sequential Tests

Igor V. Nikiforov

***Abstract*—We address the problem of minimax detecting and isolating abrupt changes in random signals. The criterion of optimality consists in minimizing the maximum mean detection/isolation delay for a given maximum probability of false isolation and mean time before a false alarm. It seems that such a criterion has many practical applications especially for safety-critical applications, in monitoring dangerous industrial processes and also when the decision should be done in a hostile environment. The redundant strapdown inertial reference unit integrity monitoring problem is discussed. An asymptotic lower bound for the mean detection/isolation delay is given.**

***Index Terms*—Asymptotic optimality, lower bound, minimax change detection/isolation, navigation system integrity monitoring, recursive algorithm.**

NOMENCLATURE

t, n	Current time instants (discrete time).
$k+1$	Change time (fault onset time).
l, j	Type of change (type of fault).
K	Total number of hypotheses.
$\|X\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$	Norm of X.
N, M	Stopping (alarm) time.
ν	Final decision.
$\mathbb{E}(Y)$	Expectation of the random value Y.
$\mathbb{E}(Y \mid X_1, \ldots, X_t)$	Conditional expectation of the random value Y given $X_1, \ldots, X_t$.
$\Pr(B)$	Probability of the event B.
P, $f(x)$	Distribution and its density.
$\mathcal{N}(\theta, \Sigma)$	Normal law with mean vector θ and covariance matrix Σ.

I. INTRODUCTION

The problem of detecting and isolating abrupt changes in random signals has many important applications in signal processing and automatic control. Mathematically, it is the generalization of abrupt change detection (see results and references in [9], [19], [11], [1], [6], [7]) to the case of multiple $(K \geq 2)$ hypotheses. An optimal solution to the problem of abrupt change diagnosis (detection/isolation) and a nonrecursive algorithm that asymptotically attains the lower bound were obtained in [13] by using a minimax approach. The character feature of this approach is a pessimistic estimation of the detection/isolation delay ("worst case" mean detection/isolation delay) and an optimistic estimation of the probability of false isolation (it is assumed that the change occurs at the onset time to avoid the theoretical difficulties). A multiple hypothesis Shiryayev sequential probability ratio test by adopting a dynamic programming approach has been proposed by Malladi and Speyer in [10]. This algorithm minimize a certain Bayesian criterion that includes the measurement cost, the cost of a false alarm, and the cost of miss-alarm in the dynamic programming scheme. Next, Lai [8]

Manuscript received April 3, 2000; revised July 1, 2002. The material in this correspondence was presented in part at the 15th Triennial World Congress of the IFAC, Barcelona, Spain, July 2002.

The author is with LM2S, Université de Technologie de Troyes, BP 2060-10010, Troyes Cedex, France (e-mail: nikiforov@utt.fr).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2003.818398

0018-9448/03$17.00 © 2003 IEEE