

SRC TR 86-5

**Robust Minimas Source Coding
With A Fidelity Criterion**

by

Evangelos Geraniotis

ROBUST MINIMAS SOURCE CODING WITH A
FIDELITY CRITERION

by

Evangelos Geraniotis

ROBUST MINIMAX SOURCE CODING WITH A FIDELITY CRITERION

Evangelos Geraniotis

Department of Electrical Engineering
and Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

The problem of minimax robust source coding under a fidelity criterion for sources whose statistics belong to uncertainty classes determined by 2- alternating Choquet capacities is examined. We consider (i) single-letter difference distortion criteria for discrete memoryless sources whose probability distributions belong to capacity classes and (ii) the mean- square error distortion criterion for stationary Gaussian sources whose spectral measures belong to capacity classes. Both block source codes and trellis source codes are considered. It is shown that there exists an ensemble of block source codes and an ensemble of trellis codes such that for all rates larger than a critical rate and all sources in the class the average distortion converges to any prescribed fidelity level exponentially with increasing block length or constraint length, respectively. Besides the rate distortion function, the distortion exponent of the class is also evaluated.

I. INTRODUCTION

For sources whose statistical description (i.e., the probability distribution which governs the source statistics) is known, Shannon [1] in his renowned source coding theorem showed that, if the code rate is larger than a critical rate (termed the rate distortion function), we can by using block source codes represent the source in any given alphabet and asymptotically satisfy a fidelity (distortion) criterion. Furthermore it was shown (e.g. [2], [3]) that the asymptotic convergence to a given distortion level is exponential in the code length. Similar results for trellis source codes were established in [4] and [3].

For sources whose statistics are not perfectly known but the determining quantity (e.g., the probability distribution) belongs to a class the rate distortion function was found in [5] to be $\sup_{s \in S} R_s(D)$ whose S is the class of probability distributions of the sources and $R_s(D)$ is the rate distortion function at a distortion level D for a particular member s of the class.

As far as coding is concerned two approaches have been followed. According to the first approach termed *universal coding* and described in [6] - [10] (we have not attempted to compile a complete listing of all the papers on the subject) the source is represented or approximated as a finite composite of stationary ergodic subsources and a union code is formed from the codes which are optimal for these representative subsources. Then, all sources in the class have asymptotically optimal coded performance. This approach is applicable to a large number of cases (e.g., general alphabets and distortion measures), since it is independent of the source statistics. Two disadvantages of this approach are: (i) a large number of representative subsources may be necessary and (ii) the construction of the representative subsources for a given class can be very complicated.

The second approach termed *minimax robust coding* is based on a worst-case design. The least-favorable source is singled out and we subsequently code for it. Then the average distortion approaches any prescribed fidelity level exponentially with increasing block length for all sources in the class. The disadvantage of this approach is that the asymptotic coded performance is not optimal for all but the least-favorable source in the class. However, this approach is attractive because it requires only one representative source for the class (the least-favorable one) which can be explicitly found in several interesting cases. This approach was considered in [11] for some different classes of sources than those considered in this paper. For stationary Gaussian sources with spectral uncertainty within classes similar to those considered in this paper the rate distortion function over the class was derived in [12] but no minimax source coding theorems were established. Finally, minimax noiseless block source coding was considered in [13], and [14]. Again, no effort to compile a complete listing of all the papers on the subject has been made.

In this paper we apply the minimax coding approach for block and trellis source codes which satisfy a fidelity criterion to discrete-memoryless sources (DMS's) and discrete-time stationary Gaussian sources (SGS's) which belong to classes determined by *2-alternating Choquet capacities* [15]. Our choice of these uncertainty models is justified in two ways. First, important uncertainty models like contaminated mixtures [16], total variation neighborhoods [16], band models [17] - [18], and extended p-point models [18] are capacity classes and have played an important role in hypothesis testing [19] and filtering [20]. Second, the least-favorable sources can be explicitly found for the uncertainty classes described by any of the above models. In this paper we restrict attention to DMS's and discrete-time SGS's (continuous-time SGS's are also discussed), because these are the simplest nontrivial cases of interest with which our techniques can be

illustrated. Our results can be extended to other classes of sources; e.g., homogenous first-order Markov sources. Similar results for the dual minimax robust channel coding problem are described in [21].

This paper is organized as follows. Minimax robust coding under a fidelity criterion is discussed in Section II for difference distortion criteria and discrete-memoryless sources with uncertainty in the probability distribution, and in Section III for the mean-square-error distortion criterion and stationary Gaussian sources with spectral uncertainty. In each of these sections we first present the uncertainty models that we consider and introduce the necessary notation. Next, we formulate the mismatch source coding problem and establish the appropriate coding theorems for both block and trellis source codes. Finally, we derive the coding theorems for minimax robust source coding over the uncertainty class. In particular, we show that there exist an ensemble of block source codes and an ensemble of trellis source codes such that, provided the code rate is larger than a critical rate, the average distortion converges exponentially to any prescribed fidelity level with increasing block length or constraint length, respectively, for all sources in the uncertainty class. Then, in Section IV a brief summary of this paper and some conclusions are presented.

II. ROBUST CODING FOR DISCRETE MEMORYLESS SOURCES

A. Uncertainty Classes Generated by Choquet Capacities

Suppose that U is the source alphabet, V is the representation (or user) alphabet, and \mathbf{F} , \mathbf{G} are the σ -algebras generated by subsets of U and V , respectively. A discrete memoryless source is characterized by the probability measure $Q(A)$ $A \in \mathbf{F}$. We assume that the probability measures Q are only known to lie in a convex class generated by a Choquet 2-alternating capacity [15]

$$\mathbf{Q}_w = \{Q \in \mathbf{Q} \mid Q(A) \leq w(A), A \in \mathbf{F}\} \quad (1)$$

where \mathbf{Q} denotes the class of all probability measures on (U, \mathbf{F}) , and w is a 2-alternating capacity on (U, \mathbf{F}) with $w(U) = 1$.

A *Choquet 2-alternating capacity* [15] on (U, \mathbf{F}) is a finite set function, which is increasing, continuous from below, continuous from above on closed sets, and satisfies $w(\emptyset) = 0$ and $w(A \cup B) + w(A \cap B) \leq w(A) + w(B)$ for all $A, B \in \mathbf{F}$. Notice that any finite measure w is a 2-alternating capacity; in this case the uncertainty class generated by (1) reduces to $\mathbf{Q}_w = \{w\}$. If we further assume that U is compact then all the uncertainty models mentioned in the Introduction are capacity classes. If U is not compact [e.g., $U = (-\infty, \infty)$] only the band model can be defined in terms of a capacity.

Examples of 2-alternating capacity classes are: the band model [17] defined by

$$\mathbf{Q}_{w_1} = \{Q \in \mathbf{Q} \mid Q_0(A) \leq Q(A) \leq Q_1(A), \forall A \in \mathbf{F}\}, \quad (2a)$$

where Q_0 and Q_1 are known measures (not necessarily probability measures) with $Q_0(U) \leq 1 \leq Q_1(U)$; the ϵ -contaminated mixtures model [16] defined by

$$\mathbf{Q}_{w_2} = \{Q \in \mathbf{Q} \mid Q(A) = (1-\epsilon)Q_0(A) + \epsilon P(A), \forall A \in \mathbf{F}, P \in \mathbf{Q}\}, \quad (2b)$$

where Q_0 is a known probability measure and the number ϵ in $(0,1)$ is the degree of uncertainty in the model; and the total variation model [16] defined by

$$Q_w = \{Q \in \mathcal{Q} \mid |Q(A) - Q_0(A)| \leq \epsilon \forall A \in \mathbf{F}\}, \quad (2c)$$

where Q_0 is a known probability measure and the number ϵ in $(0,1)$ is again the degree of uncertainty in the model. Then (2a) - (2c) can be expressed in the form (1) if we set

$$w_1(A) = \min\{1 - Q_0(A^c), Q_1(A)\} \quad (3a)$$

for the band model,

$$w_2(A) = (1 - \epsilon)Q_0(A) + \epsilon \quad (3b)$$

for the ϵ -contaminated model, and

$$w_3(A) = \min\{Q_0(A) + \epsilon, 1\} \quad (3c)$$

for the total variation model. See [18] for a description of the p-point capacity class.

In the sequel we will need the following fundamental result due to Huber and Strassen [19]:

Lemma 1: If w is a 2-alternating capacity on (U, \mathbf{F}) , Q_w is a convex class of probability measures determined by it as in (1), and λ is the Lebesgue measure on U , then there exists a unique Lebesgue measurable function $\pi_w: U \rightarrow [0, \infty]$ with the defining property that for all $x \in [0, \infty]$ and A_x defined by $A_x = \{\pi_x > x\}$

$$x \lambda(A_x) + w(A_x^c) \leq x \lambda(A) + w(A^c), \forall A \in \mathbf{F}. \quad (4)$$

Furthermore there exists a measure \hat{Q} in Q_w such that for all $x \in [0, \infty]$.

$$\hat{Q}(\{\pi_w \leq x\}) = w(\{\pi_w \leq x\}), \quad (5)$$

which means that \hat{Q} makes π_w stochastically smallest over all Q in Q_w , and π_w is a

version of $d\hat{Q}/d\lambda$, the generalized Radon-Nikodym derivative of \hat{Q} with respect to λ ; that is, $d\hat{Q}/d\lambda$ may be infinite on sets of λ measure 0.

The function π_w is termed the *Huber-Strassen derivative* of w with respect to λ (w may not be a measure). The probability measure \hat{Q} singled out by Lemma 1 is termed the *least-favorable measure* of the class \mathcal{Q}_w . Let $\hat{Q} = \hat{Q}' + \hat{Q}''$ be the Lebesgue decomposition of \hat{Q} , where \hat{Q}' is absolutely continuous with respect to λ and \hat{Q}'' is singular with respect to λ (that is, it concentrates all its mass on sets of λ measure 0). Then,

$$\hat{Q}'(A) = \int_A \pi_w d\lambda \quad (6a)$$

and

$$\hat{Q}''(A) = w(A \cap \{\pi_w = \infty\}), \quad (6b)$$

for all $A \in \mathcal{F}$. For the band model the Huber-Strassen derivative $\pi_w = \hat{q}$ is defined as

$$\hat{q}_1(u) = \max\{q_0(u), \min\{c, q_1(u)\}\}, \quad (7a)$$

where $q_j = dQ_j/d\lambda$ is the Radon-Nikodym (R-N) derivative of Q_j of (2a) for $j = 0, 1$ and c is chosen so that $\hat{Q}_1(U) = 1$. For the ϵ -contaminated model the corresponding definition is

$$\hat{q}_2(u) = \max\{(1-\epsilon)q_0(u), c\}, \quad (7b)$$

where q_0 is the R-N derivative of Q_0 of (2b) and c is chosen so that $\hat{Q}_2(U) = 1$. Similarly for the total variation neighborhood model we have

$$\hat{q}_3(u) = \max\{c', \min\{c'', q_0(u)\}\}, \quad (7c)$$

where q_0 is the R-N derivative of the Q_0 of (2c) and c', c'' are chosen so that $\hat{Q}_3(U) = 1$. See [18] for the definition of \hat{q} for the p-point class model.

It should be noted that Huber-Strassen derivatives of *generalized capacities* [a generalized capacity is defined in the same way as a 2-alternating capacity except that it is required to be continuous from above on compact (and not just closed) sets] with respect to σ -finite (and not just finite) measures can be constructed [22, Chapter IV]. One of the implications of this extension is that several of the most useful examples of capacity classes (e.g., ϵ -mixtures, variation neighborhoods) are generalized capacities when U is σ -compact (and not just compact). Then, if U is σ -compact and thus λ is σ -finite, Lemma 1 still holds.

For the proofs of the minimax robust coding theorems in Section II.C below we need to assume that the least-favorable measure \hat{Q} of the capacity class \mathcal{Q}_w is absolutely continuous with respect to the Lebesgue measure λ on U (i.e., $\hat{Q} \ll \lambda$). This assumption is satisfied, provided that $Q_1 \ll \lambda$ for the band class and $Q_0 \ll \lambda$ for the ϵ -contaminated and total variation neighborhood classes.

We would also like to mention at this point that if U is a discrete alphabet then $\hat{q}(u)$ for $u \in U$ becomes a probability mass function (pmf) and all the results involving the capacities described above still hold, provided that we replace the integrals with respect to u with sums and the Radon-Nikodym derivatives with pmf's. This duality becomes possible if we replace the Lebesgue measure on U (in the continuous case) with the measure which assigns equal mass to all the elements of U (in the discrete case) in Lemma 1 and apply the Huber-Strassen theory to this case. See [23] for a more extensive discussion of this duality. Therefore, in the sequel we will be working with continuous amplitude sources, pdf's, and integrals, but the results will still be valid for the corresponding situations with discrete amplitude sources.

B. Mismatch Source Coding Theorems for Block and Trellis Codes

In this paper we consider DMS's with $V = (-\infty, \infty)$ and single-letter difference distortion measures $d(u, v) = d(v - u)$, $u \in U$, $v \in V$ which satisfy the bounded variance condition

$$\int_U d^2(u) dQ(u) \leq d_0 \quad (8)$$

for some finite positive number d_0 and all Q in \mathcal{Q}_w . Let D be the maximum level of average distortion per letter that we can tolerate when we represent letters from the source alphabet U with letters from the user alphabet V . We define the average distortion per letter associated with a difference distortion criterion $d(\cdot)$ which satisfies (8), a conditional probability density function (pdf) $p(v | u)$, and a source with probability measure Q in \mathcal{Q}_w by

$$D(p, Q) = \int_U \int_V p(v | u) d(u, v) \lambda(dv) dQ(u). \quad (9)$$

Furthermore, if we assume that $p(v | u) = p(v - u)$ for $u \in U$, $v \in V$, (9) is equivalent to

$$D(p, Q) = D(p) \cdot Q'(U) \leq D(p) = \int_V d(z) p(z) \lambda(dz). \quad (10)$$

Equation (10) follows from (9) after performing the substitution $z = v - u$ and taking into account the facts that $V = (-\infty, \infty)$ and $Q'(U) \leq Q(U) = 1$. Notice that in (10) the average distortion $D(p, Q)$ is upperbounded by $D(p)$ which does not depend on Q . This fact will be critical for establishing the results of the next Sections. The need for the aforementioned assumptions will become clear in subsection II.C (during the proof of the main results stated in Theorems 3 and 4).

Suppose now that in the presence of uncertainty about Q the user mistakenly assumes that (or attempts to estimate Q and comes with an estimate that) \tilde{Q} is the

$$E_0(\rho, p, \tilde{q}; Q) = -\ln \left\{ \int_U \frac{1}{\tilde{q}(u)} \left[\int_V \tilde{p}(u | v)^{\frac{1}{1+\rho}} \tilde{p}(v) \lambda(dv) \right]^{1+\rho} dQ(u) \right\} \quad (15a)$$

$$= -\ln \left\{ \int_U \left[\int_V p(v | u)^{\frac{1}{1+\rho}} \left[\int_U p(v | u') \tilde{q}(u') \lambda(du') \right] \lambda(dv) \right]^{\frac{\rho}{1+\rho}} dQ(u) \right\} \quad (15b)$$

and d_0 was introduced in (8). For this theorem to be valid it is required that $I(p, \tilde{q}; Q) > 0$ and $E_0(\rho, p, \tilde{q}; Q) < 0$ for all ρ in $[-1, 0]$. However, these inequalities are satisfied for all nontrivial p , \tilde{q} , and Q .

Remark 1. The quantities $I(p, \tilde{q}; Q)$ and $E_0(\rho, p, \tilde{q}; Q)$ represent the mismatch mutual information function and the mismatch distortion exponent, respectively.

Remark 2. We consider Theorem 1 and Theorem 2, which follows, important in two ways: as being fundamental intermediate results necessary for the proof of the main Theorems 3 and 4 below, and as interesting independent results which characterize source coding with a fidelity criterion (for both block and trellis codes) in the case of mismatch (i.e., when the actual probability measure of the source is different than the estimate employed in the encoding procedure).

Proof: Our proof basically follows from a modification of the proof of the source coding theorem for the matched case $\tilde{q} = q$ ($Q = \tilde{Q}$ has a pdf in this case) given in [3, Sections 7.5.1 and Lemma 7.2.2]. Thus we only present here these points of the proof of [3] which were considerably modified. It was shown in [3] that we can express the average distortion D_c achieved using a particular code $C = \{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_M\}$ where $\underline{v}_m \in V^n$, $m = 1, 2, \dots, M$, and $M = \lceil e^{nR} \rceil$ as

$$D_c \leq D(p, Q) + d_0 \left[\int_{U^n} \int_{V^n} p_n(\underline{v} | \underline{u}) B(\underline{u}, \underline{v}; C) \lambda_n(\underline{v}) dQ_n(\underline{u}) \right]^{\frac{1}{2}} \quad (16)$$

In (16) $dQ_n(\underline{u}) = \prod_{i=1}^n dQ(u_i)$, $p_n(\underline{v} | \underline{u}) = \prod_{i=1}^n p(v_i | u_i)$ for $u \in U^n$, $v \in V^n$,

$\lambda_n(d\underline{u}) = \prod_{i=1}^n \lambda(du_i)$, and

$$B(\underline{u}, \underline{v}; C) = \begin{cases} 1; & d_n(\underline{u}, \underline{v}) < \min_{v' \in C} d_n(\underline{u}, v') \\ 0; & d_n(\underline{u}, \underline{v}) \geq \min_{v' \in C} d_n(\underline{u}, v') \end{cases} \quad (17)$$

Next we proceed to bound \bar{D}_c the average of D_c over the ensemble of block source codes described in Theorem 1. The members of this ensemble are assigned the product

distribution $\tilde{P}(C) = \prod_{m=1}^M \tilde{p}_n(\underline{v}_m)$ where $\tilde{p}_n(\underline{v}) = \prod_{i=1}^n \tilde{p}(v_i)$ and

$\tilde{p}(v) = \int_U p(v | u) \tilde{q}(u) \lambda(du)$ for $v \in V$. Then, since

$\tilde{p}_n(\underline{u} | \underline{v}) = p_n(\underline{v} | \underline{u}) \tilde{q}_n(\underline{u}) / \tilde{p}_n(\underline{v})$, we can write E_c the quantity inside the brackets in (16) as

$$E_c = \int_{U^n} \frac{1}{\tilde{q}_n(\underline{u})} \left[\int_{V^n} \tilde{p}_n(\underline{u} | \underline{v}) B(\underline{u}, \underline{v}; C) \tilde{p}_n(\underline{v}) \lambda_n(d\underline{v}) \right] dQ_n(\underline{u}), \quad (18)$$

and apply for ρ in $[-1, 0]$ Holder's inequality:

$$\int f g d\mu \leq \left(\int f^\alpha d\mu \right)^{\frac{1}{\alpha}} \left(\int g^\beta d\mu \right)^{\frac{1}{\beta}}, \quad (19)$$

where $1 < \alpha < \infty$, $1 < \beta < \infty$, and $\alpha^{-1} + \beta^{-1} = 1$, for

$f = \tilde{p}_n(\underline{u} | \underline{v})$, $\alpha = 1/(1 + \rho)$, $g = B(\underline{u}, \underline{v}; C)$, $\beta = 1/(-\rho)$, and $d\mu = \tilde{p}_n(\underline{v}) \lambda_n(d\underline{v})$

to obtain

$$E_c \leq \int_{U^n} \frac{1}{\tilde{q}_n(\underline{u})} \left[\int_{V^n} \tilde{p}_n(\underline{u} | \underline{v})^{\frac{1}{1+\rho}} \tilde{p}_n(\underline{v}) \lambda_n(d\underline{v}) \right]^{1+\rho}$$

$$\left[\int_{V_n} B(\underline{u}, \underline{v}; C) \bar{p}_n(\underline{v}) \lambda(d\underline{v}) \right]^{-\rho} dQ_n(\underline{u}) \quad (20)$$

Averaging E_c over the code ensemble and applying Jensen's inequality yields

$$\bar{E}_c \leq \exp \left[-nE_0(\rho, p, \tilde{q}; Q) \right] \left[\int_{V_n} B(\underline{u}, \underline{v}; C) \bar{p}_n(\underline{v}) \lambda_n(d\underline{v}) \right]^{-\rho} \quad (21)$$

where \bar{X} denotes averaging X over the code ensemble. In [3, p.393] it was shown that

$$\int_{V_n} B(\underline{u}, \underline{v}; C) \bar{p}_n(\underline{v}) \lambda_n(d\underline{v}) \leq M^{-1} \leq e^{-nR}. \quad (22)$$

Finally, by using inequalities (10) and (11), averaging (16) over the ensemble of block source codes, and substituting from (21) and (22) we obtain (14). Then, condition (12), definition (13), and the aforementioned positivity and negativity requirements follow from the requirement that the exponent $E_0(\rho, p, \tilde{q}; Q) - \rho R$ be strictly positive for ρ in $[-1, 0]$ and the fact that

$$I(p, \tilde{q}; Q) = \left[\partial E_0(\rho, p, \tilde{q}; Q) / \partial \rho \right]_{\rho=0} \quad (23)$$

in the same way as for the usual source coding theorem (e.g. [3, Lemmas 7.2.2 and 7.2.3])

We would like at this point to emphasize the necessity of the assumption that Q is absolutely continuous with respect to λ (i.e., $\tilde{Q} \ll \lambda$). This was essential in deriving equations (18), (20), (21) and (22).

We now show that the positivity and negativity requirements on $I(p, \tilde{q}; Q)$ and $E_0(\rho, p, \tilde{q}; Q)$ for all ρ in $[-1, 0]$, respectively, are satisfied for all pairs of (\tilde{q}, Q) . To show the positivity of $I(p, \tilde{q}; Q)$ we only need to apply the inequality $\ln x \geq 1 - x^{-1}$ ($x > 0$) and use the fact that $p(v | u)$ and $\bar{p}(v)$ are pdf's. To show the negativity of $E_0(\rho, p, \tilde{q}; Q)$ given by (15a) for all ρ in $[-1, 0]$ we first apply Holder's inequality [see (19)] for $f = \bar{p}(u | v)$, $\alpha = 1/(1+\rho)$, $g = 1$, $\beta = 1/(-\rho)$, and

$d\mu = \tilde{p}(v)\lambda(dv)$, then we multiply both members of the resulting inequality by $1/\tilde{q}(u)$, we integrate over U with respect to $dQ(u)$, and finally take the negative logarithm.

Similarly, for *trellis* source codes used as described in [3, Section 7.4]) with the necessary modifications for the case of mismatch, the following result holds:

Theorem 2: Under the assumptions of Theorem 1, consider the ensemble of trellis codes of constraint length K and rate $R = \frac{1}{n} \ln M$ nats per source symbol satisfying (12) which is generated by assigning N letters from the alphabet V independently and according to $\tilde{p}(v) = \int_U p(v|u)\tilde{q}(u)\lambda(du)$, $v \in V$, to the branches of the trellis. Then, the average distortion \bar{D}_c over the ensemble of trellis source codes is upper-bounded by $D_K(\rho, p, \tilde{q}; Q)$ given by

$$D_K(\rho, p, \tilde{q}; Q) = D + \frac{d_0 M^{\frac{1}{2}(K-1)\rho}}{1 - M^{-\left[E_0(\rho, p, \tilde{q}; Q) / R - \rho \right]}} \quad (24)$$

where $-1 \leq \rho \leq E_0(\rho, p, \tilde{q}; Q) / R$.

Proof: It is a modification of the proof of the source coding theorem for trellis codes (see [3, Section 7.5.2]) for the matched case ($\tilde{q} = q$) which takes into account the mismatch arguments established during the proof of Theorem 1; therefore we do not repeat it here.

C. Minimax Robust Source Coding Theorems for Block and Trellis Codes

In this section we assume that the probability measure which governs the statistics of the source is only known to lie in a class of the form (1) described in Section II.A. The source encoder employs a measure \tilde{Q} in a way described in Theorems 1 and 2. The

goal is to choose \tilde{Q} so that for all code rates larger than a critical rate the probability of erroneous representation approaches zero with increasing blocklength for all sources in the class.

To prove the main results of this section we need to assume that the following condition is satisfied:

Condition C_0 The conditional pdf p which, for a given probability measure \tilde{Q} and the corresponding density \tilde{q} , minimizes $I(p, \tilde{q}; \tilde{Q})$ under the constraint (11) is of the form $p(v | u) = p(v - u)$ for all $u \in U$ and $v \in V$.

What we really require with condition C_0 is that

$$\arg \min_{p \in \mathbf{P}_D} I(p, \tilde{q}; \tilde{Q}) = \arg \min_{p \in \mathbf{P}'_D} I(p, \tilde{q}; \tilde{Q}) \quad (25)$$

Where

$$\mathbf{P}_D = \{p \in \mathbf{P} \mid D(p) \leq D\},$$

$$\mathbf{P}'_D = \{p \in \mathbf{P} \mid p(v | u) = p(v - u), \forall (u, v) \in U \times V, D(p) \leq D\},$$

and \mathbf{P} is the class of all conditional pdf's. Therefore, if condition C_0 is satisfied, we only need to consider the smaller class \mathbf{P}'_D for which the average distortion $D(p, Q)$ is upper-bounded by $D(p)$ [see (10)]. To restrict attention to minimizations over the smaller class \mathbf{P}'_D turns out to be necessary for Theorems 3 and 4 below to be valid. Condition C_0 is not so restrictive when we consider difference distortion criteria and continuous amplitude sources with $V = (-\infty, \infty)$. For example, *discrete memoryless Gaussian* sources with uncertainty in their probability distribution described by capacity classes satisfy the above condition and so do the *stationary Gaussian sources* with spectral uncertainty considered in Section III.

We now state and prove the main results of this section:

Theorem 3: Suppose the probability distribution Q belongs to a class of the form (1) and \hat{Q} (let $\hat{Q} \ll \lambda$ and $\hat{q} = d\hat{Q}/d\lambda$) is the element of the class singled out by Lemma 1. Then the following inequalities are true for all conditional pdf's p , all Q in \mathcal{Q}_w , and all ρ in $[-1,0]$:

$$I(p, \hat{q}; Q) \leq I(p, \hat{q}; \hat{Q}) \leq I(p, q; \hat{Q}) \quad (26)$$

and

$$E_0(\rho, p, \hat{q}; Q) \geq E_0(\rho, p, \hat{q}; \hat{Q}). \quad (27)$$

Suppose further that condition C_0 is satisfied for any fidelity level D . We consider pdf's p in the set P_D' described above. Then the operating point (\hat{p}, \hat{q}) where $\hat{p} = \arg \min_p I(p, \hat{q}; \hat{Q})$ for p satisfying $D(p) \leq D$ and the source determined by \hat{Q} form a saddle point for $\min_{(p, \tilde{q})} \max_Q I(p, \tilde{q}; Q)$ under the fidelity constraint $D(p, Q) \leq D(p) \leq D$; i.e.,

$$I(\hat{p}, \hat{q}; Q) \leq I(\hat{p}, \hat{q}; \hat{Q}) \leq I(\hat{p}, q; \hat{Q}). \quad (28)$$

The triple $(\hat{p}, \hat{q}; \hat{Q})$ also represents a least-favorable operating point for the distortion exponent as eq. (27) applied for $p = \hat{p}$ indicates. Finally, the condition

$$R > I(\hat{p}, \hat{q}; \hat{Q}) \quad (29)$$

is sufficient and necessary to guarantee that for the ensemble of block source codes of length n and rate R determined by \hat{q} (as described in Theorem 1, just set $\tilde{q} = \hat{q}$) the average distortion converges to the fidelity level D exponentially with increasing n for all sources in the class.

Remark 3. The quantity $I(\hat{p}, \hat{q}, \hat{Q})$ represents the rate distortion function of the class defined by (1).

Remark 4. Besides the inequality in (28) the following inequality also holds

$$I(\hat{p}, q; Q) \leq I(\hat{p}, \hat{q}; \hat{Q}) \leq I(p, \hat{q}; \hat{Q}) \quad (30)$$

under the fidelity constraint $D(p, Q) \leq D(p) \leq D$, which implies that $I(\hat{p}, \hat{q}; \hat{Q}) = \max_Q \min_p I(p, q; Q)$ for $Q \ll \lambda$, that is, the rate distortion function for the source determined by \hat{Q} is the worst-case rate distortion function over all sources in \mathbf{Q}_w [in the usual notation: $R_{\hat{Q}}(D) = \sup_{Q \in \mathbf{Q}_w} R_Q(D)$].

Proof: We first prove the inequalities (26) and (27). To prove the right-hand side inequality in (26) we use Jensen's inequality to show that $I(p, \hat{q}; \hat{Q}) - I(p, q; \hat{Q}) \leq 0$. For the left-hand side inequality in (26) we notice that it is equivalent to:

$$\int_U G(\hat{q}) dQ \leq \int_U G(\hat{q}) d\hat{Q}, \quad (31)$$

where $G(\hat{q}) = \int_V p(v | u) \ln \frac{p(v | u)}{\int_U \hat{q}(u') p(v | u') \lambda(du')}$. Since $G(\hat{q})$ is a

decreasing function of $\hat{q} = \pi_w$, and according to Lemma 1 \hat{Q} makes π_w stochastically smallest over all Q in \mathbf{Q}_w , the inequality in (31) is satisfied. Similarly, to prove the inequality in (27) notice that it is equivalent [via (15b)] to the inequality

$$\int_U H(\hat{q}) dQ \leq \int_U H(\hat{q}) d\hat{Q} \quad (32)$$

where $H(\hat{q}) = \left\{ \int_V p(v | u)^{\frac{1}{1+\rho}} \left[\int_U \hat{q}(u') p(v | u') \lambda(du') \right]^{\frac{\rho}{1+\rho}} \lambda(dv) \right\}^{1+\rho}$. How-

ever, since $G(\hat{q})$ is a decreasing function of $\hat{q} = \pi_w$, for ρ in $[-1, 0]$ and according to Lemma 1 \hat{Q} makes π_w , stochastically smallest over all Q in \mathbf{Q}_w , (32) is satisfied and so

is (27).

Next we prove inequality (28). The left-hand side inequality in (28) follows from the left-hand side inequality in (26) for $p = \hat{p}$. Then, the right-hand side inequality in (28) follows from the fact that, because of the definition of \hat{p} , $I(\hat{p}, \hat{q}, \hat{Q}) \leq I(p, \hat{q}, \hat{Q})$ and from the right-hand side inequality in (26). The constraint $D(\hat{p}) \leq D$ is satisfied since the minimization of $I(p, \hat{q}; \hat{Q})$ was performed under the constraint $D(p) \leq D$.

Condition C_0 is critical because it enables us to consider pdf's of the form $p(v | u) = p(v - u)$ and thus use eq. (10) which guarantees that $D(p, Q)$ is upper-bounded by $D(p)$ which is independent of Q .

We now proceed to the final stage of the proof of Theorem 3. First, because of (28) condition (29) implies that $R > I(\hat{p}, \hat{q}; Q)$ for all Q in \mathcal{Q}_w . Furthermore as discussed above $D(\hat{p}) \leq D$ is satisfied independent of Q . Thus Theorem 1 applied for $\bar{p} = \hat{p}$ implies that for the ensemble of block source codes of rate R and length n (for which the n letters of each codeword are chosen from the user alphabet V independently and according to \hat{p} [where $\hat{p}(v) = \int_U \hat{q}(u) \hat{p}(v | u) \lambda(du)$, $v \in V$], while the $[e^{nR}]$ codewords are chosen independently and with equal probability) the average distortion converges to the fidelity level D exponentially with increasing n . Since this is true for all q in the class under consideration the sufficiency of condition (29) is established. To prove its necessity, notice that, according to the usual converse source coding theorem under a fidelity criterion for the matched case, $R < I(\hat{p}, \hat{q}; \hat{Q})$ implies that fidelity level D can not be reached for the source determined by \hat{Q} , which is a member of the aforementioned class. This completes the proof of Theorem 3.

The proof of eq. (30) stated in Remark 4 is a result of the inequalities:

$$I(\hat{p}, q; Q) \leq I(\hat{p}, \hat{q}; Q) \leq I(\hat{p}, \hat{q}; \hat{Q}) \leq I(p, \hat{q}; \hat{Q})$$

where $D(\hat{p}) \leq D$ and $D(p) \leq D$. The first inequality follows from an application of Jensen's inequality, the second inequality was proved as part of eq. (28), and the third inequality follows from the definition of \hat{p} as the minimizing argument for $I(p, \hat{q}; \hat{Q})$.

At this point we discuss the choice of the *operating point* that is of a triplet of the form (ρ, p, \tilde{q}) , where ρ is the parameter in $[-1, 0]$ involved in the distortion exponent $E_o(\rho, p, \tilde{q}; Q) - \rho R$, p is a conditional pdf in $\mathbf{P}'_{\mathbf{D}}$, and \tilde{q} characterizes the ensemble of block source codes. Thus, if our main objective is to operate at the minimum required rate [see (29)] then the operating point should be $(\hat{\rho}, \hat{p}, \hat{q})$ where $\hat{\rho} = \arg \max_{\rho} [E_o(\rho, \hat{p}, \hat{q}; \hat{Q}) - \rho R]$. However, if our main objective is to minimize the average distortion, then $(\tilde{\rho}, \tilde{p}, \hat{q})$ where $(\tilde{\rho}, \tilde{p}) = \arg \max_{(\rho, p)} [E_o(\rho, p, \hat{q}; \hat{Q}) - \rho R]$ for ρ in $[-1, 0]$ and p satisfying $D(p) \leq D$ should be the operating point and the rate R should satisfy $R > I(\tilde{p}, \hat{q}; \hat{q})$ instead of (29).

Notice that in contrast to the corresponding result for the dual robust channel coding problem (see [21]) the operating point $(\tilde{\rho}, \tilde{p}, \hat{q})$ and the source determined by \hat{Q} do not form a saddle-point for $\max_{(\rho, p, \tilde{q})} \min [E_o(\rho, p, \tilde{q}; Q) - \rho R]$. This due to the fact that in general it is not true that $E_o(\rho, p, q; Q) \geq E_o(\rho, p, \tilde{q}; Q)$ for $\tilde{q} \neq q$.

As a final comment for Theorem 3, notice that the minimum and maximum involved [minimizing argument \hat{p} , maximizing argument $(\tilde{\rho}, \tilde{p})$] exist, since the functions $I(p, q; Q)$ and $[E_o(\rho, p, q; Q) - \rho R]$, for $Q \ll \lambda$, are convex in ρ and concave in (ρ, p) , respectively, $\rho \in [-1, 0]$ and p belongs to a convex class.

For trellis source codes a similar result holds:

Theorem 4: Under the assumptions of Theorem 3 condition (29) guarantees that for the

ensemble of trellis source codes of constraint length K and rate $R = \frac{1}{2} \ln M$ the average distortion converges to the fidelity level D exponentially with increasing K for all sources in the class. Furthermore, if we define $(\rho', p') = \arg \min_{(\rho, p)} D_K(\rho, p, \hat{q}; \hat{Q})$, where $-1 \leq \rho \leq E_0(\rho, p, \hat{q}; \hat{Q})/R$ and $D(p) \leq D$, then the following inequalities hold for all Q in \mathcal{Q}_w :

$$D_K(\rho', p', \hat{q}; Q) \leq D_K(\rho', p', \hat{q}; \hat{Q}) \leq D_K(\rho, p, \hat{q}; \hat{Q}). \quad (33)$$

Proof: We first prove the inequalities in (33). The left-hand inequality in (33) follows from the inequality in (27) applied for $\rho = \rho'$ and $p = p'$, and the fact that $D_K(\rho, p, \hat{q}; Q)$ is a decreasing function of $E_0(\rho, p, \hat{q}; Q)$. The right-hand inequality follows from the definition of (ρ', p') .

To complete the proof of Theorem 4 notice that (29) together with the left-hand inequality in (28) implies that $R > I(\hat{p}, \hat{q}; Q)$ for all Q in the capacity class. Furthermore, any ρ which satisfies $-1 \leq \rho \leq E_0(\rho, \hat{p}, \hat{q}; \hat{Q})/R$ also satisfies [because of (27)] $-1 \leq \rho \leq E_0(\rho, \hat{p}, \hat{q}; Q)/R$. Consequently, Theorem 2 applied for $\tilde{p} = \hat{p}$ guarantees that for the ensemble of trellis source codes (for which the N symbols from the alphabet V are assigned independently and according to $\hat{p}(v)$ to the branches of the trellis) the average distortion converges to the fidelity level D exponentially with increasing K . Since this is true for all Q in the uncertainty class, the proof is completed.

As discussed at the end of the proof of Theorem 3 the choice of the operating point depends on our objective. For trellis source codes, if our main objective is to minimize the required rate, then the operating point should be $(\hat{\rho}, \hat{p}, \hat{q})$ where $\hat{\rho} = \arg \min_{\rho} [E_0(\rho, \hat{p}, \hat{q}; \hat{Q}) - \rho R]$; otherwise the operating point should be (ρ', p', \hat{q}) where (ρ', p') is defined as in Theorem 4 and the rate R should satisfy $R > I(p', \hat{q}; \hat{q})$

instead of (29).

Finally notice that all the minima involved in Theorem 4 [the minimizing arguments are (ρ', p') and $\hat{\rho}$] exist, since the functions $D_K(\rho, p, \hat{q}; \hat{Q})$ and $D_K(\rho, \hat{p}, \hat{q}; \hat{Q})$ are convex in (ρ, p) and ρ , respectively, $\rho \in [-1, 0]$ and p belongs to a convex class.

III. ROBUST CODING FOR STATIONARY GAUSSIAN SOURCES

In this section we describe the problem formulation and the results for robust minimax coding under a fidelity criterion for stationary Gaussian sources with spectral uncertainty. We do not include the proofs of the results, since they basically follow the same steps as the proofs of the corresponding results of Section II and involve techniques similar to those used there. We start with the description of spectral uncertainty classes generated by Choquet capacities.

Suppose that $U = V = (-\infty, \infty)$ for the source and user alphabets and the discrete-time stationary Gaussian source (SGS) is characterized by the probability density function $q_n(\underline{u})$ for $\underline{u} \in U^n$

$$q_n(\underline{u}) = (2\pi)^{-\frac{n}{2}} |\underline{\Phi}_n|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \underline{u}^T [\underline{\Phi}_n]^{-1} \underline{u}\right\}. \quad (34)$$

In (34) $|\cdot|$ denotes the determinant of a matrix and $\underline{\Phi}_n$ is a correlation matrix of order n , which because of the stationarity is a symmetric Toeplitz matrix, associated to the spectral density $\phi(\omega)$, $\omega \in [-\pi, \pi]$.

Suppose that the spectral density ϕ is the Radon-Nikodym derivative of a spectral measure Φ defined on sets $A \in \mathbf{B}$ where \mathbf{B} is the σ -algebra generated by subsets of $\Omega = [-\pi, \pi]$. The spectral measure Φ is only known to lie in the convex class Φ_w defined by

$$\Phi_w = \{\Phi \in \Phi \mid \Phi(A) \leq w(A), \forall A \in \mathbf{B}; \Phi(\Omega) = w(\Omega)\}. \quad (35)$$

In (35) Φ is the class of all spectral measures on (Ω, \mathbf{B}) and w is a 2-alternating capacity on (Ω, \mathbf{B}) . We impose on the spectral measures Φ the additional constraint $\Phi([- \pi, \pi]) = w([- \pi, \pi]) = 2\pi\sigma^2$, which is a *fixed source variance* constraint and transforms the normalized spectral measures $\Phi(A)/(2\pi\sigma^2)$ into probability measures; this

is necessary for the validity of the Huber-Strassen theory of least-favorability. Another implication of the fixed variance constraint $\Phi([-π,π]) = 2πσ^2$ is that $\phi(\omega)$ has a finite supremum for $\omega \in [-π,π]$. Let Δ denote the global supremum over all ϕ whose spectral measures Φ belong to Φ_w .

All the results about Choquet capacities and uncertainty classes of probability measures generated by them presented in Section II.A are also valid for the spectral uncertainty classes. Let $\hat{\phi}$ and $\hat{\Phi}$ denote the Huber-Strassen derivative and the least-favorable spectral measure in this case. In analogy to Section III.A we assume that $\hat{\Phi} \ll \lambda$, i.e., that $\hat{\Phi}$ is absolutely continuous with respect to λ , the Lebesgue measure on $\Omega = [-π,π]$. This assumption is satisfied if $\Phi_1 \ll \lambda$ for the band class and $\Phi_0 \ll \lambda$ for the ϵ -contaminated and total variation neighborhood classes.

For a *single-letter mean-square-error distortion measure* $d(u,v) = (v-u)^2$, $(u,v) \in U \times V$ the average distortion constraint takes the form:

$$E \{ \| U - V \|^2 \} \leq nD, \tag{36}$$

where $\| \cdot \|$ is the Euclidean norm of the n-dimensional random vector $\underline{U} - \underline{V}$, the expectation E is with respect to the n-dimensional distribution of $(\underline{U}, \underline{V})$, and D is a fixed fidelity level.

Suppose that in the presence of uncertainty about Φ the user mistakenly assumes that $\tilde{\Phi}$ is the spectral measure governing the statistics of the discrete-time stationary Gaussian source. Let $\tilde{\phi}$ denote the R-N derivative of $\tilde{\Phi}$ and $\tilde{\phi}$ denote the spectral density of $\tilde{\Phi}$, that is, we assume that $\tilde{\Phi} \ll \lambda$, where λ is the Lebesgue measure on Ω . Let Q_n and \tilde{Q}_n ($\tilde{Q}_n \ll \lambda$) be the n-th order probability measures induced by the spectral measures Φ and $\tilde{\Phi}$, respectively.

The above situation is characterized by mismatch as in the case described in Theorem 1. Therefore we can apply Theorem 1 to this special case. For the evaluation of the mismatch mutual information and mismatch distortion exponent functions it is now advantageous to follow the technique [24, Section 4.5.2] and make the problem equivalent to that of n independent zero-mean Gaussian DMS's. This involves a unitary transformation of \underline{u} and \underline{v} associated with $\tilde{\Phi}_n$ which preserves the mutual information relationships and the mean-square-error (MSE) distortion constraint.

Furthermore, because of the Gaussian statistics we restrict attention to auxiliary conditional pdf's $p_n(\underline{v} | \underline{u})$ of the form

$$p_n(\underline{v} | \underline{u}) = (2\pi)^{-n/2} |\underline{A}_n \underline{R}_n|^{-\frac{1}{2}} \left\{ \exp \left\{ -\frac{1}{2} (\underline{v} - \underline{A}_n \underline{u})^T [\underline{A}_n \underline{R}_n]^{-1} (\underline{v} - \underline{A}_n \underline{u}) \right\} \right\} \quad (37)$$

where $\underline{A}_n = \text{diag}(a_1, a_2, \dots, a_n)$ ($a_i > 0$ for $i = 1, 2, \dots, n$) is associated with a spectral density $a(\omega)$, $\omega \in [-\pi, \pi]$, and the n -th order Toeplitz matrix \underline{R}_n is associated with spectral density $r(\omega)$, $\omega \in [-\pi, \pi]$. We considered the matrix \underline{A}_n instead of the identity matrix \underline{I}_n (i.e., $\underline{v} - \underline{A}_n \underline{u}$ instead of $\underline{v} - \underline{u}$) because $p_n(\underline{v} | \underline{u})$, defined by (37) for $\underline{A}_n = \underline{I}_n$ and satisfying (36), is too restrictive to allow for the minimization of the mutual information function.

Once the SGS has been decomposed to n independent Gaussian DMS's we can apply the theory of [24, Section 4.5.2], (34), (37), the definitions (13) and (15a) - (15b) of Theorem 1 and the discrete-time version of the Toeplitz Distribution theorem [25] to to put the asymptotic (in the limit of large n) mismatch mutual information function and mismatch distortion exponent in the form:

$$I(a, r, \tilde{\phi}; \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \ln \left[1 + \frac{a(\omega) \tilde{\phi}(\omega)}{r(\omega)} \right] + \frac{a(\omega) [\phi(\omega) - \tilde{\phi}(\omega)]}{a(\omega) \tilde{\phi}(\omega) + r(\omega)} \right\} \lambda(d\omega), \quad (38)$$

and

$$E_0(\rho, a, r, \tilde{\phi}; \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \rho \ln \left[1 + \frac{a(\omega) \tilde{\phi}(\omega)}{(1+\rho)r(\omega)} \right] + \ln \left[1 + \frac{\rho a(\omega) [\phi(\omega) - \tilde{\phi}(\omega)]}{(1+\rho)[a(\omega)\phi(\omega) + r(\omega)]} \right] \right\} \lambda(d\omega). \quad (39)$$

and the average distortion constraint in the form:

$$D(a, r, \tilde{\phi}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ [a(\omega) - 1]^2 \tilde{\phi}(\omega) + a(\omega)r(\omega) \} \lambda(d\omega) \leq D, \quad (40)$$

Notice that as explained in [24, Section 4.5.2] for the matched case the quantities above depend on ϕ , the R-N derivative of Φ , and not on Φ itself; they only depend on $\tilde{\phi}$ because we assumed that $\tilde{\Phi} \ll \lambda$. For the quantities above we have that $I(a, r, \tilde{\phi}; \phi) \geq 0$ and $E_0(\rho, a, r, \tilde{\phi}; \phi) \leq 0$ for $0 \leq \rho \leq 1$ and all pairs (a,r). These inequalities can be proved in the same way as the corresponding inequalities $I(p, \tilde{q}; Q) \geq 0$ and $E_0(\rho, p, \tilde{q}; Q) \leq 0$ in Section II.B.

Next we consider the pair (\tilde{a}, \tilde{r}) which minimizes $I(a, r, \tilde{\phi}; \phi)$, the asymptotic mutual information function, for the matched case ($\tilde{\phi} = \phi$). This pair has been shown in [24, Section 4.5.2] to be defined in terms of a parameter $\tilde{\theta}$ as:

$$\tilde{a}(\omega) = \begin{cases} 1 - \tilde{\theta}/\phi(\omega); & \text{if } \tilde{\theta} < \tilde{\phi}(\omega) \\ 0; & \text{if } \tilde{\phi}(\omega) \leq \tilde{\theta} \end{cases} \quad (41a)$$

and

$$\tilde{r}(\omega) = \tilde{\theta}; \quad \omega \in [-\pi, \pi]. \quad (41b)$$

The parameter $\tilde{\theta}$ is determined by the condition $D(\tilde{a}, \tilde{r}, \tilde{\phi}) = D$ or equivalently:

$$D(\tilde{\theta}, \tilde{\phi}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\tilde{\theta}, \tilde{\phi}(\omega)\} \lambda(d\omega) = D. \quad (42)$$

The parameter $\tilde{\theta}$ lies in the range $[0, \tilde{\Delta}]$ where $\tilde{\Delta}$ is the essential supremum of $\tilde{\phi}$ (see [24]). Similarly, the condition $D(\tilde{a}, \tilde{r}, \phi) \leq D$ is transformed into

$$D(\tilde{\theta}, \phi) = \frac{1}{2\pi} \int_{\{\phi(\omega) \leq \tilde{\theta}\}} \phi(\omega) \lambda(d\omega) + \int_{\{\tilde{\theta} < \phi(\omega)\}} \left\{ \tilde{\theta} + \left[\frac{\tilde{\theta}}{\phi(\omega)} \right]^2 [\phi(\omega) - \tilde{\phi}(\omega)] \right\} \lambda(d\omega) \leq D. \quad (43)$$

We now state the main corresponding to Theorem 1 result for stationary Gaussian sources:

Theorem 5: Consider a discrete-time stationary Gaussian source with n -th order distribution Q_n induced by the spectral measure $\Phi(\omega)$, $\omega \in [-\pi, \pi]$, and a n -th order conditional pdf $\tilde{p}_n(\underline{v} | \underline{u})$ for $(\underline{u}, \underline{v}) \in U^n \times V^n$ of the form (36) induced by the spectral densities $\tilde{a}(\omega)$ and $\tilde{r}(\omega)$, $\omega \in [-\pi, \pi]$, defined by (41a)-(41b), where the parameter $\tilde{\theta}$ satisfies the average distortion constraint (42). Assume that for a given source sequence $\underline{u} \in U^n$ the source encoder chooses the codeword $\underline{v} \in V^n$ which minimizes $\|\underline{v} - \underline{u}\|$. Consider the ensemble of block source codes of length n and rate R whose codewords are chosen independently with equal probability and the n letters of each codeword are chosen from the user alphabet V according to $\tilde{p}(\underline{v}) = \int_{U^n} \tilde{q}_n(\underline{u}) \tilde{p}_n(\underline{v} | \underline{u}) \lambda_n(d\underline{u})$, $\underline{v} \in V^n$, where the pdf $\tilde{q}_n(\underline{u})$ is induced by the spectral density $\tilde{\phi}$. Then, if

$$D(\tilde{\theta}, \phi) \leq D(\tilde{\theta}, \tilde{\phi}), \quad (44)$$

where the two quantities are defined by (42) - (43), and the rate R satisfies

$$R > I(\tilde{\theta}, \tilde{\phi}; \phi), \quad (45)$$

where,

$$I(\tilde{\theta}, \tilde{\phi}; \phi) = \frac{1}{4\pi} \int_{\{\tilde{\theta} \leq \phi(\omega)\}} \left\{ \ln \frac{\phi(\omega)}{\tilde{\theta}} + \left[1 - \frac{\tilde{\theta}}{\phi(\omega)} \right] \frac{1}{\phi(\omega)} [\phi(\omega) - \tilde{\phi}(\omega)] \right\} \lambda(d\omega), \quad (46)$$

then the average distortion \bar{D}_c over the ensemble of block source codes is for large n

upperbounded by

$$D < D + \sqrt{3\Delta} \exp\left\{-\frac{1}{2}n [E_o(\rho, \tilde{\theta}, \tilde{\phi}; \phi) - \rho R]\right\}, \quad (47)$$

where Δ is the global supremum of ϕ for the class of (35) and for $\rho \in [-1, 0]$:

$$E_o(\rho, \tilde{\theta}, \tilde{\phi}; \phi) = \int_{\{\theta \leq \tilde{\phi}(\omega)\}} \left\{ \rho \ln \left[\frac{\rho \tilde{\theta} + \tilde{\phi}(\omega)}{(1+\rho)\tilde{\theta}} \right] + \ln \left[\frac{\rho[\tilde{\phi}(\omega) - \tilde{\theta}]}{(1+\rho)\tilde{\phi}(\omega)^2} \right] [\phi(\omega) - \tilde{\phi}(\omega)] \right\} \lambda(d\omega). \quad (48)$$

For of this theorem to be valid it is required that $I(\tilde{\theta}, \tilde{\phi}; \phi) > 0$ and $E_o(\rho, \tilde{\theta}, \tilde{\phi}; \phi) < 0$ for all ρ in $[-1, 0]$. These inequalities are satisfied for all nontrivial $\tilde{\theta}$, $\tilde{\phi}$, and ϕ .

Remark 5: The functions $I(\tilde{\theta}, \tilde{\phi}; \phi)$ and $E_o(\rho, \tilde{\theta}, \tilde{\phi}; \phi)$ represent the mismatch rate distortion function and the mismatch distortion exponent, respectively.

The corresponding result for trellis source codes is:

Theorem 6: Under the assumptions of Theorem 5, consider the ensemble of trellis codes of constraint length K and rate $R = \frac{1}{N} \ln M$ satisfying (45) which is generated by assigning N letters from the user alphabet V according to the pdf $\bar{p}_N(\underline{v}) = \int_{U^n} \bar{q}_N(\underline{u}) \bar{p}_N(\underline{v} | \underline{u}) \lambda_N(d\underline{u})$, $v \in V^N$. Then, provided that eq. (44) is satisfied, the average distortion \bar{D}_c over the ensemble of trellis source codes is upperbounded by

$$D_K(\rho, \tilde{\theta}, \tilde{\phi}; \phi) = D + \frac{\sqrt{3\Delta} M^{\frac{1}{2}(K-1)\rho}}{1 - M^{-|E_o(\rho, \tilde{\theta}, \tilde{\phi}; \phi)/R - \rho|}} \quad (49)$$

where $-1 \leq \rho \leq E_o(\rho, \tilde{\theta}, \tilde{\phi}; \phi)/R$ and the parameter $\tilde{\theta}$ is determined by (42).

In the presence of uncertainty about the statistics of the source, in particular in the presence of spectral uncertainty within the classes defined by (35), the goal is to choose $\tilde{\phi}$

involved in Theorems 5 and 6 above so that, if the code rate for the codes of the ensembles considered in these two theorems is larger than a critical rate, then the average distortion converges asymptotically to D for all sources in the class.

The result corresponding to Theorem 3 is:

Theorem 7: Suppose the spectral measure Φ (let $\phi = d\Phi/d\lambda$) belongs to a class of the form (35) and $\tilde{\Phi}$ (with $\tilde{\Phi} \ll \lambda$ and $\tilde{\phi} = d\tilde{\Phi}/d\lambda$) is the element of the class singled out by Lemma 1. Then $(\tilde{\theta}, \tilde{\phi}; \tilde{\phi})$ where $\tilde{\theta}$ satisfies $D(\tilde{\theta}, \tilde{\phi}) = D$ [apply (42) for $\tilde{\phi} = \tilde{\phi}$] is a saddle point for $\min_{(\theta, \phi)} \max_{\phi} I(\theta, \tilde{\phi}; \phi)$ under the fidelity constraint $D(\theta, \phi) \leq D$, i.e.,

$$I(\tilde{\theta}, \tilde{\phi}; \phi) \leq I(\tilde{\theta}, \tilde{\phi}; \tilde{\phi}) \leq I(\theta, \phi; \tilde{\phi}), \quad (50)$$

where $D(\tilde{\theta}, \phi) \leq D$ and $D(\theta, \tilde{\phi}) \leq D$ for all $\phi = d\Phi/d\lambda$ with Φ in Φ_w ; it is also a least-favorable operating point for the distortion exponent, that is,

$$E_0(\rho, \tilde{\theta}, \tilde{\phi}; \phi) \geq E_0(\rho, \tilde{\theta}, \tilde{\phi}; \tilde{\phi}) \quad (51)$$

for all ρ in $[-1, 0]$. Finally the condition

$$R > I(\tilde{\theta}, \tilde{\phi}; \tilde{\phi}) \quad (52)$$

is sufficient and necessary to guarantee that for the ensemble of block source codes of length n and rate R (described in Theorem 5, just set $\tilde{\phi} = \tilde{\phi}$) the average distortion converges to the fidelity level D exponentially with increasing n for all sources in the class.

Remark 6: The quantity $I(\tilde{\theta}, \tilde{\phi}; \tilde{\phi})$ given parametrically by

$$I(\tilde{\theta}, \tilde{\phi}; \tilde{\phi}) = \frac{1}{4\pi} \int_{\{\theta \leq \tilde{\phi}(\omega)\}} \ln \frac{\tilde{\phi}(\omega)}{\tilde{\theta}} \lambda(d\omega) \quad (53a)$$

$$D(\hat{\theta}, \hat{\phi}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \min\{\hat{\theta}, \hat{\phi}(\omega)\} \lambda(d\omega) = D \quad (53b)$$

represents the rate distortion function of the class determined by (35). For the same class of spectral measures this function was derived in [12] directly from the definition $\sup_{s \in S} R_s(D)$. No coding theorems were derived in [12]. In contrast, our emphasis in this Section was on the mismatch and the minimax robust theorems for block and trellis source coding.

The discussion for the choice of the operating point is similar to that which followed the proof of Theorem 3 and we do not repeat it here. The corresponding result for trellis codes is:

Theorem 8: Under the assumptions of Theorem 7 condition (52) guarantees that for the ensemble of trellis source codes of constraint length K and rate R the average distortion converges to the fidelity level D exponentially with increasing K for all sources in the class. Furthermore, if we define $(\rho', \theta') = \arg \min_{(\rho, \theta)} D_K(\rho, \theta, \hat{\phi}; \hat{\phi})$ where $-1 \leq \rho \leq E_0(\rho, \theta, \hat{\phi}; \hat{\phi})/R$ and $D(\theta, \hat{\phi}) \leq D$, then the following inequalities are true.

$$D_K(\rho', \theta', \hat{\phi}; \hat{\phi}) \leq D_K(\rho', \theta', \hat{\phi}; \hat{\phi}) \leq D_K(\rho, \theta, \hat{\phi}; \hat{\phi}). \quad (54)$$

It should be noted that the rate distortion function for the class of discrete-time SGS's with spectral uncertainty, which is given parametrically by (53a) - (53b), can also serve as an upper bound for the rate distortion function of class of discrete-time stationary ergodic non-Gaussian sources with the same spectral characteristics. This is the case, since the Gaussian source is known to have ([24, Section 4.6.2]) the largest rate distortion function among the stationary ergodic sources with the same spectral characteristics.

Finally, all the results of this section can be extended to *continuous-time* stationary Gaussian *bandlimited* (e.g., with spectral densities defined on $\Omega = [-\omega_0, \omega_0]$) sources. Since Huber-Strassen derivatives of capacities with respect to σ -finite (and not finite) measures can be constructed [22, Chapter IV], these results can be extended to *non-bandlimited* [that is, with spectral densities defined on $\Omega = (-\infty, \infty)$] sources.

IV. SUMMARY AND CONCLUSIONS

We have addressed the problem of minimax robust coding for sources with uncertainty in their statistical description. First, the mismatch source coding problem under a fidelity criterion was formulated and the appropriate coding theorems were established. Then, for uncertainty classes determined by 2-alternating capacities coding theorems were proved for discrete memoryless sources with uncertainty in the probability distribution and single-letter difference distortion criteria; and for stationary Gaussian sources with spectral uncertainty and the mean-square-error distortion criterion. It was established that there exist random block source codes and random trellis source codes such that the average distortion converges to the prescribed fidelity level exponentially with increasing block length or constraint length, respectively, for all sources in the class, provided that the code rates are larger than a critical rate. The rate distortion function for the class of sources and the distortion exponent were evaluated. These quantities, as well as the ensembles of random block and trellis source codes were characterized in terms of a Radon-Nikodym type derivative between the upper measure of the uncertainty class and a Lebesgue measure defined on the appropriate set.

REFERENCES

- [1] C. E. Shannon, "Coding theorems of a discrete source with a fidelity criterion," *IRE National Convention Record*, pt. 4, 142-163, March 1959.
- [2] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.
- [3] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw Hill, 1979.
- [4] F. Jelinek, "Tree encoding of memoryless time-discrete sources with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 584-590, 1969.
- [5] D. J. Sakrison, "The rate distortion function for a class of sources," *Information and Control*, vol. 15, pp. 165-195, 1969.
- [6] J. Ziv, "Coding and sources with unknown statistics," *IEEE Trans. Information Theory*, vol. IT-18, pp. 384-394, 1972.
- [7] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 783-795, 1973.
- [8] R. M. Gray and L. D. Davisson, "Source coding without the ergodic assumption," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 502-516, 1974.
- [9] M. B. Pursley and L. D. Davisson, "Variable rate coding for nonergodic sources and classes of ergodic sources subject to a fidelity constraint," *IEEE Trans. Inform. Theory*, IT-22, pp. 324-337, May 1976.
- [10] R. E. Krichevsky and V. K. Trofimov, "The performance of universal encoding," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 199-206, March 1981.
- [11] D. J. Sakrison, "Worst sources and robust codes for difference distortion measure," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 301-309, 1975.
- [12] H. V. Poor, "The rate distortion function on classes of sources determined by spectral capacities," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 19-26, January 1982.
- [13] L. D. Davisson and A. Leon-Garcia, "A source matching approach to finding minimax codes," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 166-174, March 1980.
- [14] D. Kazakos, "Robust noiseless coding through a game theoretic approach," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 576-583, July 1983.
- [15] G. Choquet, "Theory of capacities," *Ann. Inst. Fourier*, vol. 5, pp. 131-292, 1953/54.
- [16] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statistics*, vol. 36, pp. 1753-1758, 1965.
- [17] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 242-247, 1981.
- [18] K. S. Vastola and H. V. Poor, "On the p-point uncertainty class," *IEEE Trans. on Information Theory*, vol. IT-30, pp. 374-376, March 1984.
- [19] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statistics*, vol. 1, pp. 251-265, 1973.
- [20] H. V. Poor, "Minimax linear smoothing for capacities," *Ann. Probability*, vol. 10, 1982.

- [21] E. A. Geraniotis, "Minimax robust coding for channels with uncertain statistics," submitted to the *IEEE Trans. Inform. Theory*, 1984.
- [22] K. S. Vastola "Topics in robust statistical signal processing," Report R-965 (also Ph. D. dissertation), Coordinated Science Laboratory, Univ. of Illinois, Urbana, September 1982.
- [23] E. A. Geraniotis and D. Kazakos, "Robust noiseless fixed-length source coding", submitted to the *IEEE Trans. on Information Theory*, 1985.
- [24] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice Hall, 1971.
- [25] V. Grenader and G. Szego, *Toeplitz Forms Theory and Applications*, Univ. of California Press, 1958.