INFORMATION AND CONTROL **60**, 168–191 (1984)

A Game Theoretic Approach to Robust Filtering*

P. PAPANTONI-KAZAKOS

The University of Connecticut, Storrs, Connecticut 06268

A game theoretic approach to the filtering or smoothing problem is presented. A family of stationary information carrying processes and generalized models for the noise channel and the filter is considered. Sufficient conditions for the existence of saddle-point type solutions are stated. In addition, the solution for a special case of noise channel, a family of information carrying processes, and a nonlinear filter are found.

1. INTRODUCTION

The general objective of filtering of smoothing is the extraction of information data from noisy observations. The filtering problem and its solution are well established, when the stochastic processes that generate the information data and the noise are parametrically described. The reader may refer in this case to the books by Wiener (1949) and Hannan (1970). In this paper, we are concerned with the formulation and the solution of the filtering and smoothing problems when the statistical description of the stochastic processes that generate the information data and the noise is nonparametric. In particular, we will consider certain compact classes of stochastic processes, and we will formulate the problem as a stochastic game with saddle-point solution. To do that, we will first introduce our notation and our general assumptions.

We will name the stochastic process that generates the information data, *information carrying process*. We will denote this process by $[\mu, A, X]$, where μ is the measure, A is the alphabet of the process, and X is its name. We will assume that the process is discrete-time, and we will denote by x a given infinite sequence from this process. We will denote by x^l a given length l subsequence from the process and by X^l a sequence of l consequtive random variables from the process. Finally, we will denote $X_i^j = \{X_i, ..., X_j\}$; $j \geqslant i$, where X_i the random variable indicating the ith datum from the process and we will denote by $x_i^j = \{x_i, ..., x_j\}$, $j \geqslant i$, a given sequence of $j + 1 - i$ data from the process.

* This work was supported by the Air Force Office of Scientific Research under Grant AFOSR-78-3695C.

0019-9958/84 $3.00

We will represent the noisy environment by a stationary channel $[A, \nu, R[$, where A is the alphabet of the information carrying process $[\mu, A, X]$ and R is the real line indicating the alphabet of the process induced by $[\mu, A, X]$ and the channel. We will name the channel $[A, \nu, R]$, *noise channel.* We will denote by ν_x the measure induced by the channel and the infinite sequence x from the *information carrying process* $[\mu, A, X]$. We will denote by $[\mu\nu^{-1}, R, Y]$ the process induced by the process $[\mu, A, X]$ and the stationary channel $[A, \nu, R]$. We will name $[\mu\nu^{-1}, R, Y]$, *observation process,* and we will denote by y, Y^l, y^l, Y_i^j, y_i^j data sequences from this process, exactly as with the process $[\mu, A, X]$.

Adopting a generalized approach, we will represent the operation whose objective is the extraction of the data generated by the information carrying process $[\mu, A, X]$, by a stationary channel $[R, \sigma, B]$. B denotes the output to the channel alphabet, and for any given sequence y from the process $[\mu\nu^{-1}, R, Y]$, the channel induces in general a measure σ_y. We will name the channel $[R, \sigma, B]$, *information channel.* We will denote by $[\mu\nu^{-1}\sigma^{-1}, B, Z]$ the stochastic process induced by the observation process $[\mu\nu^{-1}, R, Y]$ and the information channel $[R, \sigma, B]$. We will name the process $[\mu\nu^{-1}\sigma^{-1}, B, Z]$, *matching process.*

The system described above, is exhibited in Fig. 1. The overall system performance should be represented by some appropriate measure of closeness between the information carrying process $[\mu, A, X]$ and the matching process $[\mu\nu^{-1}\sigma^{-1}, B, Z]$. An appropriate such measure, applicable to arbitrary stochastic processes, is the rho-bar distance. Let $\rho(x_i, z_i)$ be a distortion measure between the data values x_i, z_i. Given two sequences x^n, z^n, let us define a distortion measure $\rho_n(x^n, z^n)$ through the expression

$$\rho_n(x^n, z^n) = n^{-1} \sum_{i=1}^{n} \rho(x_i, z_i). \tag{1}$$

Given two stochastic processes $[\mu, A, X]$ and $[\lambda, A, Z]$, let μ^n, λ^n denote n dimensional restrictions of the measures μ and λ, respectively. Let $\mathscr{P}^n$ denote the family of all joint measures with marginals μ^n and λ^n. Then, the rho-bar distance $\bar{\rho}(\mu, \lambda)$ between the processes $[\mu, A, X]$ and $[\lambda, A, Z]$, and the

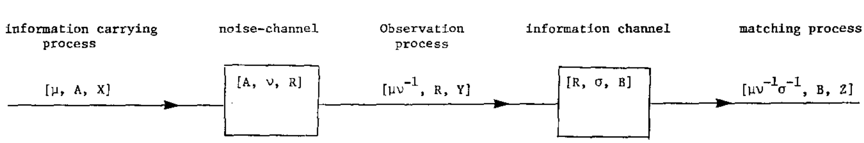

FIG. 1. Overall system.

Prohorov distance $\Pi_{\rho,n}(\mu^n, \lambda^n)$ between the n dimensional restrictions μ^n and λ^n are, respectively, defined as

$$\bar{\rho}(\mu, \lambda) = \sup_n \inf_{p^n \in \mathscr{P}^n} \int_{A^n \times A^n} \rho_n(X^n, Z^n)\, dp^n(X^n, Z^n) \tag{2}$$

$$\Pi_{\rho,n}(\mu^n, \lambda^n) = \inf_{p^n \in \mathscr{P}^n} \inf\{\delta : p^n(x^n, y^n : \rho_n(x^n, y^n) > \delta) \leqslant \delta\}.$$

If the distortion measure $\rho(\cdot, \cdot)$ is a metric on the real line, the rho-bar and Prohorov distances in (2) are also metrics on the corresponding sets of measures (Papantoni-Kazakos & Gray, 1979).

We note that for the correct formalization of the rho-bar distance, we selected identical alphabets for both the processes involved. Therefore, to appropriately express the rho-bar distance between the stochastic processes $[\mu, A, X]$ and $[\mu\nu^{-1}\sigma^{-1}, B, Z]$, we will assume that the two alphabets A and B are identical. Then, we select as the performance measure of the overall system in Fig. 1, the rho-bar distance $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$. The choice of the distortion measure $\rho(\cdot, \cdot)$ is left arbitrary at this point.

Assuming that the process $[\mu, A, X]$ and the stationary channel $[A, \nu, R]$ are given and that the distortion measure $\rho(\cdot, \cdot)$ has been selected a priori, the information channel $[R, \sigma, B]$ that minimizes the rho-bar distance $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$ can be found, at least theoretically. This is, in general, possible if an appropriate class $\mathscr{S}$ of channels $[R, \sigma, B]$ is first selected. Then the channel σ^* in $\mathscr{S}$ that minimizes the rho-bar distance $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$ is called the solution of the filtering or smoothing problem. A special such solution is, for example, the Wiener filter (Weiner, 1949).

Our objective in this paper is the design of a robust overall system. We wish to achieve good performance for a variety of information carrying processes and possibly a variety of noise channels. Our measure of performance is the rho-bar distance $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$. The variable in our control is the information channel $[R, \sigma, B]$. We will formalize the problem as a saddle-point game, and we will search for a saddle-point solution.

2. Game Formalization

Let us initially assume that the stationary noise channel $[A, \nu, R]$ remains fixed. Then, let us consider the following game played between nature and the system designer. The game starts with certain information available to both players. This information includes two given sets $\mathscr{M}$ and $\mathscr{S}$, as well as a given payoff function $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$. The knowledge of the payoff function implies knowledge of the distortion measure $\rho(\cdot, \cdot)$. The sets $\mathscr{M}$ and $\mathscr{S}$ are such that $\mu \in \mathscr{M}$ and $\sigma \in \mathscr{S}$.

Based on the above common information, nature selects some $\mu \in \mathscr{M}$. The system designer selects independently a stationary channel $\sigma \in \mathscr{S}$. After those selections have been completed, both nature and the system designer announce their choices. Following this announcement, the system designer pays nature a penalty equal to $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$. According to the rules of the game, whenever nature selects the measure μ, it guarantees for itself a gain equal to

$$\inf_{\sigma \in \mathscr{S}} \bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1}). \tag{3}$$

At the same time, the highest loss that the system designer may suffer whenever he selects the stationary channel σ is equal to

$$\sup_{\mu \in \mathscr{M}} \bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1}). \tag{4}$$

It is clearly concluded from expressions (3) and (4) that the highest gain nature can guarantee for itself is

$$\sup_{\mu \in \mathscr{M}} \inf_{\sigma \in \mathscr{S}} \bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1}). \tag{5}$$

The most optimistic loss that the system designer may expect is

$$\inf_{\sigma \in \mathscr{S}} \sup_{\mu \in \mathscr{M}} \bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1}). \tag{6}$$

If the expressions in (5) and (6) can be equal for some pair (μ^*, σ^*), then the change of either one of the members in the pair is of no advantage to either one of the players. If such a pair (μ^*, σ^*) exists, it is called the saddle point of the game. The corresponding value $\bar{\rho}(\mu^*, \mu^*\nu^{-1}\sigma^{*-1})$ is then called the saddle value of the game. Clearly, the following expression is also then satisfied:

$$\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{*-1}) \leqslant \bar{\rho}(\mu^*, \mu^*\nu^{-1}\sigma^{*-1}) \leqslant \bar{\rho}(\mu^*, \mu^*\nu^{-1}\sigma^{-1}) \qquad \forall \mu \in \mathscr{M}, \forall \sigma \in \mathscr{S}. \tag{7}$$

In our search for saddle-point solutions of the game stated above, we will use the notions of convexity, closeness, and compactness of sets of measures. Convexity is defined in the usual sense (Royden, 1963). The definitions of closeness and compactness require the use of metrics on measure spaces. Let $\gamma(\cdot, \cdot)$ be a metric on the real line. Then, as we stated in the introduction, the distances $\bar{\gamma}(\mu, \lambda)$ and $\Pi_{\gamma,n}(\mu^n, \lambda^n)$ in (2) are also metrics on the space of measures. We thus proceed with the following definition, that also includes the notion of zero-memory channels.

DEFINITION 1. Let $\mathscr{M}$ be a set of stationary measures μ. Let $\mathscr{S}$ be a set of stationary channels σ. Let $\gamma(\cdot, \cdot)$ be a metric on the real line. Then,

(A) The set $\mathscr{M}$ is *closed and compact*, if it is closed and compact (Royden, 1963) with respect to the metric $\bar{\gamma}(\mu, \lambda)$, $\lambda \in \mathscr{M}$.

(B) The set $\mathscr{S}$ is *closed*, if it is closed (Royden, 1963) with respect to the metric $\Pi_{\gamma,1}(\sigma^1_{1y}, \sigma^1_{2y})$, $\forall y$, $\sigma_1, \sigma_2 \in \mathscr{S}$, where σ^1_{1y} and σ^1_{2y} denote the 1-dimensional restrictions of the measures σ_{1y} and σ_{2y}, respectively.

(C) The stationary channel $[R, \sigma, B]$ is zero memory, iff

$$\sigma^n_y(Z^{n-1}_0 \in B_0 \times B_1 \times \cdots \times B_{n-1}) = \prod_{j=0}^{n-1} \sigma^1_y(Z_j \in B_j)$$

$$\forall n, \forall B_j \in \mathscr{B}, 0 \leqslant j \leqslant n-1. \tag{8}$$

To simplify as well as generalize the presentation of some additional notions needed, we will denote by $K(\mu, \sigma)$ some real valued nonnegative function on the measure μ and the stationary channel σ. Such a function represents the payoff in the game described in this section, and it may be the rho-bar distance $\bar{\rho}(\mu, \mu v^{-1}\sigma^{-1})$. Then, we proceed with

DEFINITION 2. Let $\mathscr{M}$ and $\mathscr{S}$ be two given sets, such that $\mu \in \mathscr{M}$ and $\sigma \in \mathscr{S}$. Let $\mathscr{S}$ be a set of zero memory channels. Let $K(\mu, \sigma)$ be some real-valued nonnegative function between $\mu \in \mathscr{M}$ and $\sigma \in \mathscr{S}$. Then

(A) $K(\mu, \sigma)$ is *finite* on $\mathscr{M} \times \mathscr{S}$, iff there exists some finite positive number α, such that

$$K(\mu, \sigma) \leqslant \alpha \qquad \forall \mu \in \mathscr{M}, \forall \sigma \in \mathscr{S}.$$

(B) $K(\mu, \sigma)$ is *continuous* on $\mathscr{M} \times \mathscr{S}$, iff

(i) $K(\mu, \sigma)$ is continuous in μ on $\mathscr{M}$ with respect to the metric $\bar{\gamma}$; that is, iff: Given $\mu_0 \in \mathscr{M}$, given $\sigma \in \mathscr{S}$, given $\varepsilon > 0$, there exists $\delta(\varepsilon, \mu_0, \sigma) > 0$ such that

$$\mu \in \mathscr{M}, \bar{\gamma}(\mu_0, \mu) < \delta(\varepsilon, \mu_0, \sigma) \to |K(\mu_0, \sigma) - K(\mu, \sigma)| < \varepsilon;$$

(ii) $K(\mu, \sigma)$ is continuous in σ on $\mathscr{S}$, with respect to the metric $\Pi_{\gamma,1}$; that is, denoting by y infinite sequences of channel inputs, and given $\sigma_0 \in \mathscr{S}$, given $\mu \in \mathscr{M}$, given $\varepsilon > 0$, given $\eta > 0$, there exist set $A(\sigma_0, \mu, \varepsilon) \in R^\infty$ and $\delta(\varepsilon, \mu, \sigma) > 0$, such that

$$\mu(A(\sigma_0, \mu, \varepsilon)) > 1 - \eta,$$

$$\sigma \in \mathscr{S}, \qquad \Pi_{\gamma,1}(\sigma^1_{0y}, \sigma^1_y) < \delta(\varepsilon, \mu, \sigma_0)$$

$$\forall y \in A(\sigma_0, \mu, \varepsilon) \to |K(\mu, \sigma_0) - K(\mu, \sigma)| < \varepsilon.$$

(C) $K(\mu, \sigma)$ is *concave-convex* on $\mathscr{M} \times \mathscr{S}$, iff:

(i) $K(\lambda\mu_1 + (1-\lambda)\mu_2, \sigma) \geqslant \lambda K(\mu_1, \sigma) + (1-\lambda) K(\mu_2, \sigma)$, $\forall \mu_1, \mu_2 \in \mathscr{M}$, $\forall \sigma \in \mathscr{S}$, $\forall \lambda: 0 \leqslant \lambda \leqslant 1$, and $[\lambda\mu_1 + (1-\lambda)\mu_2] \in \mathscr{M}$,

(ii) $K(\mu, \lambda\sigma_1 + (1-\lambda)\sigma_2) < \lambda K(\mu, \sigma_1) + (1-\lambda) K(\mu, \sigma_2)$, $\forall \mu \in \mathscr{M}$, $\forall \sigma_1, \sigma_2 \in \mathscr{S}$, $\sigma_1 \neq \sigma_2$ $\forall \lambda$: $0 < \lambda < 1$ and $[\lambda\sigma_1 + (1-\lambda)\sigma_2] \in \mathscr{S}$.

In part (C) of Definition 2, the conditions $[\lambda\mu_1 + (1-\lambda)\mu_2] \in \mathscr{M}$ and $[\lambda\sigma_1 + (1-\lambda)\sigma_2] \in \mathscr{S}$ for every λ, $0 \leqslant \lambda \leqslant 1$, are always satisfied if the sets $\mathscr{M}$ and $\mathscr{S}$ are convex.

Based on Definitions 1 and 2, we can now express Theorem 1 whose proof is in Appendix A.

THEOREM 1. Let $\mathscr{M}$ be a set of measures μ. Let $\mathscr{S}$ be a set of stationary, zero-memory channels. Let $K(\mu, \sigma)$ be a real valued nonnegative function between μ and σ.

Let $\mathscr{M}$ and $\mathscr{S}$ be nonempty, convex, and closed sets. Let, in addition, the set $\mathscr{M}$ be compact. Let $K(\mu, \sigma)$ be finite, continuous, and concave–convex on $\mathscr{M} \times \mathscr{S}$. Then, there exists saddle point (μ^*, σ^*) on $\mathscr{M} \times \mathscr{S}$, that is,

$$\inf_{\sigma \in \mathscr{S}} \sup_{\mu \in \mathscr{M}} K(\mu, \sigma) = \sup_{\mu \in \mathscr{M}} \inf_{\sigma \in \mathscr{S}} K(\mu, \sigma) = K(\mu^*, \sigma^*).$$

In this section we laid the foundations for a game-oriented solution to the general problem stated in the Introduction. The appropriateness of any specific payoff function $K(\mu, \sigma)$ must be studied within the guidelines provided by Theorem 1. If, as initially suggested, the payoff function is the rho-bar distance $\bar{\rho}(\mu, \mu\nu^{-1}\sigma^{-1})$ then its properties must be studied on an appropriately selected set $\mathscr{M} \times \mathscr{S}$. This task will be undertaken in the following section.

3. Analysis-Sufficient Conditions

Let us consider the overall system in Fig. 1. Let us assume that the noise channel $[A, \nu, R]$ is fixed and it is stationary. Let us assume that the alphabets A and B are identical. Let us denote by $\mathscr{F}_s$ the family of all stationary processes whose alphabet is A. Let us consider this alphabet known to the system designer. Let μ_0 be some given measure in $\mathscr{F}_s$. Let μ_0 be known to the system designer. Then, select a set $\mathscr{M}$ of measures μ in $\mathscr{F}_s$ the following way:

Select some metric $\gamma(\cdot, \cdot)$ on A, and some positive finite constant α. Then define $\mathscr{M}$ as

$$\mu \in \mathscr{M} \text{ iff:} \qquad \mu \in \mathscr{F}_s \qquad \text{and} \qquad \bar{\gamma}(\mu_0, \mu) \leqslant \alpha. \tag{A}$$

The set $\mathscr{M}$ defined above is nonempty and convex, as well as closed and compact, with respect to the metric $\bar{\gamma}$. Also, since the set $\mathscr{M}$ is a set of stationary processes and the noise channel $[A, v, R]$ is also stationary, the observation process $[\mu v^{-1}, R, Y]$ is stationary for all μ in $\mathscr{M}$.

Let us denote by $\mathscr{D}_{ls}$ the family of all stationary, zero-memory channels whose input alphabet is R, whose output alphabet is B (identical with A), and which operate on input sequences of length l per channel output element. Let Q_l be a class contained in $\mathscr{D}_{ls}$. Such a class Q_l includes members in $\mathscr{D}_{ls}$ with possibly specific properties, where such properties will be identified later. The class Q_l may include deterministic channels as well. Then, the measures $\sigma^1_{y^l}$ reduce to deterministic functions, and the Prohorov distance $\Pi_{\gamma,1}(\sigma^1_{0y^l}, \sigma^1_{y^l})$ reduces to $\gamma(\sigma^1_{0y^l}, \sigma^1_{y^l})$.

We have already selected a metric $\gamma(\cdot,\cdot)$ on the alphabet A. Let now $\xi(\cdot,\cdot)$ be some metric on the real line R. Let $\rho(\cdot,\cdot)$ and $\tau(\cdot,\cdot)$ be two distortion measures. Let $\rho(\cdot,\cdot)$ be used for information carrying processes $\mu \in \mathscr{M}$, and for matching processes $[\mu v^{-1}\sigma^{-1}, B, Z]$. Let $\tau(\cdot,\cdot)$ be used for observation processes $[\mu v^{-1}, R, Y]$. Let $\rho(\cdot,\cdot)\,\tau(\cdot,\cdot)$ be such that, given $\varepsilon > 0$ there exist $\delta_1 > 0$, $\delta_2 > 0$, $\delta_3 > 0$, $\delta_4 > 0$, such that

$$\begin{aligned} &\rho(x,y) < \delta_1 \to \gamma(x,y) < \varepsilon \\ &\gamma(x,y) < \delta_2 \to \rho(x,y) < \varepsilon \\ &\xi(x,y) < \delta_2 \to \tau(x,y) < \varepsilon \\ &\tau(x,y) < \delta_1 \to \xi(x,y) < \varepsilon. \end{aligned} \tag{9}$$

One $\gamma(\cdot,\cdot)$, $\rho(\cdot,\cdot)$ combination that satisfies the conditions in (9) is $\gamma(x,y) = |x-y|$, $\rho(x,y) = (x-y)^2$. Similarly for $\xi(\cdot,\cdot)$ and $\tau(\cdot,\cdot)$.

As in (Papantoni-Kazakos & Gray, 1979; Papantoni-Kazakos, 1981), we need a definition of empirical measures. Given a sequence x^n, we form an empirical measure μ_{x^n} through the operation

$$\mu_{x^n}(D) = \sum_{i: T^i x \in D} n^{-1}, \qquad D \in \mathscr{A}^\infty, \tag{10}$$

where $x = (..., x^n, x^n, ...)$, T indicates one step shift in time, and $\mathscr{A}^\infty$ the infinite product of the σ-algebra $\mathscr{A}$ of sets on the space on which each datum X_i assumes values. The properties of the empirical measure in (10) can be found in (Papantoni-Kazakos & Gray, 1979). Now we need

DEFINITION 3. (i) For given finite l, the stationary, zero-memory channel σ_l is *continuous* if given $y^l \in R^l$, $\varepsilon > 0$, there exists $\delta = \delta(l, y^l, \varepsilon) > 0$ such that

$$\xi_l(y^l, x^l) < \delta \to \Pi_{\gamma,1}(\sigma^1_{l,y^l}, \sigma^1_{l,x^l}) < \varepsilon.$$

(ii) The sequence $\{\sigma_l\}$ of stationary, zero-memory channels is *continuous at the measure* λ if given $\varepsilon > 0$, $\eta > 0$, there exist integers k, l_0, some $\delta > 0$, and for each $l < l_0$ some set $E^l \in R^l$ with $\lambda^l(E^l) > 1 - \eta$, such that for each $x^l \in E^l$, $y^l \in R^l$ with the property

$$\Pi_{\xi,k}(\mu^k_{x^l}, \mu^k_{y^l}) < \delta,$$

it is implied that

$$\Pi_{\gamma,1}(\sigma^1_{l,x^l}, \sigma^1_{l,y^l}) < \varepsilon.$$

In part (ii) of Definition 3, we have assumed that the measure λ is defined on the alphabet R. The above definition is exactly Definition 3 in (Papantoni-Kazakos, 1981). The sequence $\{\sigma_l\}$ is generated by varying l values. Also, $\mu^k_{x^l}$ denotes k-dimensional restrictions of the empirical measure μ_{x^l} defined by (10); $\Pi_{\xi,k}$ denotes the Prohorov distance on k-dimensional restrictions of measures, where the metric ξ is used; λ^l denotes the l-dimensional restriction of the measure λ.

A definition parallel to Definition 3 applies to the noise channel $[A, v, R]$, assuming that this channel is also a zero memory channel. Then, the metrics $\xi(\cdot, \cdot)$ and $\gamma(\cdot, \cdot)$ in Definition 3 should be reversed. Also, λ should be then substituted by μ, and R^l should be substituted by A^l.

The consideration of noise and information channels which are either continuous or continuous at some measure (Definition 3) will be valuable in securing properties which make our game approach meaningful. In the remainder of this section, we will present a number of theorems and lemmas. In the next section, we will discuss the implications in the statements of those theorems and lemmas.

THEOREM 2. *Let the alphabets A and B be identical. Let the metric $\gamma(\cdot, \cdot)$ be used on A. Let the metric $\xi(\cdot, \cdot)$ be used on R. Let the distortion measure $\rho(\cdot, \cdot)$ be used on the measures μ and $\mu v^{-1}\sigma^{-1}$.* Let the distortion measure $\tau(\cdot, \cdot)$ be used on the measure μv^{-1}. *Let $\gamma(\cdot, \cdot)$, $\xi(\cdot, \cdot)$, $\rho(\cdot, \cdot)$, $\tau(\cdot, \cdot)$ satisfy the conditions in* (9), *and let $\rho(\cdot, \cdot)$, $\tau(\cdot, \cdot)$ be nonnegative. Let, in addition, either one of the following two conditions be satisfied:*

(1) *Both $\rho(\cdot, \cdot)$ and $\tau(\cdot, \cdot)$ are bounded.*

(2) *$\tau(\cdot, \cdot)$ is bounded, $\rho(\cdot, \cdot)$ takes bounded values for bounded values of its arguments, and the alphabet A is bounded.*

Let the channel $[A, v_m, R]$ be zero memory and stationary, operating on input sequences of some fixed finite length m. Let v be also continuous (Definition 3(i)).

Let $\mu \in \mathscr{F}_s$, where $\mathscr{F}_s$ the class of stationary measures. Let the class Q_l of stationary, zero-memory channels $[R, \sigma_l, B]$ be considered, which also *operate on input sequences of fixed finite length l and which are, in addition, continuous (Definition* 3(i)).

Then, the rho-bar distance $\bar{\rho}(\mu, \mu v_m^{-1} \sigma_l^{-1})$ *is continuous on* $\mathscr{F}_s \times Q_l$ (*Definition* 2(B)), *and it is continuous in* μ *with respect to* $\bar{\rho}$.

The proof of Theorem 2 is included in Appendix A. We express some conclusions from Theorem 2 via a corollary.

COROLLARY. *Let the noise channel* v_m *be zero memory and stationary, operating on input sequences of some fixed finite length m. Let v be also continuous (Definition* 3(i)).

Let Q_l *be the class of stationary, zero-memory channels, which operate on input sequences of fixed finite length l, and which are also continuous (Definition* 3(i)). *Let* $\mathscr{F}_{\mathrm{sb}}$ *be the class of stationary measures with bounded support (with bounded alphabets). Let* $\mathscr{M}$ *and* $\mathscr{S}_l$ *be two sets such that*

$$\mathscr{M} \subset \mathscr{F}_{\mathrm{sb}}, \qquad \mathscr{S}_l \subset Q_l.$$

Let $\gamma(\cdot, \cdot)$, $\xi(\cdot, \cdot)$, $\rho(\cdot, \cdot)$, $\tau(\cdot, \cdot)$ *be as in Theorem* 2. *Let* $\tau(\cdot, \cdot)$ *be bounded, and let* $\rho(\cdot, \cdot)$ *assume bounded values for bounded values of its arguments.*

Then, the rho-bar distance $\bar{\rho}(\mu, \mu v_m^{-1} \sigma_l^{-1})$ *is continuous on* $\mathscr{M} \times \mathscr{S}_l$ (*Definition* 2(B)) *and it is continuous in* μ *with respect to* $\bar{\rho}$.

As we will see in the next section, the conditions and conclusions expressed in the corollary are important for the design of overall systems with robust performance. Let us now consider the case of stationary channels $[R, \sigma, B]$ that operate on assymptotically long input sequences per single output element. Considering the metrics $\gamma(\cdot, \cdot)$ and $\xi(\cdot, \cdot)$ as before, we first present the following definition.

DEFINITION 4. The sequence $\{\sigma_l\}$ of stationary, zero-memory channels operating on the input alphabet R is *assymptotically continuous*, if given $\varepsilon > 0$, there exist integers k, l_0, some $\delta > 0$, and for each $l > l_0$ some set $E^l \in R^l$, such that for each $x^l \in E^l$, $y^l \in R^l$ with the property

$$\Pi_{\xi,k}(\mu_{x^l}^k, \mu_{y^l}^k) < \delta,$$

it is implied that

$$\Pi_{\gamma,1}(\sigma_{l,x^l}^1, \sigma_{l,y^l}^1) < \varepsilon.$$

The notation in Definition 4 is as in Definition 3. We notice that a sequence $\{\sigma_l\}$ which is asymptotically continuous is not matched to any particular measure. This is in contrast to continuity at some measure in Definition 3. We now proceed with a theorem and a lemma. The proof of the theorem is in Appendix A.

THEOREM 3. *Given a fixed channel v_m as in Theorem 2, given nonempty, convex, and compact with respect to the metric $\bar{\gamma}$ set $\mathscr{M}$ of stationary measures μ, there exists an asymptotically continuous (Definition 4) sequence $\{\sigma_l\}$ of stationary zero-memory channels, which is also continuous at every measure μv_m^{-1} (Definition 3(ii)) such that $\mu \in \mathscr{M}$. This sequence $\{\sigma_l\}$ is then called continuous at the set $\mathscr{M}$.*

LEMMA 1. *Let the channel v_m as well as $\gamma(\cdot, \cdot)$, $\xi(\cdot, \cdot)$, $\rho(\cdot, \cdot)$, and $\tau(\cdot, \cdot)$ be as in the corollary. Let $\mathscr{M}$ be a nonempty, convex, and compact with respect to $\bar{\gamma}$ set of stationary measures, with bounded support. Let $\mathscr{S}$ be a set of stationary, zero-memory sequences $\{\sigma_l\}$ of channels which are also continuous at the set $\mathscr{M}$ (Theorem 3). Then, the rho-bar distance $\bar{\rho}(\mu, \mu v_m^{-1}\sigma^{-1})$ is continuous on $\mathscr{M} \times \mathscr{S}$.*

The proof of Lemma 1 is as the proof of Theorem 2, where the result in Theorem 2 of (Papantoni-Kazakos, 1981) is used to guarantee rho-bar stability of the measures $\mu v_m^{-1}\sigma^{-1}$.

LEMMA 2. *Let the noise channel v_m as well as $\gamma(\cdot, \cdot)$, $\xi(\cdot, \cdot)$, $\rho(\cdot, \cdot)$, and $\tau(\cdot, \cdot)$ be as in Theorem 2. Let either one of the following conditions be true:*

(1) *$\mathscr{M}$ is a nonempty convex set of stationary measures. If $\rho(\cdot, \cdot)$ is not bounded, assume that the measures in $\mathscr{M}$ also have bounded support. $\mathscr{S}$ is a nonempty convex set of stationary, zero-memory channels operating on input sequences of fixed finite length l.*

(2) *$\mathscr{M}$ is a nonempty, convex, and compact with respect to $\bar{\gamma}$ set of stationary measures. If $\rho(\cdot, \cdot)$ is unbounded, assume that the measures in $\mathscr{M}$ have bounded support. $\mathscr{S}$ is a nonempty convex set of stationary, zero-memory sequences $\{\sigma_l\}$ of channels that are also continuous at the set $\mathscr{M}$ (Theorem 3).*

Then, the Prohorov distance $\Pi_{\rho,1}(\mu, \mu v_m^{-1}\sigma^{-1})$ is concave–convex on $\mathscr{M} \times \mathscr{S}$ (Definition 2(C)). Also, the distances $\Pi_{\rho,1}(\mu, \mu v_m^{-1}\sigma^{-1})$ and $\bar{\rho}(\mu, \mu v_m^{-1}\sigma^{-1})$ are then equivalent.

The proof of Lemma 2 is in Appendix A. We complete the major part of this section by expressing a final theorem. The theorem basically summarizes the results from Theorem 1, Section 2, in conjuction with the results from the theorems and lemmas in this section.

THEOREM A. *Let the alphabets A and B be identical. Let the metric $\gamma(\cdot, \cdot)$ be used on A. Let the metric $\xi(\cdot, \cdot)$ be used on R. Let the distortion measure $\rho(\cdot, \cdot)$ be used on the measures μ and $\mu v^{-1}\sigma^{-1}$. Let the distortion measure $\tau(\cdot, \cdot)$ be used on the measure μv^{-1}. Let $\gamma(\cdot, \cdot)$, $\xi(\cdot, \cdot)$, $\rho(\cdot, \cdot)$, and $\tau(\cdot, \cdot)$ satisfy the condition (9). Let $\rho(\cdot, \cdot)$ and $\tau(\cdot, \cdot)$ be nonnegative.*

Let the channel $[A, v_m, R]$ *be considered. Let* v_m *be stationary and zero memory, operating on input sequences of fixed finite length m. Let* v_m *be also continuous* (*Definition* 3(i)).

Let $\mathscr{M}$ *be a nonempty and convex, as well as closed and compact w.r.t.* $\bar{\gamma}$ *set of stationary measures* μ. *If* $\rho(\cdot, \cdot)$ *unbounded, let the measures* μ *in* $\mathscr{M}$ *also have bounded support.*

Let $\mathscr{S}_l$ *be a nonempty, convex, and closed w.r.t.* $\Pi_{\gamma,1}$ *set of stationary, zero-memory channels* σ_l *that operate on input sequences of fixed finite length l and that are also continuous* (*Definition* 3(i)). *If* $\rho(\cdot, \cdot)$ *unbounded, also let the output alphabet of each channel* σ_l *in* $\mathscr{S}_l$ *be bounded.*

Let $\mathscr{S}$ *be a nonempty, convex, and closed w.r.t.* $\Pi_{\gamma,1}$ *set of stationary, zero-memory channel sequences* $\{\sigma_l\}$, *that are also continuous at the set* $\mathscr{M}$ (*Theorem* 3).

Then, the rho-bar distance $\bar{\rho}(\mu, \mu v_m^{-1} \sigma^{-1})$ *is finite, continuous* (*Definition* 2(B)), *and concave–convex* (*Definition* 2(C)) *on* $\mathscr{M} \times \mathscr{S}_l$ *and on* $\mathscr{M} \times \mathscr{S}$. *Therefore, there exist then saddle-point pairs* (μ^*, σ_l^*) *and* $(\mu^*, \{\sigma_l^*\})$, *and corresponding saddle values* $\bar{\rho}(\mu^*, \mu^* v_m^{-1} \sigma_l^{*-1})$ *and* $\bar{\rho}(\mu^*, \mu^* v_m^{-1} \sigma^{*-1})$, *where* $\sigma^* \to \{\sigma_l^*\}$.

From the analysis and derivations in the present section, it is clear that a game formalization is meaningful only under certain conditions. Such restrictive conditions are mainly imposed on the noise and information channels. In particular, continuity of both channels (Definition 3) is essential for the guarantee of saddle-point solutions. We must point out here that the consideration of stationary information channels clearly includes such deterministic channels as well. Indeed, the preceeding analysis carries through then, where in the definitions of channel continuity (Definition 3), the Prohorov distance $\Pi_{\gamma,1}(\sigma^1_{l,y^l}, \sigma^1_{l,x^l})$ reduces to $\gamma(\sigma_l^1(x^l), \sigma_l^1(y^l))$; $\sigma_l^1(x^l)$ represents then some deterministic function.

Until now we assumed that the noise channel $[A, v, R]$ is given and fixed. However, deviations from such a given description are possible and realistic. It is desirable, therefore, to study the behavior of our game approach in the presence of such deviations. We first proceed with

DEFINITION 5. Let Q_m be the class of zero-memory stationary channels that operate on input sequences of fixed finite length m, per output element. Let the input alphabet of the channels in Q_m be A. Let the output alphabet be R. Let v_{0m} be some given channel in Q_m. Let $\mathscr{M}$, $\mathscr{S}$, $\mathscr{S}_l$ be sets as in Theorem A. Let (μ_0^*, σ_{0l}^*) be the saddle-point solution for the noise channel v_{0m} and for $\mu \in \mathscr{M}$, $\sigma_l \in \mathscr{S}_l$. Let $(\mu_0^*, \{\sigma_{0l}^*\})$ be the saddle-point solution for v_{0m}, $\mu \in \mathscr{M}$, $\{\sigma_l\} \in \mathscr{S}$.

Then, the solutions (μ_0^*, σ_{0l}^*), $(\mu_0^*, \{\sigma_{0l}^*\})$ are called *robust at* v_{0m} iff: Given

$\varepsilon > 0$, there exist numbers $\delta_1 > 0$, $\{\delta_l\}$: $\delta_l > 0$; $\forall l > l_0$, and some set $\Delta^m \in A^m$, such that, respectively,

(1) $v_m \in Q_m$, $\Pi_{\gamma,1}(v^1_{m,x^m}, v^1_{0m,x^m}) < \delta_1$, $\forall x^m \in \Delta^m \rightarrow |\bar{\rho}(\mu_0^*, \mu_0^* v_{0m}^{-1} \sigma_{0l}^{*-l}) - \bar{\rho}(\mu_0^*, \mu_0^* v_m^{-1} \sigma_{0l}^{*-1})| < \varepsilon$;

(2) for each $l > l_0$,

$$v_m \in Q_m, \quad \Pi_{\gamma,1}(v^1_{m,x^m}, v^1_{0m,x^m}) < \delta_l, \ \forall x^m \in \Delta^m \rightarrow |\bar{\rho}(\mu_0^*, \mu_0^* v_{0m}^{-1} \sigma_{0l}^{*-l}) - \bar{\rho}(\mu_0^*, \mu_0^* v_m^{-1} \sigma_{0l}^{*-1})| < \varepsilon.$$

It is clear that Δ^m can be selected as a high probability set for the m-dimensional restriction of the measure μ_0^*, and that then the conditions (1) and (2) in Definition 5 imply continuity of the measure $\mu_0^* v_m^{-1}$ at v_{0m}. This continuity carries then over to the rho-bar distances $\bar{\rho}(\mu_0^*, \mu_0^* v_m^{-1} \sigma_{0l}^{*-1})$, as exhibited in the proofs of Theorem 2 and Lemma 1. Therefore, the following proposition holds.

PROPOSITION 1. *Given Q_m as in Definition 5, given $v_{0m} \in Q_m$, given $\mathscr{M}$, $\mathscr{S}_l$, $\mathscr{S}$ as in Theorem* A, *the saddle-point solutions (μ_0^*, σ_{0l}^*) and $(\mu_0^*, \{\sigma_{0l}^*\})$ (as in Definition* 5) *are robust at v_{0m}.*

The conclusion from Proposition 1 is that continuity (Definition 3, Theorem 3) of both the noise and information channels, in addition to guaranteeing the existence of saddle-point solution at some noise channel v_{0m}, it also guarantees robustness at v_{0m}. The implications behind the conclusions in Theorem A deserve special focusing. They will be discussed in the next section.

4. INTERPRETATION OF SOME ASSUMPTIONS AND CONCLUSIONS

In Theorem A, the consideration of noise channels v_m that are zero memory and that operate on input sequences of fixed, finite length m, has special meaning. It means that the noisy environment is represented by distortions which are influenced by the values of a whole sequence of data from the information-carrying process. This generalizes the noise notion. A special case with $m = 1$ is the case of an additive, memoryless noise channel.

If the alphabet A is such that $A \subset R$, representing a bounded interval on the real line R, bandwidth expansion caused by the noise channel is in general implied. A bounded distortion measure $\tau(\cdot, \cdot)$ on R represents then elimination of extreme values. Such values are representing just noise rather than information-carrying data.

If the distortion measure $\rho(\cdot, \cdot)$ on A is unbounded (such as the mean square such measure), boundness of A is required. That implies that the information channel maps unbounded input sequences onto bounded intervals on the real line, if $A \subset R$. That is, the information channel eliminates then extreme values, performing some jackknifing type of operation. Clearly, deterministic information channels that perform linear operations on input data are then excluded.

In the case that the information channel operates on asymptotically long input sequences, continuity at the whole set $\mathscr{M}$ of information carrying processes is required. This continuity is projected through empirical measures on data sequences. This projection presents another reason why information channels that operate on linear transformations of input data should be excluded.

As a conclusion from the above discussion, the set $\mathscr{S}_l$ in Theorem A includes no linear deterministic channels. The set $\mathscr{S}$ in Theorem A includes no linear channels in the sense of operation on linear transformations of input data sequences.

5. A Special Case

Let A be a finite subinterval of the real line R. Let μ_0 be a given measure determining the stationary stochastic process $[\mu_0, A, X]$. Let $\gamma(x,y) \triangleq \xi(x,y) \triangleq \tau(x,y) \triangleq |x-y|$. Let $\rho(x,y) \triangleq (x-y)^2$. For this $\rho(\cdot,\cdot)$, let α be a known positive constant and let

$$\mathscr{M}: \bar{\rho}(\mu_0, \mu) \leqslant \alpha. \tag{11}$$

Let the noise channel $[A, \nu_m, R]$ be memoryless, additive, and Gaussian. That is, $\nu^l_{x^m} \triangleq \nu^l_{x^1}$ and

$$\Pr\left\{ Y_0 \leqslant y \;\middle|\; X_0 = x \right\} = \int_{-\infty}^{y} \frac{\exp\{-(u-x)^2/2\sigma^2\}}{\sqrt{2\pi}\,\sigma}\, du = \Phi\left(\frac{y-x}{\sigma}\right). \tag{12}$$

Let $g(\cdot)$ be a deterministic, nonlinear operation, defined as

$$\begin{aligned} g(u) &= B_1;\ -\infty < u \leqslant b_1, \\ &\ \vdots \\ &= B_k;\ b_{k-1} < u \leqslant b_k, \\ &\ \vdots \\ &= B_n;\ b_{n-1} < u < \infty, \end{aligned} \tag{13}$$

where $B_1 \geqslant 0$, $B_k < B_{k+1}$, and $b_k < b_{k-1}$, $\forall k$, $B_n < \infty$.

Let the nonlinear operation $g(\cdot)$ be applied on the infinite observation sequence ..., y_{-1}, y_0, y_1,..., to generate the infinite sequence ..., $g(y_{-1})$, $g(y_0)$, $g(y_1)$,... .

Then, due to the noise channel model in (12), we have

$$G(x) \triangleq E\left\{ g(Y_j) \,\middle|\, X_j = x \right\} = B_n - \sum_{k=2}^{n} (B_k - B_{k-1})\Phi\left(\frac{b_{k-1} - x}{\sigma}\right). \quad (14)$$

It can be easily seen that the function $G(x)$ in (14) is strictly monotone and bounded, since $B_1 \leqslant G(x) \leqslant B_n$, $\forall x$. Defining

$$c \triangleq \frac{B_n - B_1}{\sqrt{2\pi}}, \quad (15)$$

we easily conclude that

$$\frac{dG(x)}{dx} \leqslant c, \ \forall x,$$

and therefore

$$|G(x) - G(y)| \leqslant c\,|x - y|, \ \forall x, y. \quad (16)$$

Let us consider the nonlinearity in (13) fixed and known. We define then a convex and closed w.r.t. $\Pi_{\gamma,1}$ set $\mathscr{S}$ of zero-memory sequences $\{\sigma_l\}$ of deterministic information channels in the following way:

$\mathscr{S}$: for each l, the observation sequence y_{-l}^{-1} is first transformed to the sequence $\{g(y_j), -l \leqslant j \leqslant -1\}$. Then, the linear transformation

$$\sum_{j=-l}^{-1} a_{lj}\, g(y_j) \text{ is performed to map } G(x_0), \text{ where } \{a_{lj}; -l \leqslant j \leqslant -1\};$$

$$\sum_{j=-l}^{-1} a_{lj} = 1, \qquad a_{lj} \geqslant 0, \ \forall j. \quad \text{(B)}$$

Furthermore, given $\varepsilon > 0$ there exists positive integer

$$l_0 \colon \sum_{j=-l}^{-1} a_{lj}^2 < \varepsilon, \qquad \forall l > l_0.$$

At this point we should observe that since the function $G(x)$ in (14) is strictly monotone, if $G(X_0)$ is estimated, X_0 can be recovered uniquely from this estimate. Also, in (B) we have selected arbitrarily a class of predictive filters in $G(x_0)$. We now proceed with

LEMMA 3. *The sequence* $\{\sigma_l\}$ *of deterministic channels in* (B) *is asymptotically continuous* (*Definition* 4).

The proof of Lemma 3 is in Appendix B. The sequence $\{\sigma_l\}$ is asymptotically continuous on R^l for each l larger than some l_0. Let μ be some measure in the set $\mathscr{M}$ defined by expression (11). Let us then denote by $P_\mu(\lambda)$, $-\pi \leqslant \lambda \leqslant \pi$ the spectral density of the process $[\mu, A, X]$. Let μG^{-1} denote the measure induced by the measure μ in $\mathscr{M}$ and the strictly monotone nonlinearity $G(\cdot)$ defined by expression (14). Let us denote by $P_{G\mu}(\lambda)$, $-\pi \leqslant \lambda \leqslant \pi$ the spectral density of the process $[\mu G^{-1}, B, W]$, where $W = G(X)$ and $B = [B_1, B_n]$. Let us denote by $\mathscr{M}'$ the convex, closed, and compact set of spectral densities

$$\mathscr{M}': (2\pi)^{-1} \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - P_{G\mu}^{1/2}(\lambda)]^2 \, d\lambda \leqslant ca, \tag{17}$$

where $P_{G\mu_0}(\lambda)$ the spectral density induced by the given measure μ_0, and the nonlinearity $G(\cdot)$, and c the constant in (15).

We can now express Lemma 4, whose proof is in Appendix B.

LEMMA 4. *Let* $\rho(x, y) \triangleq (x - y)^2$. *Let* $\mathscr{M}$ *be the set of measures given by* (11). *Let* $G(\cdot)$ *be defined by* (14). *Let* $\mathscr{M}'$ *be the set of spectral densities given by* (17). *Then, the set* $\mathscr{M}$ *is contained in the set* $\mathscr{M}$; *that is,* $\mathscr{M} \subset \mathscr{M}'$.

Adopting the noise channel described in the beginning of this section, considering the nonlinearity $g(\cdot)$ in (13), and a subsequent linear predictive filter described by the coefficients $\{a_j, j \leqslant -1\}$, let us use the transformation $\sum_{j \leqslant -1} a_j g(Y_j)$ to map $G(X_0)$, where $G(\cdot)$ is given by (14). Let the set $\{a_j, j \leqslant -1\}$ be in $\mathscr{S}$, where $\mathscr{S}$ is descriped by conditions (B). For some μ in $\mathscr{M}$, it is easy to see that the mean square error (corresponding to $\rho(x, y) \triangleq (x - y)^2$) then takes the form

$$E_\mu \left\{ G(X_0) - \sum_{j \leqslant -1} a_j G(X_j) \right\}^2. \tag{18}$$

Considering now the monotone function $G(\cdot)$ fixed and given by (14), considering the larger convex and closed set $\mathscr{M}'$ in (17) rather than the original set $\mathscr{M}$, and adopting the set $\mathscr{S}$ of information channels in (B), we have defined a new concave—convex problem which is determined by the payoff function in (18) and which satisfies all the conditions in Theorem A. Specifically,

$$\begin{aligned} &\sup_{\mu G^{-1} \in \mathscr{M}'} \inf_{\sigma \in \mathscr{S}} E_\mu \left\{ G(X_0) - \sum_{j \leqslant -1} a_j G(X_j) \right\}^2 \\ &= \inf_{\sigma \in \mathscr{S}} \sup_{\mu G^{-1} \in \mathscr{M}'} E_\mu \left\{ G(X_0) - \sum_{j \leqslant -1} a_j G(X_j) \right\}^2. \end{aligned} \tag{19}$$

We will first fix the measure μ and we will find the optimal linear filter at μ; that is, the optimal set $\{a_j\}$. Denoting by $H_\mu(\lambda)$ the transfer function of the filter at μ, we have the well-known optimal result (Hannan, 1970) for predictive filtering,

$$\|1 - H_\mu(\lambda)\|^2 = 2\pi P_{G\mu}^{-1}(\lambda) \cdot \exp\left\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log P_{G\mu}(\lambda)\, d\lambda\right\}$$

$$\text{for almost all } \lambda \text{ in } [-\pi, \pi], \tag{20}$$

$$e(\mu, G) = (2\pi) \exp\left\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log P_{G\mu}(\lambda)\, d\lambda\right\}. \tag{21}$$

In (20) and (21) $P_{G\mu}(\lambda)$ is the spectral density induced by the measure μ and the nonlinearity $G(\cdot)$, and $e(\mu, G)$ is the value of the payoff function in (18) for the linear predictive filter in (20). Due to expressions (20) and (21), the saddle-point solution on $\mathscr{M}' \times \mathscr{S}$, for the payoff function in (18) will be specified by the measure μG^{-1} which satisfies the supremum of $e(\mu, G)$ in (21) on $\mathscr{M}'$ and the corresponding filter in (20). We express the above conclusions in

PROPOSITION 2. *Let $G(\cdot)$ be the nonlinearity in* (14). *Let the sets $\mathscr{M}'$, $\mathscr{S}$ be as in* (17) *and* (B), *respectively. Then, the payoff function $E_\mu\{G(X_0) - \sum_{j \leqslant -1} a_j G(X_j)\}^2$ has a saddle-point solution on $\mathscr{M}' \times \mathscr{S}$. This solution is satisfied by the pair (μ^*, H^*) such that*

$$\int_{-\pi}^{\pi} \log P_{G\mu^*}(\lambda)\, d\lambda = \sup_{\mu G^{-1} \in \mathscr{M}'} \int_{-\pi}^{\pi} \log P_{G\mu}(\lambda)\, d\lambda,$$

$$\|1 - H^*(\lambda)\|^2 = 2\pi P_{G\mu^*}^{-1}(\lambda) \cdot \exp\left\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log P_{G\mu^*}(\lambda)\, d\lambda\right\}$$

$$\text{a.e. } \lambda \in [-\pi, \pi].$$

The pair (μ^*, H^*) is equivalently the saddle-point solution on $\mathscr{M}' \times \mathscr{S}$ for the payoff function $E_{\mu,\nu}\{X_0 - G^{-1}(\sum_{j \leqslant -1} a_j g(Y_j))\}^2$, where ν the additive Gaussian noise channel of this section and $g(\cdot)$ the nonlinearity in (13).

As it is clear from Proposition 2, the problem has now been reduced to the investigation of the spectrum $P_{G\mu}(\lambda)$ in

$$\mathscr{M}': (2\pi)^{-1} \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - P_{G\mu}^{1/2}(\lambda)]^2\, d\lambda \leqslant c\alpha$$

that satisfies the supremum of $\int_{-\pi}^{\pi} \log P_{G\mu}(\lambda)\, d\lambda$. We thus express Lemma 5 whose proof is in Appendix B.

LEMMA 5. *The pair* (μ^*, H^*) *in Proposition* 2 *is*

$$\mu^*: P_{G\mu^*}(\lambda) = \gamma^* + 2^{-1}P_{G\mu_0}(\lambda)[1 + 4\gamma^* P_{G\mu_0}^{-1}(\lambda)]^{1/2} + 2^{-1}P_{G\mu_0}(\lambda)$$

$$H^*: \|1 - H^*(\lambda)\|^2 = (2\pi) P_{G\mu^*}^{-1}(\lambda) \cdot \exp\left\{(2\pi)^{-1}\int_{-\pi}^{\pi} \log P_{G\mu^*}(\lambda)\, d\lambda\right\}$$

$$a.e.\ \lambda \in [-\pi, \pi],$$

$$: H(0) = 1,$$

where γ^* *is the unique positive constant which satisfies the expression*

$$(2\pi)^{-1}\int_{-\pi}^{\pi} \{\gamma + 2^{-1}P_{G\mu_0}(\lambda) - 2^{-1}P_{G\mu_0}(\lambda)[1 + 4\gamma P_{G\mu_0}^{-1}(\lambda)]^{1/2}\}\, d\lambda \overset{\gamma<\gamma^*}{\underset{\gamma>\gamma^*}{\lesseqgtr}} c\alpha.$$

The only restriction $P_{G\mu_0}(\lambda)$ for the existence of the solution in Lemma 6 is that $P_{G\mu_0}(\lambda) \neq 0$, $\forall \lambda$. The result in Lemma 5 was obtained without any energy restrictions on the spectral densities $P_{G\mu}(\lambda)$. Indeed, even if the spectral densities of the original measures μ have fixed common energy, this does not hold for the transformed measures μG^{-1}.

We will conclude by observing that we can always define $B_n = 1$ and $B_1 = 0$, in (13). Then the constant c in (15) is equal to $(2\pi)^{-1/2}$, and the sphere $\mathscr{M}'$ of spectral densities $P_{G\mu}(\lambda)$ in (17) has smaller radius than the sphere $\mathscr{M}$ in (11). In general therefore, the solution in Lemma 5 induces then smaller mismatch errors. This introduces an additional advantage presented by the nonlinearity $G(\cdot)$. The other advantage is, as already stated, asymptotic continuity, thus the existence of saddle-point–saddle-value solution.

APPENDIX A

Proof of Theorem 1. Let us define

$$K(\mu) \triangleq \inf_{\sigma \in \mathscr{S}} K(\mu, \sigma), \qquad \mu \in \mathscr{M},$$

$$K(\mu_\sigma, \sigma) \triangleq \sup_{\mu \in \mathscr{M}} K(\mu, \sigma), \qquad \sigma \in \mathscr{S}.$$

Due to the convexity of the set $\mathscr{S}$, its closeness with respect to the metric $\Pi_{\gamma,1}$, and due to the continuity of $K(\mu, \sigma)$ with respect to $\Pi_{\gamma,1}$ and its strict convexity in σ on $\mathscr{S}$, the function $K(\mu)$ above exists and it is unique. Due to the convexity of $\mathscr{M}$, its closeness and compactness with respect to the metric $\bar{\gamma}$, and due to the continuity of $K(\mu, \sigma)$ with respect to the metric $\bar{\gamma}$ and its concavity in μ on $\mathscr{M}$, the function $K(\mu_\sigma, \sigma)$ above exists and it is unique. Let now $\mu_1, \mu_2 \in \mathscr{M}$, and λ: $0 < \lambda < 1$. Then,

$$\begin{aligned} K(\lambda\mu_1 + (1-\lambda)\mu_2) &\triangleq \inf_{\sigma\in\mathscr{S}} K(\lambda\mu_1 + (1-\lambda)\mu_2, \sigma) \\ &\geqslant \inf_{\sigma\in\mathscr{S}} \{\lambda K(\mu_1, \sigma) + (1-\lambda) K(\mu_2, \sigma)\} \\ &\geqslant \lambda \inf_{\sigma\in\mathscr{S}} K(\mu_1 \sigma) + (1-\lambda) \inf_{\sigma\in\mathscr{S}} K(\mu_2 \sigma) \\ &= \lambda K(\mu_1) + (1-\lambda) K(\mu_2). \end{aligned}$$

Thus $K(\mu)$ is concave on $\mathscr{M}$. It is also continuous in μ on $\mathscr{M}$ with respect to the metric $\bar{\gamma}$, and $\mathscr{M}$ is closed and compact with respect to $\bar{\gamma}$. Thus, the following supremum exists and it is unique.

$$K(\mu^*) = \sup_{\mu\in\mathscr{M}} K(\mu) = \sup_{\mu\in\mathscr{M}} \inf_{\sigma\in\mathscr{S}} K(\mu, \sigma).$$

We now obtain

$$\begin{aligned} \inf_{\sigma\in\mathscr{S}} \sup_{\mu\in\mathscr{M}} K(\mu, \sigma) &= \inf_{\sigma\in\mathscr{S}} K(\mu_\sigma, \sigma) \leqslant \inf_{\sigma^0\in\mathscr{S}} K(\mu_\sigma, \sigma^0) = K(\mu_\sigma) \\ &\leqslant K(\mu^*) \triangleq \sup_{\mu\in\mathscr{M}} \inf_{\sigma\in\mathscr{S}} K(\mu, \sigma). \end{aligned}$$

But it is always true that

$$K(\mu^*) \triangleq \sup_{\mu\in\mathscr{M}} \inf_{\sigma\in\mathscr{S}} K(\mu, \sigma) \leqslant \inf_{\sigma\in\mathscr{S}} \sup_{\mu\in\mathscr{M}} K(\mu, \sigma).$$

From the last two inequalities, we thus obtain

$$\inf_{\sigma\in\mathscr{S}} \sup_{\mu\in\mathscr{M}} K(\mu, \sigma) = \sup_{\mu\in\mathscr{M}} \inf_{\sigma\in\mathscr{S}} K(\mu, \sigma) = K(\mu^*, \sigma^*),$$

where

$$\mu^*: K(\mu^*) = \sup_{\mu\in\mathscr{M}} K(\mu)$$

$$\sigma^*: K(\mu^*, \sigma^*) = \inf_{\sigma\in\mathscr{S}} K(\mu^*, \sigma).$$

Proof of Theorem 2. (i) Directly from Theorem 1 in (Papantoni-Kazakos, 1981) we have that if the conditions stated in the present theorem are satisfied, then: For any stationary μ, and for given $\varepsilon > 0$, we have: There exists $\delta(\varepsilon, \mu) > 0$ such that

$$\mu' \text{ stationary and } \bar{\rho}(\mu, \mu') < \delta(\varepsilon, \mu) \rightarrow \bar{\tau}(\mu v_m^{-1}, \mu' v_m^{-1}) < \varepsilon.$$

Similarly, for any given $\mu\nu_m^{-1}$ stationary, and given $\varepsilon > 0$, we have:
There exists $\delta(\varepsilon, \mu\nu_m^{-1}) > 0$ such that

$$\mu'\nu_m^{-1} \text{ stationary and } \bar{\tau}(\mu\nu_m^{-1}, \mu'\nu_m^{-1}) < \delta(\varepsilon, \mu\nu_m^{-1}) \to \bar{\rho}(\mu\nu_m^{-1}\sigma_l^{-1}, \mu'\nu_m^{-1}\sigma_l^{-1}) < \varepsilon.$$

Therefore, for any given μ and given $\varepsilon_0 > 0$, there exists some $\delta = \delta(\varepsilon_0, \mu, \sigma_l, \nu_m) > 0$ such that

$$\mu' \text{ stationary and } \bar{\rho}(\mu, \mu') < \delta \to \bar{\rho}(\mu\nu_m^{-1}\sigma_l^{-1}, \mu'\nu_m^{-1}\sigma_l^{-1}) < \varepsilon_0. \tag{A.1}$$

Using now the metric $\gamma(\cdot, \cdot)$ we have

$$\begin{aligned}\bar{\gamma}(\mu, \mu\nu_m^{-1}\sigma_l^{-1}) &\leqslant \bar{\gamma}(\mu, \mu') + \bar{\gamma}(\mu', \mu\nu_m^{-1}\sigma_l^{-1}) \leqslant \bar{\gamma}(\mu, \mu') \\ &\quad + \bar{\gamma}(\mu', \mu'\nu_m^{-1}\sigma_l^{-1}) + \bar{\gamma}(\mu'\nu_m^{-1}\sigma_l^{-1}, \mu\nu_m^{-1}\sigma_l^{-1}) \\ &\to \bar{\gamma}(\mu, \mu\nu_m^{-1}\sigma_l^{-1}) - \bar{\gamma}(\mu', \mu'\nu_m^{-1}\sigma_l^{-1}) \\ &\leqslant \bar{\gamma}(\mu, \mu') + \bar{\gamma}(\mu'\nu_m^{-1}\sigma_l^{-1}, \mu\nu_m^{-1}\sigma_l^{-1}).\end{aligned} \tag{A.2}$$

Due to the symmetry of expression (A.2), we obtain

$$|\bar{\gamma}(\mu, \mu\nu_m^{-1}\sigma_l^{-1}) - \bar{\gamma}(\mu', \mu'\nu_m^{-1}\sigma_l^{-1})| \leqslant \bar{\gamma}(\mu, \mu') + \bar{\gamma}(\mu'\nu_m^{-1}\sigma_l^{-1}, \mu\nu_m^{-1}\sigma_l^{-1}). \tag{A.3}$$

Utilizing now (A.1), (A.3), and the conditions in (9), we have: Given $\varepsilon_1 > 0$, there exists $\delta_1 > 0$ such that

$$\begin{aligned}|\bar{\gamma}(\mu, \mu\nu_m^{-1}\sigma_l^{-1}) - \bar{\gamma}(\mu', \mu'\nu_m^{-1}\sigma_l^{-1})| < \delta_1 \to |\bar{\rho}(\mu, \mu\nu_m^{-1}\sigma_l^{-1}) \\ - \bar{\rho}(\mu', \mu'\nu_m^{-1}\sigma_l^{-1})| < \varepsilon_1.\end{aligned} \tag{A.4}$$

Given $\varepsilon_2 > 0$, there exists $\delta_2 > 0$ such that

$$\begin{aligned}\bar{\rho}(\mu, \mu') + \bar{\rho}(\mu'\nu_m^{-1}\sigma_l^{-1}, \mu\nu_m^{-1}\sigma_l^{-1}) < \delta_2 \to \bar{\gamma}(\mu, \mu') \\ + \bar{\gamma}(\mu'\nu_m^{-1}\sigma_l^{-1}, \mu\nu_m^{-1}\sigma_l^{-1}) < \varepsilon_2.\end{aligned} \tag{A.5}$$

So, given $\varepsilon > 0$, find $\delta_1(\varepsilon)$ in (A.4) and select $\varepsilon_2 = \delta_1(\varepsilon)$, where ε_2 given in (A.5). Then find $\delta_2 = \delta_2(\varepsilon_2) = \delta_2(\delta_1(\varepsilon))$, where δ_2 is given in (A.5). Then select the ε_0 in (A.1) such that $\varepsilon_0 = \delta_2(\delta_1(\varepsilon))/2$. Denote $\delta(\varepsilon_0)$ the δ in (A.1). Select

$$\delta' = \min\left(\varepsilon_0 = \frac{\delta_2(\delta_1(\varepsilon))}{2},\ \delta(\varepsilon_0) = \delta\left(\frac{\delta_2(\delta_1(\varepsilon))}{2}\right)\right).$$

Then $\bar{\rho}(\mu, \mu') < \delta' \to |\bar{\rho}(\mu, \mu\nu_m^{-1}\sigma_l^{-1}) - \bar{\rho}(\mu', \mu'\nu_m^{-1}\sigma_l^{-1})| < \varepsilon$ and continuity with respect to the measure μ has been proved.

(ii) Let μ be fixed. Let $\sigma_{0l}, \sigma_l \in Q_l$. For given x^l, let $p_{x^l}^1$ denote the joint measure with marginals $\sigma^1_{0l,x^l}, \sigma^1_{l,x^l}$, which satisfies the Prohorov distance $\Pi_{\gamma,1}(\sigma^1_{0l,x^l}, \sigma^1_{l,x^l})$.

Let w, z denote outcomes generated by the stationary processes $\mu v_m^{-1}\sigma_{0l}^{-1}$ and $\mu v_m^{-1}\sigma_l^{-1}$, respectively. Denote by p' some joint measure with marginals the 1-dimensional restrictions of the measures $\mu v_m^{-1}\sigma_{0l}^{-1}$, $\mu v_m^{-1}\sigma_l^{-1}$. Let E^l be such that $E^l \in R^l$ and $\mu v_m^{-1}(E^l) > 1-\delta_1$, for some $0<\delta_1<1$. Let

$$x^l \in E^l \to \Pi_{\gamma,1}(\sigma^1_{0l,x^l}, \sigma^1_{l,x^l}) < \delta_0.$$

Then

$$p^1(w,z\colon \gamma(w,z) \geqslant \delta_0) = \int_{R^l} \mu v_m^{-1}(x^l)\, p^1_{x^l}(w,z\colon \gamma(w,z)\geqslant \delta_0) \leqslant \delta_0+\delta_1. \quad (A.6)$$

Given $\varepsilon_1>0$, select $\delta_1=\varepsilon_1/2$ and $\delta_0=\varepsilon_1/2$. Then, from (A.6) we conclude

$$\Pi_{\gamma,1}(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \varepsilon_1, \quad (A.7)$$

where in (A.7) we express the Prohorov distance between the 1-dimensional restrictions of the measures $\mu v_m^{-1}\sigma_{0l}^{-1}$ and $\mu v_m^{-1}\sigma_l^{-1}$.

Due to the properties of $\gamma(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$ in the theorem, given $\varepsilon_2>0$, there exists $\delta>0$ such that

$$\Pi_{\gamma,1}(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \delta \to \bar\rho(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \varepsilon_2. \quad (A.8)$$

Thus, select $\varepsilon_1=\delta$ and $\delta_0=\delta/2$ and conclude

$$\Pi_{\gamma,1}(\sigma^1_{0l,x^l}, \sigma^1_{l,x^l}) < \frac{\delta}{2} \qquad \forall x^l \in E^l \to \bar\rho(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \varepsilon_2. \quad (A.9)$$

Now, due to

$$\begin{aligned}\bar\gamma(\mu, \mu v_m^{-1}\sigma_{0l}^{-1}) &\leqslant \bar\gamma(\mu, \mu v_m^{-1}\sigma_l^{-1}) + \bar\gamma(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1})\\ &\to |\bar\gamma(\mu, \mu v_m^{-1}\sigma_{0l}^{-1}) - \bar\gamma(\mu, \mu v_m^{-1}\sigma_l^{-1})|\\ &\leqslant \bar\gamma(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1})\end{aligned}$$

as well as (A.4) and condition (similar to (A.5)): Given $\varepsilon_3>0$, there exists $\delta_3>0$ such that

$$\bar\rho(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \delta_3 \to \bar\gamma(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \varepsilon_3,$$

we conclude: Given $\varepsilon>0$, there exists $\delta(\varepsilon)>0$ such that

$$\bar\rho(\mu v_m^{-1}\sigma_{0l}^{-1}, \mu v_m^{-1}\sigma_l^{-1}) < \delta(\varepsilon) \to |\bar\rho(\mu, \mu v_m^{-1}\sigma_{0l}^{-1}) - \bar\rho(\mu, \mu v_m^{-1}\sigma_l^{-1})| < \varepsilon. \quad (A.10)$$

Select in (A.8) $\varepsilon_2 = \delta(\varepsilon)$ and thus $\delta = (\delta(\varepsilon))$, $\varepsilon_1 = \delta$, $\delta_0 = \delta/2$, and continuity with respect to the channel σ_l has been proven.

Proof of Theorem 3. Due to the compactness of the set $\mathscr{M}$, given $\eta > 0$, a finite coverage of the set $\mathscr{M}$ can be found consisting of $N(\eta/2)$ spheres $\bar{\gamma}(\mu_i v^{-1}, \mu v^{-1}) < \eta/2$, each centered at the measure $\mu_i v^{-1}$, $1 \leqslant i \leqslant N(\eta/2)$.

Given l, if $E_i^l \in R^l$ and is such that $\mu_i^l(E_i^l) > 1 - (\eta/2)$, then $\mu^l(E_i^l) > 1 - \eta$; $\forall \mu$: $\bar{\gamma}(\mu_i v^{-1}, \mu v^{-1}) < \eta/2$. Then, given η, l, select $E^l = \bigcup_{1 \leqslant i \leqslant N(\eta/2)} E_i^l$ for the asymptotically continuous sequence $\{\sigma_l\}$. Clearly $\{\sigma_l\}$ is then also continuous at every μv^{-1}, $\mu \in \mathscr{M}$.

Proof of Lemma 2. (i) Given fixed $\mu \in \mathscr{M}$, the distance $\Pi_{\rho,1}(\mu, \mu v_m^{-1} \sigma^{-1})$ is strictly convex with respect to σ because it is strictly convex with respect to $\mu v_m^{-1} \sigma^{-1}$.

(ii) Given fixed $\sigma \in \mathscr{S}$, let $\mu_1, \mu_2 \in \mathscr{M}$. Form the measure $\mu = \varepsilon\mu_1 + (1-\varepsilon)\mu_2$, for any given ε: $0 < \varepsilon < 1$. Then, $\mu \in \mathscr{M}$.

The measure μ induces the matching process measure $\varepsilon\mu_1 v_m^{-1}\sigma^{-1} + (1-\varepsilon)\mu_2 v_m^{-1}\sigma^{-1}$. Let p^1 be the joint measure which has for marginals the 1-dimensional restrictions of the measures $\varepsilon\mu_1 + (1-\varepsilon)\mu_2$ and $\varepsilon\mu_1 v_m^{-1}\sigma^{-1} + (1-\varepsilon)\mu_2 v_m^{-1}\sigma^{-1}$, and which also satisfies the Prohorov distance $\Pi_{\rho,1}(\varepsilon\mu_1 + (1-\varepsilon)\mu_2,\ \varepsilon\mu_1 v_m^{-1}\sigma^{-1} + (1-\varepsilon)\mu_2 v_m^{-1}\sigma^{-1})$.

Let w, z be realizations from the processes $\varepsilon\mu_1 + (1-\varepsilon)\mu_2$ and $\varepsilon\mu_1 v_m^{-1}\sigma^{-1} + (1-\varepsilon)\mu_2 v_m^{-1}\sigma^{-1}$, respectively. Let $p_{\mu_1}^1$, $p_{\mu_2}^1$ be the joint measures satisfying, respectively, the Prohorov distances $\Pi_{\rho,1}(\mu_1, \mu_1 v_m^{-1}\sigma^{-1})$ and $\Pi_{\rho,1}(\mu_2, \mu_2 v_m^{-1}\sigma^{-1})$. Then, for any $\delta > 0$, we have

$$p^1(w, z: \rho(w, z) \geqslant \delta) = \varepsilon p_{\mu_1}^1(w, z: \rho(w, z) \geqslant \delta) + (1-\varepsilon)\, p_{\mu_2}^1(w, z: \rho(w, z) \geqslant \delta). \tag{A.11}$$

But

$$\Pi_{\rho,1}(\mu_1, \mu_1 v_m^{-1}\sigma^{-1}) = \inf\{\delta: p_{\mu_1}^1(w, z: \rho(w, z) \geqslant \delta) < \delta\} \tag{A.12}$$

and similarly for $\Pi_{\rho,1}(\mu_2, \mu_2 v_m^{-1}\sigma^{-1})$.

Let

$$\begin{aligned} \Pi_{\rho,1}(\mu_1, \mu_1 v_m^{-1}\sigma^{-1}) &= \delta_1 \\ \Pi_{\rho,1}(\mu_2, \mu_2 v_m^{-1}\sigma^{-1}) &= \delta_2. \end{aligned} \tag{A.13}$$

Then, it is clear that in (A.11) the infinum δ such that $p^1(w, z: \rho(w, z) \geqslant \delta) < \delta$, cannot be less than $\varepsilon\delta_1 + (1-\varepsilon)\delta_2$. Therefore

$$\begin{aligned} &\Pi_{\rho,1}(\varepsilon\mu_1 + (1-\varepsilon)\mu_2,\ \varepsilon\mu_1 v_m^{-1}\sigma^{-1} + (1-\varepsilon)\mu_2 v_m^{-1}\sigma^{-1}) \\ &\qquad \geqslant \varepsilon\Pi_{\rho,1}(\mu_1, \mu_1 v_m^{-1}\sigma^{-1}) + (1-\varepsilon)\,\Pi_{\rho,1}(\mu_2, \mu_2 v_m^{-1}\sigma^{-1}). \end{aligned}$$

(iii) Equivalence between $\Pi_{\rho,1}(\mu, \mu v_m^{-1}\sigma^{-1})$ and $\bar{\rho}(\mu, \mu v_m^{-1}\sigma^{-1})$ is due to stationarity of both measures μ and $\mu v_m^{-1}\sigma^{-1}$ in conjuction with $\rho(\cdot, \cdot)$ boundness.

Appendix B

Proof of Lemma 3. The nonlinearity $g(\cdot)$ clearly maps probability masses per datum. Thus, we only have to show that given $\varepsilon > 0$, there exists some l such that for any finite k and for any two disjoint selections $\{a_{i_j}; 1 \leqslant j \leqslant k\}$, $\{a_{m_j}; 1 \leqslant j \leqslant k\}$ of k filter coefficients, we have: $|\sum_{j=1}^k a_{i_j} - \sum_{j=1}^k a_{m_j}| < \varepsilon$. But

$$\left\{\sum_{j=1}^k a_{i_j} - \sum_{j=1}^k a_{m_j}\right\}^2 \leqslant k \sum_{j=1}^k (a_{i_j} - a_{m_j})^2$$

$$\leqslant k \left\{\sum_{j=1}^k [\max(a_{i_j}, a_{m_j})]^2 - \sum_{j=1}^k [\min(a_{i_j}, a_{m_j})]^2\right\}$$

$$< k \sum_{j=1}^k [\max(a_{i_j}, a_{m_j})]^2.$$

But, given $\varepsilon/k > 0$, there exists l_0: $\sum_{j=-l}^{-1} a_j^2 < \varepsilon/k$; $\forall l > l_0$. So, then $\{\sum_{j=1}^k a_{i_j} - \sum_{j=1}^k a_{m_j}\}^2 < \varepsilon$, and the lemma is proved.

Proof of Lemma 4. Let $\mu \in \mathscr{M}$, where $\mathscr{M}$ as in (11). Let $\rho(x, y) \triangleq (x-y)^2$. Let p be the joint measure that satisfies the rho-bar distance $\bar{\rho}(\mu_0, \mu)$. Let $G(\cdot)$ be as in (14). Then,

$$\bar{\rho}(\mu_0 G^{-1}, \mu G^{-1}) \leqslant \int [G(X) - G(Y)]^2 \, dp(X, Y)$$

$$\leqslant c \int [X - Y]^2 \, dp(X, Y) = c\bar{\rho}(\mu_0, \mu), \qquad \text{(B.1)}$$

where the second inequality in (B.1) is due to (16). From (B.1) we have

$$\mu \in \mathscr{M} \to \bar{\rho}(\mu_0 G^{-1}, \mu G^{-1}) \leqslant c\alpha. \qquad \text{(B.2)}$$

Defining the set $\mathscr{M}''$ as

$$\mathscr{M}'': \bar{\rho}(\mu_0 G^{-1}, \mu G^{-1}) \leqslant c\alpha, \qquad \text{(B.3)}$$

we have, due to (B.2), $\mathcal{M} \subset \mathcal{M}''$. But for $\rho(x, y) \triangleq (x-y)^2$, we have directly from (Gray, Neuhoff, & Shields, 1975),

$$\bar{\rho}(\mu_0 G^{-1}, \mu G^{-1}) \geqslant (2\pi)^{-1} \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - P_{G\mu}^{1/2}(\lambda)]^2 \, d\lambda + [m_{G\mu_0} - m_{G\mu}]^2, \tag{B.4}$$

where $m_{G\mu} = E_\mu G(X)$. Therefore,

$$\bar{\rho}(\mu_0 G^{-1}, \mu G^{-1}) \leqslant c\alpha \to (2\pi)^{-1} \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - P_{G\mu}^{1/2}(\lambda)]^2 \, d\lambda \leqslant c\alpha. \tag{B.5}$$

Defining

$$\mathcal{M}' : 2(\pi)^{-1} \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - P_{G\mu}^{1/2}(\lambda)]^2 \, d\lambda \leqslant c\alpha. \tag{B.6}$$

We have, due to (B.5), $\mathcal{M}'' \subset \mathcal{M}'$. Since also $\mathcal{M} \subset \mathcal{M}''$, we finally have $\mathcal{M} \subset \mathcal{M}'$.

Proof of Lemma 5. Fixing temporarily the constraint to $(2\pi)^{-1} \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - P_{G\mu}^{1/2}(\lambda)]^2 \, d\lambda = \beta$, we formalize the variational problem

$$f(\varepsilon) = \int_{-\pi}^{\pi} \log[P_{G\mu}^*(\lambda) + \varepsilon P_1(\lambda)] \, d\lambda$$
$$- \nu \int_{-\pi}^{\pi} [P_{G\mu_0}^{1/2}(\lambda) - \{P_{G\mu}^*(\lambda) + \varepsilon P_1(\lambda)\}^{1/2}]^2 \, d\lambda, \qquad \nu > 0,$$

where ν is the Langrange multiplier.

If $P_{G\mu}^*(\lambda)$ is the solution of the concave problem under the fixed constraint, then $f'(0) = 0$ for all $P_1(\lambda)$. But

$$f'(0) = \int_{-\pi}^{\pi} P_1(\lambda)\{[P_{G\mu}^*(\lambda)]^{-1} + \nu P_{G\mu_0}^{1/2}(\lambda)[P_{G\mu}^*(\lambda)]^{-1/2} - \nu\} \, d\lambda = 0.$$

So,

$$[P_{G\mu}^*(\lambda)]^{-1} + \nu P_{G\mu_0}^{1/2}(\lambda)[P_{G\mu}^*(\lambda)]^{-1/2} - \nu = 0$$
$$\to [P_{G\mu}^*(\lambda)]^{-1/2} = \tfrac{1}{2}\{-\nu P_{G\mu_0}^{1/2}(\lambda) + [\nu P_{G\mu_0}(\lambda) + 4\nu]^{1/2}\}, \qquad \nu > 0. \tag{B.7}$$

From (B.7) we obtain in a straight forward manner,

$$P_{G\mu}^*(\lambda) = \gamma + 2^{-1} P_{G\mu_0}(\lambda) + 2^{-1} P_{G\mu_0}(\lambda)[1 + 4\gamma P_{G\mu_0}^{-1}(\lambda)]^{1/2}, \qquad \gamma > 0, \tag{B.8}$$

where γ some constant which will be determined through the appropriate constraint. From expression (B.8) we easily find

$$(2\pi)^{-1} \int_{-\pi}^{\pi} \{[P^*_{G\mu}(\lambda)]^{1/2} - P^{1/2}_{G\mu_0}(\lambda)\}^2 \, d\lambda$$

$$= \gamma + (2\pi)^{-1} \int_{-\pi}^{\pi} \{2^{-1} P_{G\mu_0}(\lambda) - 2^{-1} P_{G\mu_0}(\lambda)[1 + 4\gamma P^{-1}_{G\mu_0}(\lambda)]^{1/2}\} \, d\lambda. \tag{B.9}$$

It is easily seen that both the expressions in (B.8) and (B.9) are monotonically increasing with increasing positive parameter γ. Thus, the conclusion in the lemma.

RECEIVED: August 23, 1983; ACCEPTED: July 10, 1984

REFERENCES

GRAY, R. M., NEUHOFF, D., AND SHIELDS P. C. (1975), A generalization of Ornstein's distance with applications to information theory, *Ann. Probab.* **3** No. 2, 315–324.

HANNAN, E. J. (1970), "Multiple Time Series," Wiley, New York.

PAPANTONI-KAZAKOS, P., AND GRAY, R. M. (1979), "Robustness of estimation on stationary observations, *Ann. Probab.* **7** No. 6, 989–1002.

PAPANTONI-KAZAKOS, P. (1981), Sliding block encoders that are rho-bar continuous functions of their input, *IEEE Trans. Inform. Theory* **IT-27** No. 3, 372–376.

ROYDEN, H. L. (1963), "Real Analysis," Macmillan Co., New York.

WIENER, N. (1949), "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," MIT Press, Cambridge, Mass.