Transaction Papers

Extremal Properties of Likelihood-Ratio Quantizers

John N. Tsitsiklis

***Abstract*—Let there be M hypotheses $H_1, \cdots, H_M$, and let Y be a random variable, taking values in a set $\mathcal{Y}$, with a different probability distribution under each hypothesis. A quantizer $\gamma: \mathcal{Y} \mapsto \{1, \cdots, D\}$ is applied to form a quantized random variable $\gamma(Y)$. We characterize the extreme points of the set of possible probability distributions of $\gamma(Y)$, as γ ranges over all quantizers. We then establish optimality properties of likelihood-ratio quantizers for a very broad class of quantization problems, including problems involving the maximization of an Ali–Silvey distance measure and the Neyman–Pearson variant of the decentralized detection problem.**

I. INTRODUCTION

SUPPOSE that there are M hypotheses $H_1, \cdots H_M$, and that Y is a random variable with a different probability distribution under each hypothesis. In classical detection theory [28], one observes one or more realizations of the random variable Y and attempts to infer the nature of the true hypothesis. In several contexts, however, practical considerations dictate that the observations must be quantized and statistical inferences are constrained to depend only on the quantized observations. We are then led to the problem of finding a quantizer which is optimal with respect to a performance measure of interest. This problem has been studied extensively in the quantization literature [12], [18], [1], [10], [16], [3]. It also arises in the area of decentralized detection [25], [23] whereby a set of sensors obtain some observations and transmit a summary of their observations to a fusion center that makes a final selection of one of the candidate hypotheses (see [24] for a survey and more references).

Throughout the literature on quantization and decentralized detection, there is a recurrent theme. In particular, for several specific choices of a performance criterion, it has been shown that likelihood ratio quantizers (LRQ's) are optimal [25], [10], [17], [13], [27], [24], [30]. Motivated by such results, this paper studies the geometry of the set of all quantizers, establishes some extremal properties of LRQ's, and derives some very broad conditions under which LRQ's are optimal.

The contribution of this paper is twofold. First, the results of a multitude of published papers are shown to be immediate consequences of a simple general principle. Second, a number of new results are derived.

Paper approved by the Editor for Random Access and Distributed Communications Systems. Manuscript received November 15, 1989; revised April 29, 1991. This work was supported by the ARO under Grant DAAL03-86-K-0171 and by the NSF under Grant ECS-8552419, with matching funds from Bell Communications Research and Du Pont. The paper was presented at the 29th IEEE Conference on Decision and Control, Honolulu, HI, December 1990.

The author is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 9209470.

Summary of the paper

In Section II, we define a quantizer as a function γ from the range of the random variable Y into a finite set $\{1, \cdots, D\}$. We also define randomized quantizers. To each quantizer γ, we associate a vector $q(\gamma)$ that describes the probability distribution of $\gamma(Y)$ under each hypothesis. We let $\overline{Q}$ be the set of all vectors $q(\gamma)$, and we use a result of Liapunoff to show that $\overline{Q}$ is convex and compact. As a corollary, we obtain a general result on the existence of optimal quantizers.

In Section III, we focus on the case of binary hypotheses. After a definition of likelihood ratio quantizers, we explore the geometry of the set $\overline{Q}$. We show that $q(\gamma)$ is an extreme point of $\overline{Q}$ if and only if γ is an LRQ with a certain "canonical" property. Furthermore, we characterize the extreme points of certain "sections" of the set $\overline{Q}$. As a corollary, we establish the optimality of LRQ's for a broad class of performance criteria.

In Section IV, we establish the optimality of LRQ's when the performance criterion is an Ali–Silvey distance measure [2], [18], thus generalizing the results of [17] and [13] and providing an alternative view of the results of [10] and [30].

In Section V, we generalize the results of Sections III–IV to the case of M hypotheses, for arbitrary M. In particular, we show that $q(\gamma)$ is an extreme point of $\overline{Q}$ if and only if there exists a sequence $\{\gamma_n\}$ of LRQ's (suitably defined) such that $q(\gamma_n)$ converges to $q(\gamma)$. We then derive some implications on the nature of optimal solutions to certain quantization problems.

Finally, in Section VI, we consider decentralized detection problems of the type introduced in [25]. We concentrate on a Neyman–Pearson variant of the problem and establish the optimality of LRQ's, thus providing an alternative derivation of results of [29] and [30].

II. QUANTIZERS

Let $\mathcal{Y}$ be some set endowed with a σ-field $\mathcal{F}$, and let $\mathcal{P}_1, \cdots, \mathcal{P}_M$ be M probability measures on the measurable space $(\mathcal{Y}, \mathcal{F})$. We associate each measure $\mathcal{P}_i$, $i = 1, \cdots, M$, with a hypothesis H_i on the distribution of a $\mathcal{Y}$-valued random variable Y. Accordingly, we use the notation $\Pr(A|H_i)$ to indicate the $\mathcal{P}_i$-measure of an event A.

Let D be a positive integer that will be held constant throughout the paper. We define a *deterministic quantizer* as an $\mathcal{F}$-measurable mapping $\gamma: \mathcal{Y} \mapsto \{1, \cdots, D\}$. We use Γ to denote the set of all deterministic quantizers.

If the quantized version $\gamma(Y)$ of the random variable Y is to be used for choosing between the hypotheses $H_1, \cdots, H_M$, then the distribution of $\gamma(Y)$ under each hypothesis becomes of interest. Since $\gamma(Y)$ is finite-valued, its distribution is specified by the finite set of scalars

$$q_d(\gamma|H_i) = \Pr(\gamma(Y) = d|H_i),\quad i = 1, \cdots, M, \quad d = 1, \cdots, D. \tag{2.1}$$

For $i = 1, \cdots, M$, and for any $\gamma \in \Gamma$, let

$$q(\gamma|H_i) = (q_1(\gamma|H_i), \cdots, q_D(\gamma|H_i)). \tag{2.2}$$

Thus, $q(\gamma|H_i)$ is a D-dimensional vector describing the probability distribution of $\gamma(Y)$ under hypothesis H_i. Finally, for any $\gamma \in \Gamma$, we define a vector $q(\gamma) \in \Re^{MD}$ by letting

$$q(\gamma) = (q(\gamma|H_1), \cdots, q(\gamma|H_M)). \tag{2.3}$$

Any two quantizers γ, $\gamma' \in \Gamma$ satisfying $q(\gamma) = q(\gamma')$ are equally helpful for the purpose of distinguishing between the different hypotheses. Thus, instead of studying quantizers directly, we can concentrate on the corresponding vectors $q(\gamma)$. Accordingly, we define

$$Q = \{q(\gamma)|\gamma \in \Gamma\}. \tag{2.4}$$

As is well known from Neyman–Pearson detection theory [28], randomization can improve performance in some detection problems. For this reason, we generalize our earlier definition, as follows. Let K be an arbitrary positive integer, and let $\gamma_1, \cdots, \gamma_K$ be some deterministic quantizers. Let $p_1, \cdots, p_K$ be some nonnegative scalars whose sum is equal to 1. Consider a random variable W (defined on some auxiliary probability space) which, under either hypothesis, takes the value k with probability p_k, and is statistically independent from Y. We define a function γ: $\mathcal{Y} \times \{1, \cdots, K\} \mapsto \{1, \cdots, D\}$ by letting

$$\gamma(Y, W) = \gamma_W(Y).$$

This function γ will be referred to as the (randomized) quantizer corresponding to $(\gamma_1, \cdots, \gamma_K, p_1, \cdots, p_K)$. Intuitively, it corresponds to picking at random and then applying one of the deterministic quantizers $\gamma_1, \cdots, \gamma_K$. Note that

$$\Pr(\gamma(Y, W) = d|H_i) = \sum_{k=1}^{K} p_k \Pr(\gamma_k(Y) = d|H_i),\quad \forall\, d, i. \tag{2.5}$$

In the sequel, we drop the argument W and write $\gamma(Y)$ instead of $\gamma(Y, W)$. However, it should be kept in mind that $\gamma(Y)$ is not always completely determined by Y.

Let $\overline{\Gamma}$ be the set of all (randomized) quantizers that can be constructed as in the preceding paragraph. By considering the case $K = 1$, it is seen that Γ can be identified with a subset of $\overline{\Gamma}$. In the sequel, we use the term "quantizer" to refer to elements of $\overline{\Gamma}$ and "deterministic quantizer" to refer to elements of Γ.

For any $\gamma \in \overline{\Gamma}$, we define $q_d(\gamma|H_i)$, $q(\gamma|H_i)$ and $q(\gamma)$, by means of (2.1)–(2.3). Then, if γ is the randomized quantizer corresponding to $(\gamma_1, \cdots, \gamma_K, p_1, \cdots, p_K)$, (2.5) implies that

$$q(\gamma) = \sum_{k=1}^{K} p_k q(\gamma_k). \tag{2.6}$$

Let

$$\overline{Q} = \{q(\gamma)|\gamma \in \overline{\Gamma}\}.$$

As is apparent from (2.6), we have

$$\overline{Q} = \mathrm{co}(Q), \tag{2.7}$$

where co$(\cdot)$ stands for the convex hull.

Existence of Optimal Quantizers

The following result is part of what is known as Lyapunoff's theorem.

Proposition 2.1: The sets Q and $\overline{Q}$ are compact.

Proposition 2.1 was proved in [14] for the case $D = 2$, and in [8] for a general value of D. A simpler proof, for the case $D = 2$, was subsequently given in [15]. In view of the evident importance of this result, and for completeness, a proof of Proposition 2.1 is provided in the Appendix. This proof consists of a simple modification of the argument in [15].

Let $J : \Re^{MD} \mapsto \Re$ be a continuous function, and suppose that the performance of a quantizer γ is captured by the value of $J(q(\gamma))$. An optimal quantizer can be defined as one that minimizes $J(q(\gamma))$ over the set $\overline{\Gamma}$; equivalently, we are dealing with the minimization of $J(q)$ over the set $\overline{Q}$. Then, Proposition 2.1 implies the existence of an optimal quantizer. This existence result remains valid if we only optimize over the set Γ of deterministic quantizers.

III. Extremal Quantizers—Binary Hypotheses

Throughout this and the next section, we assume that the number M of hypotheses is equal to 2. As mentioned in the Introduction, we are particularly interested in likelihood ratio quantizers (LRQ's for short), which we now define formally.

Let A be a measurable subset of $\mathcal{Y}$ such that $\mathcal{P}_1(A) = 1$, and such that $\mathcal{P}_2$ is absolutely continuous with respect to $\mathcal{P}_1$ on the set A. The (generalized) *likelihood ratio* is a measurable function $L : \mathcal{Y} \mapsto [0, \infty]$ satisfying

$$L(y) = \begin{cases} (d\mathcal{P}_2/d\mathcal{P}_1)(y), & \text{if } y \in A, \\ \infty, & \text{otherwise}, \end{cases} \tag{3.1}$$

where $d\mathcal{P}_2/d\mathcal{P}_1$ stands for (a version of) the Radon–Nikodym derivative of $\mathcal{P}_2$ with respect to $\mathcal{P}_1$ on the set A.

Definition 3.1: a) We define the *threshold set* T as the set of all vectors $t = (t_1, \cdots, t_{D-1}) \in [0, \infty]^{D-1}$, satisfying $0 \le t_1 \le \cdots \le t_{D-1} \le \infty$. For any $t \in T$, the associated intervals $I_1, \cdots, I_D$ are defined by $I_1 = [0, t_1]$, $I_2 = [t_1, t_2], \cdots, I_{D-1} = [t_{D-2}, t_{D-1}]$, $I_D = [t_{D-1}, \infty]$.
b) Let $t \in T$. We say that a quantizer $\gamma \in \Gamma$ is a *monotone LRQ* with threshold vector t, if

$$\Pr(\gamma(Y) = d \quad \text{and} \quad L(Y) \notin I_d|H_i) = 0, \qquad \forall\, d, i.$$

c) We say that a quantizer is an *LRQ* if there exists a permutation mapping $\pi: \{1, \cdots, D\} \mapsto \{1, \cdots, D\}$ such that $\pi \circ \gamma$ is a monotone LRQ. We use $\overline{\Gamma}_L$ to denote the set of all LRQ's.

With this definition, an LRQ is obtained from a monotone LRQ, after renaming the elements of $\{1, \cdots, D\}$. Suppose now that γ is a monotone LRQ. With our definition, the quantized value $\gamma(Y)$ is forced (modulo a zero measure event) to be equal to d whenever $L(Y)$ belongs to the interior of the set I_d. On the other hand, the value of $\gamma(Y)$ has some freedom when $L(Y)$ belongs to the common boundary of two intervals, that is, when $L(Y)$ is equal to some threshold. Also, notice that we allow different thresholds to be equal. Thus, we may have $t_{d-1} = t_d$, in which case the interval I_d has empty interior.

Extreme Points of Sections of $\overline{Q}$

We introduce some more notation. For any $\alpha \in \mathfrak{R}^D$, we define

$$\overline{Q}_\alpha = \{q(\gamma) \in \overline{Q} \mid q(\gamma|H_1) = \alpha\}.$$

Thus, each $\overline{Q}_\alpha$ is a "section" of the compact convex set $\overline{Q}$. It follows that $\overline{Q}_\alpha$ is compact and convex for every α.

Proposition 3.1: For any $\alpha \in \mathfrak{R}^D$, the following hold:

a) the set $\overline{Q}_\alpha$ has a finite number of extreme points;

b) if $q(\gamma)$ is an extreme point of Q_α, then γ is an LRQ.

Proof: Let $\overline{\Delta}$ be the set of all randomized quantizers whose range is $\{1, 2\}$ (instead of $\{1, \cdots, D\}$). Any $\delta \in \overline{\Delta}$ can be viewed as a statistical test for choosing between H_1 and H_2. For any $s \in [0, 1]$, let

$$R(s) = \min_{\delta \in \overline{\Delta}} \Pr(\delta(Y) = 1|H_2) \quad \text{subject to } \Pr(\delta(Y) = 1|H_1) = s. \tag{3.2}$$

In classical terminology, the mapping $s \mapsto 1 - R(1 - s)$ coincides with the receiver operating characteristic (ROC) curve. It is well known that the minimum in (3.2) is attained.

If $t \in [0, \infty]$, we say that an element δ of $\overline{\Delta}$ is a *likelihood-ratio test* (LRT) with threshold t if

$$\Pr(\delta(Y) = 1 \text{ and } L(Y) > t|H_i) = 0, \qquad i = 1, 2,$$
$$\Pr(\delta(Y) = 2 \text{ and } L(Y) < t|H_i) = 0, \qquad i = 1, 2.$$

The Neyman-Pearson lemma asserts the following. (A proof is naturally omitted.)

Lemma 3.1: There exists a nondecreasing function $\mathcal{X} : [0, 1] \mapsto [0, \infty]$ such that:

a) if δ attains the minimum in (3.2), then δ is an LRT with threshold $\mathcal{X}(s)$;

b) if $0 \le s < 1$, if δ is an LRT with threshold $\mathcal{X}(s)$, and if $\Pr(\delta(Y) = 1|H_1) = s$, then δ attains the minimum in (3.2);

c) if $s = 1$, and $\delta(y) = 1$ if and only if $L(y) < \infty$, then δ attains the minimum in (3.2).

The proof of the proposition rests on the following lemma:

Lemma 3.2: Fix some $\alpha = (\alpha_1, \cdots, \alpha_D) \in \mathfrak{R}^D$ such that $\overline{Q}_\alpha$ is nonempty. Fix also some vector $c = (c_1, \cdots, c_D) \in \mathfrak{R}^D$, and suppose that no two components of c are equal. Consider the problem

$$\text{minimize} \sum_{d=1}^{D} c_d \beta_d \quad \text{subject to } (\alpha, \beta) \in \overline{Q}_\alpha. \tag{3.3}$$

Then, the problem (3.3) has a unique solution $\beta^* = (\beta_1^*, \cdots, \beta_D^*)$, and β^* is completely determined by the ordering of the components of c. Furthermore, if $q(\gamma) = (\alpha, \beta^*)$, then γ is an LRQ.

Proof: Let us first prove the result for the special case where $c_1 > c_2 > \cdots > c_D$. Let $(\alpha, \beta) \in \overline{Q}_\alpha$, and choose some $\gamma \in \overline{\Gamma}$ such that $q(\gamma) = (\alpha, \beta)$. For $d = 1, \cdots, D$, we define $\delta^d \in \overline{\Delta}$ by

$$\delta^d(y) = \begin{cases} 1, & \text{if } \gamma(y) \le d, \\ 2, & \text{if } \gamma(y) > d. \end{cases} \tag{3.4}$$

Then,

$$\alpha_1 + \cdots + \alpha_d = \Pr(\delta^d(Y) = 1|H_1), \qquad \forall\, d,$$
$$\beta_1 + \cdots + \beta_d = \Pr(\delta^d(Y) = 1|H_2), \qquad \forall\, d.$$

Using the definition (3.2) of the function R we obtain

$$R(\alpha_1 + \cdots + \alpha_d) \le \beta_1 + \cdots + \beta_d. \qquad \forall\, d.$$

It follows that

$$\begin{aligned}\sum_{d=1}^{D} c_d \beta_d &= (c_1 - c_2)\beta_1 + (c_2 - c_3)(\beta_1 + \beta_2) \\ &\quad + \cdots + (c_{D-1} - c_D)(\beta_1 + \cdots + \beta_{D-1}) + c_D \\ &\ge (c_1 - c_2)R(\alpha_1) + (c_2 - c_3)R(\alpha_1 + \alpha_2) \\ &\quad + \cdots + (c_{D-1} - c_D)R(\alpha_1 + \cdots + \alpha_{D-1}) + c_D. \end{aligned} \tag{3.5}$$

Let us define β^* by letting $\beta_1^* + \cdots + \beta_d^* = R(\alpha_1 + \cdots + \alpha_d)$, for $d = 1, \cdots, D-1$, and $\beta_1^* + \cdots + \beta_D^* = 1$. We will first show that $(\alpha, \beta^*) \in \overline{Q}_\alpha$. Since this is the unique value of β^* for which (3.5) becomes an equality, it will follow that β^* is the unique optimal solution of the problem (3.3). Subsequently, we will show that if $q(\gamma) = (\alpha, \beta^*)$, then γ must be a monotone LRQ.

Let us define a threshold vector $t \in T$ by letting $t_d = \mathcal{X}(\alpha_1 + \cdots + \alpha_d)$, $d = 1, \cdots, D-1$, where $\mathcal{X}$ is the function of Lemma 3.1. Let γ be a monotone LRQ with threshold t, and suppose that the tie-breaking rule [when $L(y)$ is equal to some threshold] is chosen so that:

i) $\Pr(\gamma(Y) \le d|H_1) = \alpha_1 + \cdots + \alpha_d$, $\quad \forall\, d$;

ii) If $L(y) = t_d = \infty$, then $\gamma(y) > d$.

Then, for $d = 1, \cdots, D-1$, the function δ^d defined by (3.4) has the properties required in Lemma 3.1(b), (c). It follows that

$$\begin{aligned}\beta_1^* + \cdots + \beta_d^* &= \Pr(\gamma(Y) \le d|H_2) = \Pr(\delta^d(Y) = 1|H_2) \\ &= R(\alpha_1 + \cdots + \alpha_d), \qquad d = 1, \cdots, D-1\end{aligned}$$

Therefore, $q(\gamma) = (\alpha, \beta^*)$, which proves that $(\alpha, \beta^*) \in \overline{Q}_\alpha$.

Now, let us suppose that $q(\gamma) = (\alpha, \beta^*)$. We will show that γ is a monotone LRQ. Let $t_d = \mathcal{X}(\alpha_1 + \cdots + \alpha_d)$. Since $\beta_1^* + \cdots + \beta_d^* = R(\alpha_1 + \cdots + \alpha_d)$, Lemma 3.1(a) applied to δ^d implies that

$$\Pr(\gamma(Y) > d \text{ and } L(Y) < t_d|H_i) = 0, \qquad \forall\, i, d,$$
$$\Pr(\gamma(Y) \le d \text{ and } L(Y) > t_d|H_i) = 0, \qquad \forall\, i, d.$$

Furthermore, $t_1 \le \cdots \le t_{D-1}$. It then follows easily that γ is a monotone LRQ.

We have proved so far Lemma 3.2 for the special case where $c_1 > c_2 > \cdots > c_D$. The general case can be reduced to this special case by "renaming" of (that is, applying a permutation to) the elements of $\{1,\cdots,D\}$ so that the coefficients c_d become strictly decreasing. The only difference is that if $q(\gamma) = (\alpha, \beta^*)$, then γ will be a nonmonotone LRQ. (It will be monotone with respect to the renamed variables). Q.E.D.

Let G_α be the set of all (α, β^*) for which there exists a vector c with unequal components such that β^* is the unique optimal solution of the problem (3.3). Since each ordering of the components of c gives rise to exactly one β^*, it follows that G_α has at most $D!$ elements. Furthermore, Lemma 3.2 has established that if $q(\gamma) \in G_\alpha$, then γ is an LRQ. The proof of the proposition will be completed by showing that G_α contains the set of extreme points of $\overline{Q}_\alpha$.

Suppose that $\overline{Q}_\alpha$ has an extreme point $x = (\alpha, \overline{\beta})$ that does not belong to G_α. Then $\overline{Q}_\alpha$ has an extreme point $x = (\alpha, \overline{\beta})$ that does not belong to the convex hull $\mathrm{co}(G_\alpha)$ of G_α. Using the separating hyperplane theorem to separate x from $\mathrm{co}(G_\alpha)$, there exists some vector $c = (c_1, \cdots, c_D)$ such that

$$\sum_{d=1}^{D} c_d \overline{\beta}_d < \min_{(\alpha,\beta)\in G_\alpha} \sum_{d=1}^{D} c_d \beta_d . \tag{3.6}$$

By slightly perturbing the components of c, we can make them distinct while retaining the validity of (3.6). This contradicts the definition of G_α. The contradiction shows that G_α contains the set of extreme points of $\overline{Q}_\alpha$. Q.E.D.

Extreme Points of $\overline{Q}$

The following is our main result.

Proposition 3.2: If $\gamma \in \overline{\Gamma}$ and $q(\gamma)$ is an extreme point of $\overline{Q}$, then γ is an LRQ. Furthermore, there exists a deterministic LRQ γ' such that $q(\gamma') = q(\gamma)$.

Proof: Suppose that $q(\gamma) = (\alpha, \beta)$ is an extreme point of $\overline{Q}$. Then, $q(\gamma)$ is also an extreme point of $\overline{Q}_\alpha$, and Proposition 3.1 implies that γ is an LRQ. Furthermore, since $\overline{Q}$ is the convex hull of Q, it follows that $q(\gamma) \in Q$. Thus, there exists a deterministic $\gamma' \in \Gamma$ such that $q(\gamma') = q(\gamma)$. Using the already proved part of the proposition, γ' is an LRQ. Q.E.D.

The converse of Proposition 3.2 is not always true. For example, consider a threshold vector t, and suppose that for some d, $\Pr(L(Y) = t_d | H_i) > 0$ for some i. If γ is an LRQ with threshold vector t and uses randomization to resolve ties whenever $L(Y) = t_d$, it is evident that $q(\gamma)$ is a convex combination of two different elements of $\overline{Q}$, and $q(\gamma)$ is not an extreme point. This suggests that for $q(\gamma)$ to be an extreme point of $\overline{Q}$, γ should not use randomization for tie-breaking. This motivates the following definition.

Definition 3.2: We say that γ is a *canonical* LRQ if it is an LRQ and there exists a function $f: [0,\infty] \mapsto \{1,\cdots,D\}$ such that $\gamma(Y) = f(L(Y))$, with probability 1, under either hypothesis.

Note that if $L(Y) < \infty$ with probability 1, and if the probability distribution of $L(Y)$ is absolutely continuous with respect to Lebesgue measure (under either hypothesis), then every LRQ is a canonical LRQ. The following result provides a complete characterization of the extreme points of $\overline{Q}$. We omit the proof because the most important parts of our subsequent results do not depend on it.

Proposition 3.3: $q(\gamma)$ is an extreme point of $\overline{Q}$ if and only if γ is a canonical LRQ.

Optimality Properties of Likelihood Ratio Quantizers

Proposition 3.4: Suppose that $f : \overline{Q} \mapsto \Re$ is continuous and convex. Then:

a) there exists a canonical LRQ γ^* that maximizes $f(q(\gamma))$ over all $\gamma \in \overline{\Gamma}$; and

b) if f is also strictly convex, and if γ^* maximizes $f(q(\gamma))$ over all $\gamma \in \overline{\Gamma}$, then γ^* is a canonical LRQ.

Proof: a) By Corollary 32.3.1 of [19], the maximum of a convex function f over the compact convex set $\overline{Q}$ is attained at an extreme point. By Proposition 3.3, such an extreme point is of the form $q(\gamma^*)$ for some canonical LRQ γ^*.

b) In the strictly convex case, the value of f at any nonextreme point has to be smaller than the value of f at some extreme point (because nonextreme points can be expressed as convex combinations of extreme points). Thus, if $f(q(\gamma^*)) = \max_{q\in\overline{Q}} f(q)$, then $q(\gamma^*)$ is an extreme point of $\overline{Q}$ and, by Proposition 3.3, γ^* is a canonical LRQ. Q.E.D.

The next proposition applies to optimal quantization problems in which the function $f(\alpha, \beta)$ is only convex in β. It also applies to problems in which the value of $\alpha = q(\gamma|H_1)$ has to obey certain constraints. Such constraints arise in certain problems in the Neyman-Pearson type. An example will be seen in Section VI.

Proposition 3.5: Suppose that $f : \overline{Q} \mapsto \Re$ is continuous, and that for any $\alpha \in \Re^D$, the restriction of f on the set $\overline{Q}_\alpha$ is convex (that is, $f(\alpha, \beta)$ is convex in β). Let A be some closed subset of $\Re^D$.

a) There exists an LRQ γ^* such that $q(\gamma^*)$ maximizes $f(\alpha, \beta)$ subject to the constraints $(\alpha, \beta) \in \overline{Q}$ and $\alpha \in A$.

b) If $f(\alpha, \beta)$ is also strictly convex in β for each α, and if $q(\gamma^*)$ maximizes $f(\alpha, \beta)$ subject to the constraints $(\alpha, \beta) \in \overline{Q}$ and $\alpha \in A$, then γ^* is an LRQ.

Proof: a) Existence of an optimal solution follows because the set $\{(\alpha, \beta) \in \overline{Q} | \alpha \in A\}$ is compact, being the intersection of a compact and closed set. Let γ^* be such that $q(\gamma^*) = (\alpha^*, \beta^*)$ is an optimal solution. In particular, $\alpha^* \in A$.

Let us consider the auxiliary problem of maximizing $f(\alpha, \beta)$ subject to $(\alpha, \beta) \in \overline{Q}$ and $\alpha = \alpha^*$. Since $f(\alpha^*, \beta)$ is a convex function of β, it follows that there exists an extreme point $(\alpha^*, \overline{\beta})$ of $\overline{Q}_{\alpha^*}$ at which the maximum is attained. By the definition of $\overline{\beta}$, we have $f(\alpha^*, \overline{\beta}) \geq f(\alpha^*, \beta^*)$. Using the optimality of (α^*, β^*), the converse inequality also holds, and we conclude that $(\alpha^*, \overline{\beta})$ maximizes $f(\alpha, \beta)$ subject to the constraints $(\alpha, \beta) \in \overline{Q}$ and $\alpha \in A$. Since $(\alpha^*, \overline{\beta})$ is an extreme point of $\overline{Q}_{\alpha^*}$, Proposition 3.1 implies that there exists an LRQ $\overline{\gamma}$ such that $q(\overline{\gamma}) = (\alpha^*, \overline{\beta})$. Such a $\overline{\gamma}$ is clearly an optimal solution of the problem under consideration.

b) This follows similarly with part b) of Proposition 3.4. Q.E.D.

IV. Ali–Silvey Distance Measures

Ali–Silvey distance measures [2] (also known as f-divergences [5]) are general measures of the distance between two probability measures defined on the same measurable space. Such distance measures are useful in several contexts, including quantization problems; see [18], [10] and [16]. In this section, we show that a quantizer that maximizes an Ali–Silvey distance measure of the quantized distributions $q(\gamma|H_1)$ and $q(\gamma|H_2)$ can always be chosen to be an LRQ.

Let $f: [0,\infty] \mapsto \Re$ be a continuous convex function satisfying

$$\lim_{x\to\infty} \frac{f(x)}{x} = 0\,. \tag{4.1}$$

Then, the Ali–Silvey distance of two probability measures $\mathcal{P}_1,\ \mathcal{P}_2$ is defined as

$$D_f(\mathcal{P}_1,\mathcal{P}_2) = \int_{\{y\in\mathcal{Y}|L(y)<\infty\}} f(L(y))\,d\mathcal{P}_1(y)\,, \tag{4.2}$$

where $L(y)$ is the generalized likelihood ratio, as defined by (3.1).

If we employ a quantizer γ, the quantized random variable $\gamma(Y)$ has a different probability distribution $q(\gamma|H_i)$ under each hypothesis H_i, $i = 1,2$. The usefulness of a quantizer γ for discriminating between the two hypotheses (H_1,H_2) can be measured in terms of the Ali–Silvey distance $F(\gamma) = D_f(q(\gamma|H_1), q(\gamma|H_2))$. Let $\alpha = (\alpha_1,\cdots,\alpha_D) = q(\gamma|H_1)$ and $\beta = (\beta_1,\cdots,\beta_D) = q(\gamma|H_2)$. Then, using (4.2), we have

$$F(\gamma) = \sum_{\{d|\alpha_d\neq 0\}} \alpha_d f\left(\frac{\beta_d}{\alpha_d}\right) = \sum_{d=1}^{D} \alpha_d f\left(\frac{\beta_d}{\alpha_d}\right), \tag{4.3}$$

where the second equality follows once we adopt the convention $0\cdot f(\infty) = 0$. Let us use $J(\alpha,\beta)$ to denote the right-hand side of (4.3). The main result of this section follows.

Proposition 4.1: The problem of finding a quantizer that maximizes the Ali–Silvey measure $F(\gamma)$ has an optimal solution which is an LRQ.

Proof: As γ ranges over the set $\overline{\Gamma}$ of all quantizers, (α,β) ranges over the set $\overline{Q}$. Thus, finding a quantizer $\gamma\in\overline{\Gamma}$ that maximizes $F(\gamma)$ over the set $\overline{\Gamma}$ is equivalent to maximizing $J(\alpha,\beta)$ over the set $\overline{Q}$. Using (4.1), we see that J is a continuous function. Furthermore, since f is assumed convex, it is clear that $J(\alpha,\beta)$ is a convex function of β, for any fixed α. The result follows from Proposition 3.5(a). Q.E.D.

Some historical comments are in order. In [10], an iterative algorithm is given which , given any quantizer, produces a new quantizer with larger or equal value of the Ali–Silvey distance. No matter how the algorithm is initialized, the algorithm always produces LRQ's. Thus, the argument of [10] implicitly contains a proof that, if an optimal quantizer exists, then there exists an LRQ which is optimal. However, the derivation in [10] rests on an assumption that the function f is twice differentiable and strictly convex. This excludes, for example, the case where we want to maximize the variational distance between the two conditional distributions $\gamma(Y)$, because we have to let $f(x) = |x-1|$, which is neither strictly convex nor differentiable. In contrast, we are only assuming the f is continuous and convex. Furthermore, we believe that the algorithmic derivation in [10] does not expose the simple reasons for which Proposition 4.1 is true. Independently, from [10], the optimality of LRQ's was established in [17] for the special case $f(x) = 1/x$, using a direct argument. Following the lines of the argument in [17], [13] established the same result for the special case $f(x) = -x^s$, $s\in[0,1]$. Finally, in independent work, Proposition 4.1 was proved in [30], under the assumption that f is twice differentiable, and the measures $\mathcal{P}_1,\ \mathcal{P}_2$ are absolutely continuous.

Examples

Kullback–Liebler Divergence: When $f(x) = -\log x$, the corresponding Ali–Silvey distance is the Kullback–Liebler divergence, which plays a prominent role in Neyman-Pearson hypothesis testing. For concrete example [23], suppose that N sensors receive i.i.d. samples $Y_1,\cdots,Y_N$ of a random variable Y. The sensors transmit quantized values $\gamma(Y_1),\cdots,\gamma(Y_N)$ to a fusion center.[1] Then, the fusion center solves a Neyman–Pearson hypothesis testing problem to decide in favor of one of the underlying hypotheses. When N is large, the probability of error by the fusion center can be approximated by $e^{-NF(\gamma)}$, where F is defined by (4.3) with $f(x) = -\log x$. This leads to the problem of finding a quantizer γ that maximizes $F(\gamma)$.

The function f is convex and satisfies (4.1). On the other hand, $f(0) = \infty$, the continuity assumption on f is not satisfied, and the existence of an optimal quantizer is not automatically guaranteed. Let us assume, however, that $\int_{\mathcal{Y}} \log L(y)\,d\mathcal{P}_1(y) = c < \infty$. We then have $J(\alpha,\beta) = \sum_{d=1}^{D} \alpha_d f(\beta_d/\alpha_d) \le c$, for any $(\alpha,\beta)\in\overline{Q}$. (This is because quantization cannot increase the value of the Kullback–Liebler divergence.) From this, it follows easily that $J(\alpha,\beta)$ is continuous on the set $\overline{Q}$. We conclude that an optimal quantizer exists and an optimal quantizer can be chosen to be an LRQ.

We can actually obtain an even stronger conclusion, as follows. It is easily shown that $J(\alpha,\beta)$ is a convex function of (α,β) (just check the Hessian matrix for nonnegative definiteness). Then, Proposition 3.4(a) shows that there exists an optimal quantizer which is a canonical LRQ. Because of the symmetry of the problem, there exists an optimal quantizer which is a monotone canonical LRQ.

Chernoff's Exponent: Let s be a constant in $(0,1)$, and consider the case where $f(x) = -x^s$. Accordingly, let

$$J(\alpha,\beta;s) = -\sum_{i=1}^{D} \alpha_i^{1-s}\beta_i^s\,.$$

Again, it is easily checked that this function is convex in (α,β); and by Proposition 3.4(a), there exists an optimal canonical LRQ. Let us define

$$J(\alpha,\beta) = \sup_{s\in(0,1)} J(\alpha,\beta;s)\,.$$

This quantity is associated with the Chernoff bound on the probability of error in hypothesis testing [6]. Thus, the problem of maximizing $J(\alpha,\beta)$ over the set $\overline{Q}$ is of definite interest [3]. (Its relevance to decentralized detection problems was shown in [23]; see also [13]). This is the same as maximizing

[1] We are restricting the sensors to use the same quantizer. It has been shown in [23] that this results to no loss of optimality, asymptotically as $N\to\infty$.

$J(\alpha,\beta,s)$ over $\overline{Q} \times (0,1)$. Assume that the maximum is attained at some (α,β^*,s^*). Then, any $(\overline{\alpha},\overline{\beta},s^*)$ is also an optimal solution provided that $(\overline{\alpha},\overline{\beta})$ maximizes $J(\alpha,\beta,s^*)$ over the set $\overline{Q}$. It follows again from Proposition 3.4(a) that there exists an optimal quantizer which is a monotone canonical LRQ.

V. The Case of Multiple Hypotheses

The results of Sections III and IV can be partially generalized to the case of multiple hypotheses, provided that the concept of an LRQ is suitably modified. Indeed, in this section, we generalize Proposition 3.3, by providing a characterization of the exposed (cf Definition 5.2 below) and of the extreme points of the set $\overline{Q}$.

We still use the model and the notation of Section II. Let $\mathcal{P} = \mathcal{P}_1 + \cdots + \mathcal{P}_M$. Notice that each $\mathcal{P}_i$ is absolutely continuous with respect to $\mathcal{P}$. We define $L_i(y) = (d\mathcal{P}_i/d\mathcal{P})(y)$.

For every $d = 1,\cdots,D$, and $i = 1,\cdots,M$, let there be given a coefficient a_{id}. Let $a_d = (a_{1d},\cdots,a_{Md})$ and $a = (a_1,\cdots,a_D)$. Let us consider quantizers of the form

$$\begin{aligned}\gamma(y) &= \arg\min_d \sum_{i=1}^M a_{id} L_i(y), \qquad \text{w.p.1}\,,\\ &= \arg\min_d a_d^T L(y), \qquad\qquad (5.1)\end{aligned}$$

where $L(y) = (L_1(y),\cdots,L_M(y))$, and the superscript T denotes transpose. Equation (5.1) generalizes the structure of optimal statistical tests in M-ary hypothesis testing. It is also a natural structure for quantization problems in the presence of a finite number of alternative hypotheses [10].

We notice that (5.1) does not define a unique quantizer because we have not provided a tie-breaking rule. Furthermore, the class of quantizers of the form (5.1) is too general. For example, by letting $a_d = 0$ for all d, we see that any quantizer $\gamma \in \overline{\Gamma}$ is of the form (5.1). We will thus concentrate on quantizers of the form (5.1) for which a tie-breaking rule is unnecessary.

Definition 5.1: A quantizer $\gamma \in \overline{\Gamma}$ is called an *unambiguous likelihood quantizer* (ULQ, for short) if it is of the form (5.1) and the set of y's for which there is a tie in (5.1) has zero $\mathcal{P}$-measure. Formally, for any $d' \neq d''$,

$$\mathcal{P}\left(\left\{y \in \mathcal{Y} | a_{d'}^T L(y) = a_{d''}^T L(y) = \min_d a_d^T L(y)\right\}\right) = 0\,.$$

A simple criterion for a quantizer of the form (5.1) to be unambiguous is available under the following assumption.

Assumption 5.1: The joint probability distribution of the random vector $(L_1(Y),\cdots,L_{M-1}(Y))$ is absolutely continuous with respect to Lebesgue measure, under either hypothesis.

Lemma 5.1: Let Assumption 5.1 hold. Let γ be a quantizer of the form (5.1) and suppose that the vectors a_d are distinct. Then, γ is an unambiguous LRQ.

Proof: Suppose that $a_d \neq a_{d'}$. Using the equality $L_1(y) + \cdots + L_M(y) = 1$, we see that we have a tie [that is, $a_d^T L(y) = a_{d'}^T L(y)$] if and only if the vector $(L_1(y),\cdots,L_{M-1}(y))$ satisfies a (nontrivial) linear equation. Equivalently, if and only if this vector belongs to a subset of $\Re^{M-1}$ of Lebesgue measure zero. Thus, under Assumption 5.1, a tie occurs with zero probability under any hypothesis, and γ is unambiguous. Q.E.D.

We will now establish an extremal property of ULQ's. We need one more definition.

Definition 5.2: Let C be a convex subset of $\Re^n$ and let $x \in C$. We say that x is an *exposed* point of C, if there exists some $c \in \Re^n$ such that $c^T x < c^T y$ for every $y \in C$ different than x.

It is evident that the exposed points of a convex set are extreme points, but the converse is not always true. However, Straszewicz's theorem [19] asserts that the set of extreme points of a closed convex set is the closure of the set of exposed points. We will use this fact later.

The following result is a counterpart of Proposition 3.3.

Proposition 5.1: $q(\gamma)$ is an exposed point of $\overline{Q}$ if and only if γ is an ULQ.

Proof: Let us fix a vector c with components c_{id}, $i = 1,\cdots,M$, $d = 1,\cdots,D$. We introduce an auxiliary Bayesian decision problem. We assume that each hypothesis H_i has the same prior probability. Furthermore, once we observe Y, we have to make a decision $\gamma(Y) \in \{1,\cdots,D\}$, and we incur a penalty of Mc_{id} if the true hypothesis is H_i and our decision is d. It is clear that the expected cost of a decision rule $\gamma \in \overline{\Gamma}$ is equal to $\sum_{i,d} c_{id} q_d(\gamma|H_i)$. In particular, minimizing the expected cost is the same as minimizing $\sum_{i,d} c_{id} x_{id}$ over all $x \in \overline{Q}$.

We derive the solution of the Bayesian decision problem. Using a standard argument, γ is optimal if and only if $\gamma(Y)$ minimizes (w.p.1) the conditional expectation of the cost, conditioned on Y; that is,

$$\gamma(Y) = \arg\min_d \sum_{i=1}^M c_{id} \Pr(H_i|Y), \qquad \text{w.p.1}\,. \tag{5.2}$$

Using Bayes' rule, $\Pr(H_i|Y)$ is proportional to $L_i(Y)$, and (5.2) becomes

$$\gamma(Y) = \arg\min_d \sum_{i=1}^M c_{id} L_i(Y), \qquad \text{w.p.1}\,. \tag{5.3}$$

Thus, if $x = q(\gamma)$, we have that x minimizes $\sum_{i,d} c_{id} x_{id}$ over the set $\overline{Q}$ if and only if γ satisfies (5.3).

Suppose that x^* is an exposed point of $\overline{Q}$. Then, there exist coefficients c_{id} such that x^* is the unique minimizer of $\sum_{i,d} c_{id} x_{id}$ over the set $\overline{Q}$. Suppose that $q(\gamma^*) = x^*$. Then, γ^* satisfies (5.3). Furthermore, if γ satisfies (5.3), then $q(\gamma) = x^*$. This shows that $q(\gamma)$ is the same for all γ that satisfy (5.3). It follows that the probability of a tie in (5.3) is equal to zero under any hypothesis. Thus, γ^* is an ULQ.

Conversely, if γ^* is an ULQ, then for some choice of coefficients c_{id}, γ^* satisfies (5.3), and the probability of a tie is zero. Let $x^* = q(\gamma^*)$. Then, x^* minimizes $\sum_{i,d} c_{id} x_{id}$ over the set $\overline{Q}$. Using the fact that γ^* is unambiguous, and by reversing the argument in the preceding paragraph, it follows that x^* must be the unique minimizer. It follows that $q(\gamma^*)$ is an exposed point of $\overline{Q}$. Q.E.D.

Using Straszewicz's theorem, we obtain the following.

Corollary 5.1: $q(\gamma)$ is an extreme point of $\overline{Q}$ if and only if there exists a sequence $\{\gamma_n\}$ of ULQ's such that $q(\gamma) = \lim_{n\to\infty} q(\gamma_n)$.

By comparing Corollary 5.1 and Proposition 3.3, we can assert that, for the case of two hypotheses, we have that γ is a canonical LRQ if and only if there exists a sequence of $\{\gamma_n\}$ ULQ's such that $q(\gamma_n)$ converges to $q(\gamma)$. (This fact can also be verified by a simple direct argument.)

The following example illustrates the manner in which an extreme point of $\overline{Q}$ can fail to be an exposed point of $\overline{Q}$. Suppose that $M = D = 3$ and that Assumption 5.1 holds. Let ℓ be a positive scalar, and let ϵ be a positive parameter. Let

$$g_\epsilon(y) = \min\{L_1(y) - \ell, \epsilon L_2(y), \epsilon L_3(y)\}.$$

We define a quantizer γ_ϵ, $\epsilon > 0$, by

$$\gamma_\epsilon(y) = \begin{cases} 1, & \text{if } g_\epsilon(y) = L_1(y) - \ell, \\ 2, & \text{if } g_\epsilon(y) = \epsilon L_2(y), \\ 3, & \text{if } g_\epsilon(y) = \epsilon L_3(y). \end{cases}$$

It is seen that as long as $\epsilon > 0$, γ_ϵ is an ULQ. We also define a quantizer γ_0 by

$$\gamma_0(y) = \begin{cases} 1, & \text{if } L_1(y) < \ell, \\ 2, & \text{if } L_1(y) > \ell \text{ and } L_2(y) < L_3(y), \\ 3, & \text{if } L_1(y) > \ell \text{ and } L_2(y) > L_3(y). \end{cases}$$

It is clear that under Assumption 5.1, $q(\gamma_\epsilon)$ converges to $q(\gamma_0)$. On the other hand, the quantizer γ_0 cannot be expressed in the form (5.1) unless $a_2 = a_3$, and therefore is not an ULQ.

The preceding example shows that the set of ULQ's is not "closed" in any meaningful sense. This is, of course, just a reflection of the fact that the set of exposed points of a closed convex set is not necessarily closed.

Optimal Quantization Problems

As in Section III and IV, we are interested in characterizing the possible optimal solutions of certain quantization problems. The main result is the following.

Proposition 5.2: Let $J: \overline{Q} \mapsto \Re$ be continuous and convex, and let $\overline{\Gamma}_U$ be the set of all ULQ's. Then,

$$\sup_{\gamma\in\overline{\Gamma}_U} J(q(\gamma)) = \max_{\gamma\in\overline{\Gamma}} J(q(\gamma)). \tag{5.4}$$

Proof: The maximum of the convex function J is attained at some extreme point of $\overline{Q}$. By Corollary 5.1, any extreme point of $\overline{Q}$ is the limit of a sequence $\{q(\gamma_n)\}$ with $\gamma_n \in \overline{\Gamma}_U$ for each n. Q.E.D.

Suppose now that f is a continuous convex function satisfying (4.1), and that $D_f(\cdot,\cdot)$ is the corresponding Ali–Silvey distance measure [cf. (4.2)–(4.3)]. Let us consider the problem of finding a quantizer γ that maximizes

$$F(q(\gamma)) = \sum_{i,j} w_{ij} D_f(q_i(\gamma), q_j(\gamma)), \tag{5.5}$$

where each w_{ij} is a positive weight. (Such quantization problems are studied, for example, in [10].

We are not able to assert any general properties of optimal solutions for the problem of maximizing the performance measure (5.5). Let us now make the additional assumption that $xf(y/x)$ is a convex function of (x, y) (as in the two examples of Section IV). Then, it is easily seen [cf. (4.3)] that F is a convex function on the set $\overline{Q}$. Thus, Proposition 5.2 implies that we can get arbitrarily close to an optimal quantizer while restricting to the class of ULQ's.

VI. Decentralized Neyman–Pearson Detection

In this section, we apply the results of Section III to characterize the optimal solutions of a decentralized Neyman-Pearson detection problem.

The problem formulation is as follows. There are two hypotheses H_1, H_2, and N sensors $S_1, \cdots, S_N$. Each sensor S_i receives an observation Y_i which is a random variable taking values in a set $\mathcal{Y}_i$. We assume that the joint probability distribution of $(Y_1, \cdots, Y_N)$, conditioned on each hypothesis, is known.

For $i = 1, \cdots, N$, let $\overline{\Gamma}_i$ be the set of all randomized quantizers of Y_i, defined as in Section II. Each sensor S_i, upon observing the value of the random variable Y_i, applies a quantizer $\gamma_i \in \overline{\Gamma}_i$, and sends a message $U_i = \gamma_i(Y_i) \in \{1, \cdots, D\}$ to a fusion center. Then, the fusion center makes a decision $U_0 = \gamma_0(U_1, \cdots, U_N) \in \{H_1, H_2\}$, where $\gamma_0 : \{1, \cdots, D\}^N \mapsto \{H_1, H_2\}$ is a deterministic function.

Let $\alpha \in (0, 1)$ be a given scalar. We consider the problem of choosing the quantizers $\gamma_1, \cdots, \gamma_N$ and the function γ_0, so as to maximize the "probability of detection" P_D by the fusion center, subject to the "probability of false alarm" P_F being bounded by α. Formally,

$$\text{maximize } \Pr(\gamma_0(\gamma_1(Y_1), \cdots, \gamma_N(Y_N)) = H_2 | H_2), \tag{6.1}$$

$$\text{subject to } \Pr(\gamma_0(\gamma_1(Y_1), \cdots, \gamma_N(Y_N)) = H_2 | H_1) \le \alpha. \tag{6.2}$$

Our main result is the following.

Proposition 6.1: Suppose that the random variables $Y_1, \cdots, Y_N$ are conditionally independent given either hypothesis. Then, there exists an optimal solution of the problem (6.1)–(6.2) such that each of the quantizers $\gamma_1, \cdots, \gamma_N$ is a monotone LRQ.

Proof: For $i = 1, \cdots, N$, $j = 0, 1$, and $\gamma_i \in \overline{\Gamma}_i$, let $q^i(\gamma_i | H_j) \in \Re^D$ be the vector of components $\Pr(\gamma_i(Y_i) = d | H_j)$, $d = 1, \cdots, D$. Let $\overline{Q}_i = \{(q^i(\gamma_i|H_1), q^i(\gamma_i|H_2)) | \gamma_i \in \overline{\Gamma}_i\}$. (These are essentially the same definitions as in Section II.)

Using the conditional independence assumption, the problem (6.1)–(6.2) is equivalent to

maximize

$$\sum_{\{u\in\{1,\cdots,D\}^N | \gamma_0(u) = H_2\}} \prod_{i=1}^N \Pr(\gamma_i(Y_i) = u_i | H_2), \tag{6.3}$$

subject to

$$\sum_{\{u\in\{1,\cdots,D\}^N | \gamma_0(u) = H_2\}} \prod_{i=1}^N \Pr(\gamma_i(Y_i) = u_i | H_1) \le \alpha. \tag{6.4}$$

For any fixed γ_0, this is the same as a constrained optimization problem defined over the set $\prod_{i=1}^N \overline{Q}_i$. The latter is a Cartesian product of compact sets (Proposition 2.1), and is therefore compact. The cost function (6.3) as well as the left-hand side of the constraining equation (6.4) are continuous. This proves the existence of an optimal solution, for any fixed γ_0. Since there is only a finite number of choices for γ_0, we conclude that the problem (6.1)–(6.2) has an optimal solution.

Let us now consider the problem facing a particular sensor S_i when the quantizers of all other sensors are fixed. Notice that the functions in (6.3)–(6.4) are linear (and therefore convex) functions of $q^i(\gamma_i|H_j)$. In particular, sensor S_i is maximizing a linear function of $q^i(\gamma_i|H_2)$, while $q^i(\gamma_i|H_1)$ is constrained to belong to a closed set. Therefore, Proposition 3.5(a) applies and shows that there exists an LRQ γ_i which is optimal for the problem facing sensor S_i.

It follows easily that there exists an optimal solution in which each γ_i is an LRQ. We can then modify each γ_i so that it becomes a monotone LRQ, without changing the information available to the fusion center (provided that γ_0 is modified accordingly). Q.E.D.

Remarks:

1) A version of Proposition 6.1 has been proved for the Bayesian counterpart of the problem (6.1)–(6.2) in [25], where decentralized detection problems were first introduced, as well as in several subsequent papers. In fact, in the Bayesian case, an elementary proof is possible.

2) The Neyman–Pearson problem considered here has been studied in several papers [21], [22], [11], [20], [26], [4] for the case $D = 2$. Some of these papers associate a Lagrange multiplier with the constraint (6.2), thus converting the problem to one which is essentially equivalent to a Bayesian one. Then, one can use the Bayesian results to assert the optimality of LRQ's. Unfortunately, such a proof is flawed for the following reason. Let $R(\alpha)$ be the optimal value (i.e., the optimal probability of detection) for the Neyman–Pearson problem (6.1)–(6.2). Unlike classical detection problems, the function R is not concave,[2] in general. Due to the lack of convexity, the optimal value in the maximization of P_D subject to $P_F \leq \alpha$ can be different from the optimal value of the maximization of $P_D - \lambda P_F$, no matter how the Lagrange multiplier λ is chosen.

3) A correct proof of Proposition 6.1, based on a direct argument, has been provided in [27] for the case $D = 2$, and in [29], [30] for the case of general D.

4) It is straightforward to generalize the proof of Proposition 6.1 to cover: a) the case of acyclic detection networks, thus providing a Neyman–Pearson counterpart of the results of [9], see [24]; b) the case where the fusion center is also allowed to use randomization.

5) In our formulation, we have allowed randomized quantizers. However, it is implicit in our formulation that the randomizations at different sensors are statistically independent. [Equations (6.3)–(6.4) would be incorrect otherwise.] If one allows the sensors to randomize cooperatively (e.g., a single coin is tossed, all sensors are informed on the outcome), the problem is "convexified" and bears a much closer relation to a Bayesian decentralized detection problem; see [24] for more details on this point.

[2] An example can be found in [20]. An explanation can be provided by observing that the left-hand side of the constraint (6.4) is not convex when viewed as a function of all the variables involved.

APPENDIX

In this appendix, we prove that the sets Q and $\overline{Q}$ are compact, by suitably extending the proof provided in [15] for the case $D = 2$.

The sets Q and $\overline{Q}$ are clearly bounded, so we only need to show that they are closed. Furthermore, since $\overline{Q}$ is the convex hull of Q, it suffices to show that Q is closed.

Let $\mathcal{P} = (\mathcal{P}_1 + \cdots + \mathcal{P}_M/M)$. Let G be the set of all measurable functions from $\mathcal{Y}$ into $\{0,1\}$. Let G^D be the Cartesian product of D copies of G. Let

$$F = \left\{ (f_1, \cdots, f_D) \in G^D | \mathcal{P}\left(\sum_{d=1}^{D} f_d(Y) = 1 \right) = 1 \right\}.$$

For any $\gamma \in \Gamma$ and $d \in \{1, \cdots, D\}$, we can let f_d be the indicator function of the set $\gamma^{-1}(d)$; that is, $f_d(y) = 1$ if and only if $\gamma(y) = d$, and $f_d(y) = 0$ otherwise. Clearly, then, $(f_1, \cdots, f_d) \in F$ and

$$q_d(\gamma|H_i) = \Pr(\gamma(Y) = d|H_i) = \int f_d(y)\, d\mathcal{P}_i(y). \quad \text{(A.1)}$$

Conversely, for any $f = (f_1, \cdots, f_d) \in F$, we define a deterministic quantizer $\gamma \in \Gamma$ as follows. If $\sum_{d=1}^D f_d(y) = 1$, then let $\gamma(y)$ be equal to the unique value of d for which $f_d(y) = 1$. If $\sum_{d=1}^D f_d(y) \neq 1$, then let $\gamma(y) = 1$. Since the event $\sum_{d=1}^D f_d(y) \neq 1$ has zero $\mathcal{P}$-measure, it is seen that (A.1) is again valid. Let $h\colon F \mapsto \Re^{MD}$ be the mapping with components

$$h_{i,d}(f) = \int f_d(y)\, d\mathcal{P}_i(y). \quad \text{(A.2)}$$

The correspondence we have established between F and Γ, together with (A.1), imply that $Q = h(F)$.

The proof will be completed by introducing a topology on G under which F is compact and h is continuous. Then, Q becomes the continuous image of a compact set and is therefore compact.

We use $\mathcal{L}_1(\mathcal{Y}; \mathcal{P})$ to denote the set of all measurable functions $f : \mathcal{Y} \mapsto \Re$ such that $\int |f(y)|\, d\mathcal{P}(y) < \infty$. Similarly, $\mathcal{L}_\infty(\mathcal{Y}; \mathcal{P})$ denotes the set of all measurable functions $f : \mathcal{Y} \mapsto \Re$ such that, after discarding a subset of $\mathcal{Y}$ of zero $\mathcal{P}$-measure, f is bounded. We view G as a subset of $\mathcal{L}_\infty(\mathcal{Y}; \mathcal{P})$. Recalling that $\mathcal{L}_\infty(\mathcal{Y}; \mathcal{P})$ is the dual of $\mathcal{L}_1(\mathcal{Y}; \mathcal{P})$, we consider the weak* topology on $\mathcal{L}_\infty(\mathcal{Y}; \mathcal{P})$, defined as the weakest topology under which the mapping

$$f \mapsto \int f(y) g(y)\, d\mathcal{P}(y) \quad \text{(A.3)}$$

is continuous for every $g \in \mathcal{L}_1(\mathcal{Y}; \mathcal{P})$. By Alaoglu's theorem [7], the unit ball in $\mathcal{L}_\infty(\mathcal{Y}; \mathcal{P})$ is weak*-compact, and it follows easily that G is also compact. Thus, G^D is also compact under the corresponding product topology. Recalling the definition of F, we see that F can be also defined as the set of all elements $(f_1, \cdots, f_D) \in G^D$ such that

$$\int_A \sum_{d=1}^{D} f_d(y)\, d\mathcal{P}(y) = \mathcal{P}(A),$$

for every measurable subset A of $\mathcal{Y}$. Equivalently, for every measurable set A,

$$\int \sum_{d=1}^{D} f_d(y) \mathcal{X}_A(y)\, d\mathcal{P}(y) = \mathcal{P}(A),$$

where $\mathcal{X}_A$ is the indicator function of A. Using the continuity of the mappings of the form (A.3), and since $\mathcal{X}_A \in \mathcal{L}_1(\mathcal{Y}; \mathcal{P})$, it follows that F is the subset of G^D on which certain continuous equality constraints are satisfied. Since G^D is compact, it follows that F is also compact.

For each i, let g_i be the Radon–Nikodym derivative of $\mathcal{P}_i$ with respect to $\mathcal{P}$. Then, $g_i \in \mathcal{L}_1(\mathcal{Y}; \mathcal{P})$ [7]. Furthermore,

$$\int f_d(y)\, d\mathcal{P}_i(y) = \int f_d(y) g_i(y)\, d\mathcal{P}(y), \qquad \forall\, i, d. \quad \text{(A.4)}$$

By the definition of the weak* topology [cf. (A.3)] and (A.4), the mapping $f \mapsto \int f_d(y)\, d\mathcal{P}_i(y)$ is continuous. Thus, the mapping h, whose components are given by (A.2), is continuous. Since $Q = h(F)$, it follows that Q is compact.

REFERENCES

[1] B. Aazhang and H. V. Poor, "On optimum and nearly optimum data quantization for signal detection," *IEEE Trans. Commun.*, vol. COM-32, pp. 745–751, 1984.

[2] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Stat. Soc.*, Series B, vol. 28, pp. 131–142, 1966.

[3] G. R. Benitz and J. A. Bucklew, "Asymptotically optimal quantizers for detection of i.i.d. data," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 316–325, 1989.

[4] M. Barkat and P. K. Varshney, "Decentralized CFAR signal detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-25, pp. 141–149, 1989.

[5] I. Csiszar, "Information-type measure of difference of probability distributions and indirect observations," *Stud. Scientarium Mathematicerium Hungarica*, vol. 2, pp. 229–318, 1967.

[6] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations," *Ann. Math. Stat.*, vol. 23, pp. 493–507, 1952.

[7] N. Dunford and J. T. Schwartz, *Linear Operators*, vol. 1. New York: Wiley, 1957.

[8] A. Dvoretzky, A. Wald, and J. Wolfowitz, "Relations among certain ranges of vector measures," *Pacific J. Math.*, vol. 1, pp. 59–74, 1951.

[9] L. K. Ekchian and R. R. Tenney, "Detection networks," in *Proc. 21st IEEE Conf. Decision Contr.*, 1982, pp. 686–691.

[10] T. J. Flynn, and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, IT-33, pp. 773–787, 1987.

[11] I. Y. Hoballah and P. K. Varshney, "Neyman–Pearson detection with distributed sensors," in *Proc. 25th IEEE Conf. Decision Contr.*, Athens, Greece, Dec. 1986, pp. 237–241.

[12] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, COM-25, pp. 479–484, 1977.

[13] D. Kazakos, V. Vannicola, and M. C. Wicks, "Optimum quantization for multisensor detection," unpublished manuscript, 1989.

[14] A. Liapounoff, "Sur les fonctions-vecteurs completement additives," *Bull. Acad. Sci. U.S.S.R.*, Ser. Math., 4, pp. 465–478, 1940.

[15] J. Lindenstrauss, "A short proof of Liapounoff's convexity theorem," *J. Math. Mech.*, vol. 15, pp. 971–972, 1966.

[16] H. V. Poor, "Fine quantization in signal detection and estimation," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 960–972, 1988.

[17] B. Picinbono and P. Duvaut, "Optimum quantization for detection," *IEEE Commun.*, vol. COM-36, pp. 1254–1258, 1988.

[18] H. V. Poor and J. B. Thomas, "Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, pp. 893–900, 1977.

[19] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.

[20] A. R. Reibman, "Performance and fault-tolerance of distributed detection networks," Ph.D. dissertation, Dept. Elec. Eng., Duke Univ., Durham, NC, 1987.

[21] R. Srinivasan, "Distributed radar detection theory," *IEE Proc.*, vol. 133, pp. 55–60, 1986.

[22] ——, "A theory of distributed detection," *Signal Processing*, vol. 11, pp. 319–327, 1986.

[23] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr. Signals, Syst.*, vol. 1, pp. 167–182, 1988.

[24] ——, "Decentralized detection," Tech. Rep. LIDS-P-1913, Lab. Inform. Decision Syst., M.I.T., Cambridge, MA., Sept. 1989; to appear in *Advances Stat. Signal Processing, Vol. 2: Signal Detection*, H. V. Poor and J. B. Thomas, Eds.

[25] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, 501–510, 1981.

[26] S. C. A. Thomopoulos, R. Viswanathan, and D. P. Bougoulias, "Optimal decision fusion in multiple sensor systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, pp. 644–653, 1987.

[27] ——, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-25, pp. 761–764, 1989.

[28] H. L. Van Trees, *Detection Estimation and Modulation Theory*, vol. I. New York: Wiley, 1968.

[29] R. Viswanathan, A. Ansari, and S. C. A. Thomopoulos, "Optimal partitioning of observations in distributed detection," Tech. Rep. in TR-SIU-EE-3-87, Dept. Elec. Eng., Southern Illinois Univ., 1987; also in *Abstracts of Papers, Int. Symp. Inform. Theory*, Kobe, Japan, June 1988, p. 197.

[30] D. J. Warren and P. K. Willett, "Optimal decentralized detection for conditionally independent sensors," in *Proc. 1989 Amer. Contr. Conf.*, pp. 1326–1329.

John N. Tsitsiklis was born in Thessaloniki, Greece, in 1958. He received the B.S. degree in mathematics in 1980, and the B.S., M.S., and Ph.D. degrees in electrical engineering, all from the Massachusetts Institute of Technology, Cambridge, in 1980, 1981, and 1984, respectively.

During the academic year 1983–1984 he was an acting Assistant Professor of Electrical Engineering at Stanford University, Stanford, CA. Since 1984 he has been with the Electrical Engineering and Computer Science Department at the Massachusetts Institute of Technology, where he is currently a Professor. His research interests are in the areas of parallel and distributed computation, systems and control theory, and operations research.

Dr. Tsitsiklis is the coauthor, with D. Bertsekas, of *Parallel and Distributed Computation: Numerical Methods* (1989). He has been a recipient of an IBM Faculty Development Award (1983), an NSF Presidential Young Investigator Award (1986), an Outstanding Paper Award by the IEEE Control Systems Society (for a paper coauthored with M. Athans, 1986), and of the Edgerton Faculty Achievement Award by M.I.T. (1989). He is an associate editor of *Applied Mathematics Letters* and *Automatica*.