Using the least-squares method, we obtain from (14)

$$\operatorname{vec} \hat{B}_i = -\left[\left(\sum_{t=1}^{T_i} Y_i(t)Y_i'(t)\right)^{-1} \otimes I_p\right]\left[\sum_{t=1}^{T_i}\left(Y_i(t)\otimes I_p\right)X_i(t)\right],\quad i = 1,2,\cdots,k. \tag{15}$$

Having an estimator for vec B_i, we may estimate $e_i(t)$ from (14) by

$$\hat{e}_i(t) = X_i(t) + \left(Y_i'(t)\otimes I_p\right)\operatorname{vec}\hat{B}_i$$

and thus matrix Σ_i by

$$\hat{\Sigma}_i = \frac{1}{T_i}\sum_{t=1}^{T_i}\hat{e}_i(t)\hat{e}_i'(t), \qquad i = 1,2,\cdots,k. \tag{16}$$

In practice, not only are the parameters of the model (1) unknown but so is the order of the model. We will now consider the estimation of the true order q_i of the autoregressive equation (1) on the basis of a realization $X_i(1), X_i(2),\cdots, X_i(T_i)$ of the process $\{X_i(t)\}$, $i = 1,2,\cdots,k$. Quinn [9] proposed an expression $\varphi(q_i)$ for the determination of the order of multivariate autoregressive models, of the following form:

$$\varphi(q_i) = \ln|\hat{\Sigma}_i| + \frac{1}{T_i}2q_i c\ln\ln T_i, \qquad c > 1, \tag{17}$$

where $\hat{\Sigma}_i$ is given by (13) or (16). Quinn shows that the estimator $\hat{q}_i^*$ of the true order of the autoregression q_i^* belonging to class π_i, which minimizes expression (17) over all $q_i \in \{1,2,\cdots,Q_i\}$ (where Q_i is an arbitrarily chosen number larger that q_i^*) is a strongly consistent estimator of the true order q_i^*, $i = 1,2,\cdots,k$. Quinn presented arguments for the acceptance of $c = p^2$.

We now propose another method of estimating the true order of the autoregressive equation. We make use of the Bayesian method for the estimation of the true order q_i^*, $i = 1,2,\cdots,k$. Suppose that the order q_i of the autoregression equation is a random variable with a known prior density function $h(q_i)$, $i = 1,2,\cdots,k$. If we have no information as to the choice of $h(q_i)$, we assume $h(q_i) = Q_i^{-1}$, where Q_i is an arbitrarily chosen number larger than q_i^*, $i = 1,2,\cdots,k$. An incorrect decision in choosing the order of model (1) results in a loss of the form

$$S(q_i, r_i) = c|q_i - r_i|, \tag{18}$$

where c is an a constant chosen beforehand, which satisfies the following conditions: $c = 1$ for $r_i \leqslant q_i$ and $c > 1$ for $r_i > q_i$, where r_i is the assumed order the model (1). The loss function of the form (18) has the following properties. The cost of choosing an order larger than the true one is proportional to the error. The cost of choosing an order smaller than the true one is smaller than cost of a higher order because lowering the order of the model progressively lowers the calculating expense involved in the data analysis.

The estimator $\hat{q}_i^*$ of the true order of the model (1) is obtained by minimizing the posterior risk

$$R(r_i) = \sum_{q_i=1}^{Q_i} S(q_i, r_i)z(q_i|x(1),\cdots,x(T_i)) \tag{19}$$

over $r_i \in \{1,2,\cdots,Q_i\}$, where $z(q_i|x(1),\cdots,x(T_i))$ is the posterior density function of the random variable q_i

$$z(q|x(1),\cdots,x(T_i)) = \frac{h(q_i)f(x(1),\cdots,x(T_i)|q_i)}{\sum_{q_i=1}^{Q_i} h(q_i)f(x(1),\cdots,x(T_i)|q_i)},$$

$i = 1,2,\cdots,k$. The function $f(x(1),\cdots,x(T_i)|q_i)$ is a density function of the form (8), $i = 1,2,\cdots,k$. From (18) and (19) we obtain that the estimator $\hat{q}_i^*$ of the true order of the model (1) can be expressed as

$$\hat{q}_i^* = \arg\left\{\min_{r_i}\sum_{q_i=1}^{Q_i} c|q_i - r_i|h(q_i)f(x(1),\cdots,x(T)|q_i)\right\}, \tag{20}$$

i.e., $\hat{q}_i^*$ is that values of the parameter r_i for which the risk function $R(r_i)$ takes the minimal value, $i = 1,2,\cdots,k$.

As the true order of the model is not known, we can assume that $q_i \in \{1,2,\cdots,Q_i\}$, $i = 1,2,\cdots,k$. For each assumed order we can obtain the estimators of the remaining parameters of the model by (12) and (13) or (15) and (16). In this manner we can compute the values of the function (8) for each $q_i \in \{1,2,\cdots,Q_i\}$, $i = 1,2,\cdots,k$, and we can find the estimator of the order of the model by minimizing the posterior risk.

References

[1] R. L. Kashyap, "Optimal feature selection and decision rules in classification problems with time series." *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 281–288, May 1978.

[2] H. Läuter and N. Miethe, "Some tests in dynamic models with a finite number of parameters," *Math. Operationsforsch. Statist., Ser. Statistics*, vol. 10, pp. 395–414, 1979.

[3] Y. Baram, "A sufficient condition for consistent discrimination between stationary Gaussian models," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 958–960, Oct. 1978.

[4] D. Kazakos, "On resolution and exponential discrimination between Gaussian stationary vector processes and dynamic model," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 294–296, Apr. 1980.

[5] D. Kazakos and P. Papantoni–Kazakos, "Spectral distance measures between Gaussian processes," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 950–959, Oct. 1980.

[6] E. J. Hannan, *Multiple Time Series*. New York: Wiley, 1970.

[7] R. L. Kashyap and R. E. Nashburg, "Parameter estimation in multivariate stochastic difference equations," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 784–797, Dec. 1974.

[8] P. Whittle, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," *Biometrika*, vol. 50, pp. 129–134, 1963.

[9] B. G. Quinn, "Order determination for a multivariate autoregression," *J. Roy. Statist. Soc., Series B*, vol. 42, pp. 182–185, 1980.

Robust Wiener Filters for Random Signals in Correlated Noise

GEORGE MOUSTAKIDES AND SALEEM A. KASSAM, MEMBER, IEEE

Abstract—**Minimax robust Wiener filtering is considered for the case in which the signal and noise spectral-density matrix is not completely specified. Results are obtained for spectral-density matrix classes which are defined by upper and lower bounds on the components of the matrix. These results form an extension of earlier results on robust Wiener filtering for the case of uncorrelated signals and noise.**

I. Introduction

For linear estimation of a random signal which is observed in additive noise, the minimum mean-squared-error solution can be

Manuscript received May 26, 1982. This research is supported in part by the Air Force Office of Scientific Research under Grant AFOSR 82-0022. This work was partially presented at the 1980 Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 1980.

G. Moustakides is with the Department of Electrical Engineering and Computer Science, Princeton University, Princeton, NJ 08544.

S. A. Kassam is with the Department of Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104.

0018-9448/83/0700-0614$01.00 ©1983 IEEE

obtained through standard procedures if the signal and noise second-order characteristics are given. When complete information on the signal and noise second-order characteristics is not available, one approach is to design minimax robust filters. In [1] and [2] robust Wiener filters and smoothers have been considered for situations in which the signal and noise processes are known to be uncorrelated. For a related problem in [3] robust state estimation for linear stochastic systems is considered, again with uncorrelated process and observation noise. Under the same restriction of uncorrelated signal and noise, a general approach for robust causal estimation in the discrete-time case has been given in [4].

The robust procedures in the above cases are procedures which minimize the worst-case error for signal and noise spectral distributions ranging over specified classes of allowable spectra. Such results are very closely related to those on minimax robust hypothesis testing [5], as discussed explicitly in [2], [6], [7], and many of the available results on robust hypothesis testing for classes of probability distributions can be translated directly into results for corresponding robust estimation problems.

In this correspondence we will consider the situation in which the signal and noise processes are possibly correlated in a Wiener filtering (smoothing) problem. Specific results will be established on robust filters for classes of allowable spectral characteristics described by upper and lower bounds on the spectra and cross-spectrum magnitudes of the signal and noise processes. The presence of correlation introduces an aspect for which no counterpart appears to exist in known results on robust hypothesis testing.

In the next section we develop a few basic results, which will be then applied in Section III to obtain the robust filters for bounded spectral classes. Although all our results are developed for the continuous-time case, exactly the same considerations lead to direct counterparts of these results for the discrete-time case.

II. Problem Statement and General Results

A. Wiener Filter Results

Assume that $S(t)$ and $N(t)$ are jointly wide-sense stationary, second-order, zero-mean processes, and that their spectral density matrix D exists, with

$$D = \begin{bmatrix} D_s(\omega) & D_{sn}(\omega) \\ D_{sn}^*(\omega) & D_n(\omega) \end{bmatrix}, \qquad -\infty < \omega < \infty.$$

Let the received process be

$$X(t) = S(t) + N(t), \qquad -\infty < t < \infty. \tag{1}$$

If $X(t)$ is passed through a filter with frequency response $H(\omega)$, then the mean-squared error $e(D, H)$ between filter output $Y(t)$ and $S(t)$ is given by

$$e(D,H) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \{D_s(\omega) - H(\omega)D_{sx}^*(\omega) - H^*(\omega)D_{sx}(\omega) + |H(\omega)|^2 D_x(\omega)\}\, d\omega. \tag{2}$$

Here $D_x(\omega)$ is the spectral density of the process $X(t)$, and $D_{sx}(\omega)$ is the cross-spectral density of $S(t)$ and $X(t)$. The optimum linear filter characteristic $H_D(\omega)$ minimizing (2) is given by

$$H_D(\omega) = \frac{D_{sx}(\omega)}{D_x(\omega)} \tag{3}$$

or, in terms of the D matrix components,

$$H_D(\omega) = \frac{D_s(\omega) + D_{sn}(\omega)}{D_s(\omega) + D_n(\omega) + 2\,\mathrm{Re}[D_{sn}(\omega)]}. \tag{4}$$

The corresponding error $e(D, H_D)$ is the minimum error $e_{\mathrm{op}}(D)$ for D, and it is given by

$$e_{\mathrm{op}}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D_s(\omega)\cdot D_x(\omega) - |D_{sx}(\omega)|^2}{D_x(\omega)}\, d\omega. \tag{5}$$

In terms of the D matrix components this is

$$e_{\mathrm{op}}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D_s(\omega)\cdot D_n(\omega) - |D_{sn}(\omega)|^2}{D_s(\omega) + D_n(\omega) + 2\,\mathrm{Re}[D_{sn}(\omega)]}\, d\omega. \tag{6}$$

Note that when $D_s(\omega) = D_n(\omega) = -D_{sn}(\omega)$, the integrand of (6) is undefined as it is written, but should in this case be interpreted as $D_s(\omega)$. Also, $H_D(\omega)$ under this condition may be arbitrarily defined, and in particular may be identically zero.

B. Robust Wiener Filters

Suppose now that the matrix-valued function D is not precisely known for all ω. We assume that D is known to belong to some class Δ of spectral density matrix functions. We require our filter frequency response function H to be in the class $\mathcal{H}$ of all bounded functions of ω. This means that the filter may be noncausal. Then a minimax robust filter for our problem is one with a frequency response $H_r \in \mathcal{H}$ satisfying

$$\min_{H\in\mathcal{H}} \sup_{D\in\Delta} e(D,H) = \sup_{D\in\Delta} e(D, H_r). \tag{7}$$

If the supremum on the right-hand side of (7) is achieved by a spectral density matrix $D^r \in \Delta$ for which the minimax filter H_r is an optimum filter, then (D^r, H_r) satisfies the saddle-point condition

$$e(D, H_r) \leqslant e(D^r, H_r) \leqslant e(D^r, H) \tag{8}$$

for all $D \in \Delta$ and all $H \in \mathcal{H}$. In this case we also have

$$e_{\mathrm{op}}(D) \leqslant e_{\mathrm{op}}(D^r) \tag{9}$$

for all $D \in \Delta$. Any $D^r \in \Delta$ satisfying (9) will be called a least-favorable (lf) spectral density matrix in Δ.

When the signal and noise are uncorrelated, it has been shown in [2] that a robust filter H_r satisfying (8) can be obtained as that filter which is optimum for a least-favorable spectral density matrix D^r satisfying (9). Of course, D^r would be diagonal in this case. This considerably simplifies the task of obtaining the robust filter for any given class Δ. In the correlated case, the components $D_s^r(\omega)$, $D_n^r(\omega)$, and $D_{sn}^r(\omega)$ of D^r may be such that $D_s^r(\omega) = D_n^r(\omega) = -D_{sn}^r(\omega) > 0$ on a nonnull ω-set, in which case the optimum filter for D^r is nontrivially not uniquely defined. We can show [8], [9] that when this condition is not encountered the optimum filter for an lf matrix D^r is a minimax robust filter satisfying (8). In the more general case our approach will be to obtain a least favorable D^r first, and then consider as candidates for the robust filter those filters which are optimum for D^r. We will prove robustness of our specific solutions directly by showing (8).

C. Maximization of $e_{\mathrm{op}}(D)$

From (6) we see that for given $D_s(\omega)$, $D_n(\omega)$, and $|D_{sn}(\omega)|$ the worst-case function for $\mathrm{Re}[D_{sn}(\omega)]$ is $-|D_{sn}(\omega)|$, since this minimizes the denominator. This can happen only when $D_{sn}(\omega) = -|D_{sn}(\omega)|$, that is, when $D_{sn}(\omega)$ is real and nonpositive. For this case the value of $e_{\mathrm{op}}(D)$ is

$$e_{\mathrm{op}}(D) = \frac{1}{2\pi}\int_{\infty}^{\infty} \frac{D_s(\omega)\cdot D_n(\omega) - |D_{sn}(\omega)|^2}{D_s(\omega) + D_n(\omega) - 2|D_{sn}(\omega)|}\, d\omega. \tag{10}$$

The integrand in (10) for given $D_s(\omega)$, $D_n(\omega)$ is

a) increasing with respect to $|D_{sn}(\omega)|$ when

$$0 \leqslant |D_{sn}(\omega)| \leqslant \min\{D_s(\omega), D_n(\omega)\};$$

b) decreasing with respect to $|D_{sn}(\omega)|$ when

$$\min\{D_s(\omega), D_n(\omega)\} \leqslant |D_{sn}(\omega)| \leqslant \sqrt{D_s(\omega)\cdot D_n(\omega)}.$$

This can be proved by taking the derivative of the integrand with respect to $|D_{sn}(\omega)|$ and treating $D_s(\omega)$, $D_n(\omega)$ as constants. Obviously for $|D_{sn}(\omega)| = \min\{D_s(\omega), D_n(\omega)\}$ we have a maximum for (10), provided that $|D_{sn}(\omega)|$ can take this value. In general we have to deal with a constrained maximization problem.

When $D_s(\omega)$ and $D_n(\omega)$ are not given, the above has to be considered for each allowable $D_s(\omega)$, $D_n(\omega)$ and the worst-case obtained. In the next section we will apply these ideas to specific classes Δ of spectral density matrices.

III. Robust Filters for Bounded Spectral Density Classes

We now consider specific classes Δ which arise from an imposition of bounds on some or all of the components $D_s(\omega)$, $D_n(\omega)$, and $|D_{sn}(\omega)|$ of $D \in \Delta$. In addition, constraints will be imposed on the variances of the signal and noise processes. Such classes are useful because in many situations total signal and noise powers can be measured, although actual spectra and cross spectra may only be reasonably assumed to lie within some confidence bands. The classes Δ we will consider here are directly related to the band models for spectra first considered in [1]. We start by considering a simple class and progress to other classes by modifying and adding constraints.

A. Signal and Noise Spectra Given, Upper Bound on $|D_{sn}(\omega)|$

Here we have classes Δ of spectral density matrices with specified diagonal elements, that is with known signal and noise spectra $D_s(\omega)$ and $D_n(\omega)$. The cross spectrum, however, is known only to be bounded above in magnitude as

$$0 \leqslant |D_{sn}(\omega)| \leqslant U(\omega), \qquad -\infty < \omega < \infty, \tag{11}$$

where $U(\omega)$ is a given bound. Note that $U(\omega)$ can be assumed to satisfy

$$U(\omega) \leqslant \sqrt{D_s(\omega)D_n(\omega)}, \qquad -\infty < \omega < \infty.$$

The lf matrix D^r now has diagonal elements $D_s(\omega)$ and $D_n(\omega)$. From the observations in Section II, we conclude that the cross spectrum $D_{sn}^r(\omega)$ in the lf matrix satisfies

$$\begin{aligned} |D_{sn}^r(\omega)| &= \min\{\min\{D_s(\omega), D_n(\omega)\}, U(\omega)\} \\ &= \min\{D_s(\omega), D_n(\omega), U(\omega)\} \end{aligned} \tag{12}$$

and, in particular,

$$D_{sn}^r(\omega) = -\min\{D_s(\omega), D_n(\omega), U(\omega)\} \tag{13}$$

for all ω. Now we take H_r to be the frequency response of a filter optimum for D^r. This filter characteristic is defined, from (4), by

$$H_r(\omega) = \frac{D_s(\omega) - \min\{D_s(\omega), D_n(\omega), U(\omega)\}}{D_s(\omega) + D_n(\omega) - 2\min\{D_s(\omega), D_n(\omega), U(\omega)\}} \tag{14}$$

when this does not reduce to 0/0; at such frequencies we can assign it any value between 0 and 1. The proof of the robustness of this H_r is a special case of a more general result we will consider next.

For the special case when $\min\{D_s(\omega), D_n(\omega)\} \leqslant U(\omega)$, all ω, (12) becomes

$$|D_{sn}^r(\omega)| = \min\{D_s(\omega), D_n(\omega)\},$$

all ω, and the frequency response of any robust filter is given by

$$H_r(\omega) = \begin{cases} 1, & D_s(\omega) > D_n(\omega); \\ A(\omega), & D_s(\omega) = D_n(\omega); \\ 0, & D_s(\omega) < D_n(\omega), \end{cases} \tag{15}$$

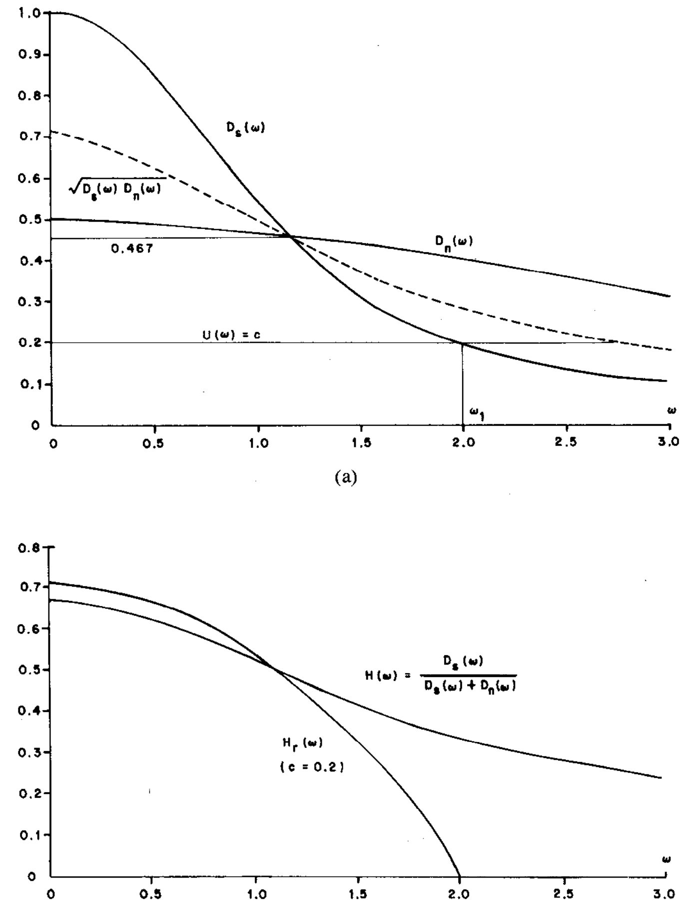

Fig. 1. (a) Illustration of spectra for example in Section III-A. (b) Robust and nominally optimum filter frequency responses for example in Section III-A.

where $A(\omega)$ can take on any value in [0, 1], for example $A(\omega) = 0$ or 1. This means that when there is considerable uncertainty about the actual value of $D_{sn}(\omega)$, for example when nothing beyond the requirement $|D_{sn}(\omega)| \leqslant \sqrt{D_s(\omega)D_n(\omega)}$ can be imposed, the simple intuitive two-level filter frequency response of (15) is minimax robust. Specific details about $D_s(\omega)$ and $D_n(\omega)$ are not used in obtaining this filter.

Example: Let the known signal and noise spectral densities be

$$D_s(\omega) = \frac{1}{\omega^2 + 1}$$

and

$$D_n(\omega) = \frac{8}{\omega^2 + 16},$$

and let the upper bound $U(\omega)$ on $|D_{sn}(\omega)|$ be a constant value c. Fig. 1(a) illustrates the situation. Let ω_1 be the value of ω when $D_s(\omega) = c$. Then $\omega_1 = \sqrt{(1/c) - 1}$). As long as $c \leqslant 0.467$ (see Fig. 1(a)), we find from the above that the robust filter frequency response is given by

$$H_r(\omega) = \begin{cases} \dfrac{D_s(\omega) - c}{D_s(\omega) + D_n(\omega) - 2c}, & |\omega| \leqslant |\omega_1|; \\ 0, & |\omega| > |\omega_1|. \end{cases}$$

This frequency response is illustrated in Fig. 1(b).

Let e_r^0 be the mean-squared error when H_r is used and $D_{sn}(\omega) = 0$. Let e^0 be the corresponding mean-squared error of the nominally optimum filter based on the assumption that $D_{sn}(\omega)$

TABLE I
PERFORMANCE OF ROBUST AND NOMINALLY OPTIMUM FILTERS FOR EXAMPLE IN SECTION III-A
MEAN-SQUARED ERROR e^0 OF NOMINALLY OPTIMUM FILTER FOR $D_{sn}(\omega) = 0$ IS 1.71

c, Upper Bound on $\lvert D_{sn}(\omega)\rvert$	e_r^0, mse of H_r for $D_{sn}(\omega) = 0$	e_r^w, Worst mse of H_r	e^w, Worst mse of Nominally Optimum Filter
0.1	1.84	2.07	2.31
0.2	2.00	2.33	2.61
0.3	2.06	2.40	2.82
0.4	2.18	2.50	2.98
$c > \sqrt{0.5}$	2.55	2.55	3.21

$= 0$. In addition, define e_r^w and e^w to be the worst performances (mean-squared errors) for these robust and nominally optimum filters, respectively. Table I shows the values for mean-squared errors which have been computed for a range of values of c. Note that the filter which is optimum for uncorrelated signal and noise has a mean-squared error under this nominal assumption of $e^0 = 1.71$. Even for a small value 0.1 for c, Table I shows that this nominally optimum filter's performance degrades significantly, whereas the robust filter's performance is relatively quite good. When c is larger that $\sqrt{0.5}$ we have effectively no information on $D_{sn}(\omega)$ except for the requirement $|D_{sn}(\omega)| \leqslant \sqrt{D_s(\omega) D_n(\omega)}$.

B. Signal Spectrum Given, Bounds on Noise Spectrum, Upper Bound on $|D_{sn}(\omega)|$

We now generalize the above result by allowing $D_n(\omega)$ to vary between upper and lower bounds, with a constraint on the noise variance. Thus we assume that upper and lower bounds $U_n(\omega)$, $L_n(\omega)$ are given and $D_n(\omega)$ satisfies

$$L_n(\omega) \leqslant D_n(\omega) \leqslant U_n(\omega) \tag{16}$$

together with

$$\int_{-\infty}^{\infty} D_n(\omega)\, d\omega = 2\pi\sigma_n^2, \tag{17}$$

where σ_n^2 is given. The magnitude of the cross spectrum is still bound above by some $U(\omega)$. Here we assume that the signal spectrum $D_s(\omega)$ is exactly known. Although we will not develop it explicitly here, the case where $D_n(\omega)$ is exactly known and $D_s(\omega)$ lies between upper and lower bounds with a signal power constraint can be treated in a very similar way.

The solution for the lf matrix D^r now requires specification of $D_n^r(\omega)$ and $D_{sn}^r(\omega)$. Note that $D_{sn}^r(\omega)$ can always be expressed in terms of $D_n^r(\omega)$ and $D_s(\omega)$, by inserting $D_n^r(\omega)$ in place of $D_n(\omega)$ in (13). Therefore in finding D^r we can reduce the problem to that of finding the worst case, $D_n^r(\omega)$, for the noise spectrum which maximizes the integral (10), in which $D_s(\omega)$ is then known and $|D_{sn}(\omega)|$ has been replaced by the function $\min\{D_s(\omega), D_n(\omega), U(\omega)\}$.

It can be shown that $D_n^r(\omega)$ is specified according to one of the following three definitions; that is, one spectral density from the three cases below will make the matrix D^r an allowable matrix in Δ.

Case A: Here $D_n^r(\omega)$ is defined by

$$D_n^r(\omega) = \begin{cases} l(\omega), & L_n(\omega) \leqslant \min\{U(\omega), D_s(\omega), U_n(\omega)\}; \\ L_n(\omega), & \text{otherwise,} \end{cases}$$

where l can be any function satisfying the constraint

$$L_n(\omega) \leqslant l(\omega) \leqslant \min\{U(\omega), D_s(\omega), U_n(\omega)\}$$

and resulting in a $D_n^r(\omega)$ with the required variance. Clearly the lf matrix is not unique in this case. Here we define H_r, an optimum filter for D^r, by

$$H_r(\omega) = \begin{cases} \dfrac{D_{sx}^r(\omega)}{D_x^r(\omega)}, & \text{when this is well-defined;} \\ 1, & \text{otherwise.} \end{cases}$$

Thus H_r comes from (4) when it does not reduce to 0/0, and is taken to be unity otherwise. Notice that this latter condition occurs for ω values where $D_n^r(\omega)$ is defined as the function $l(\omega)$. When this happens any choice for $H_r(\omega)$ results in a filter optimum for D_r, but the above specific choice makes the filter robust.

Case B: We first define the function

$$f_k(\omega) \triangleq kD_s(\omega) + (1-k)\min\{U(\omega), D_s(\omega)\},$$

where k is a nonnegative parameter.

Then $D_n^r(\omega)$ is defined by

$$D_n^r(\omega) = \begin{cases} f_k(\omega), & L_n(\omega) \leqslant f_k(\omega) \leqslant U_n(\omega); \\ L_n(\omega), & L_n(\omega) > f_k(\omega); \\ U_n(\omega), & U_n(\omega) < f_k(\omega), \end{cases}$$

if a nonnegative k exists satisfying the noise power constraint. We define H_r by

$$H_r(\omega) = \begin{cases} \dfrac{D_{sx}^r(\omega)}{D_x^r(\omega)}, & \text{when this is well-defined;} \\ \dfrac{1}{(1+k)^2}, & \text{otherwise} \end{cases}$$

Note that another way of describing $D_n^r(\omega)$ for this case is $D_n^r(\omega) = \min\{U_n(\omega), \max\{f_k(\omega), L_n(\omega)\}\}$.

Case C: For this case $D_n^r(\omega)$ is defined by

$$D_n^r(\omega) = \begin{cases} l(\omega), & D_s(\omega) < U(\omega); \\ U_n(\omega), & D_s(\omega) \geqslant U(\omega), \end{cases}$$

where

$$\min\{U_n(\omega), \max\{D_s(\omega), L_n(\omega)\}\} \leqslant l(\omega) \leqslant U_n(\omega)$$

and such that the noise power constraint is satisfied. For this case we can define H_r by

$$H_r(\omega) = \begin{cases} \dfrac{D_{sx}^r(\omega)}{D_x^r(\omega)}, & \text{when it is well-defined;} \\ 0, & \text{otherwise.} \end{cases}$$

Of course, in each of these three cases $D_{sn}^r(\omega)$ is $-\min\{D_s(\omega), D_n^r(\omega), U(\omega)\}$. It is easy to show that one of the above three definitions for $D_n^r(\omega)$ will always result in a valid spectral density matrix D^r in Δ. In the Appendix a proof is given of the robustness of the filter H_r defined in Case B above. The proofs for the other two cases are quite similar.

Note that when $U(\omega) = 0$ for all ω, that is, when it is known that signal and noise are uncorrelated, Case B will always define the least-favorable noise spectrum. In this case the result is in agreement with earlier results [1]. On the other hand, when nothing is known about the extent of correlation, so that $U(\omega) \geqslant \sqrt{D_s(\omega) U_n(\omega)}$, either Case A or Case C defines the least-favorable noise spectrum. In addition, note that when the noise spectrum is exactly known, so that $U_n(\omega) = L_n(\omega)$, all cases are valid, allowing a choice of any value between 0 and 1 for $H_r(\omega)$ when it is not well-defined by (14) (Section III-A).

As we remarked earlier, it is possible to generate similar results for the case where $D_s(\omega)$ is in a bounded class and $D_n(\omega)$ is specified. To get more general results one can impose a lower bound $L(\omega)$ on $|D_{sn}(\omega)|$ in addition to the upper bound on it.

The solution for this case has been obtained, but we omit it here because it leads to a more involved result stated in several parts, and yet is similar in form to the one we have given for $L(\omega) \equiv 0$. The complete result for this general case is given in [9]. We consider next, instead, some special cases of classes Δ where both $D_s(\omega)$ and $D_n(\omega)$ are bounded.

C. Bounds on Signal and Noise Spectra, Upper Bound on $|D_{sn}(\omega)|$: Special Cases

We now allow uncertainty in both signal and noise spectra, so that in addition to (16) and (17) for the noise spectrum $D_n(\omega)$ we assume that $D_s(\omega)$ is similarly constrained by

$$L_s(\omega) \leqslant D_s(\omega) \leqslant U_s(\omega) \tag{18}$$

and

$$\int_{-\infty}^{\infty} D_s(\omega)\, d\omega = 2\pi\sigma_s^2, \tag{19}$$

where the bounds $L_s(\omega)$, $U_s(\omega)$ and the variance σ_s^2 are given. Furthermore, we retain the upper bound constraint of (11) on $|D_{sn}(\omega)|$.

The complete solution for the least-favorable $D \in \Delta$ with a specific lower bound $L(\omega)$ also imposed on $|D_{sn}(\omega)|$, while obtainable in principle in the same way as has been illustrated for the simpler cases so far, leads to a fairly complicated definition for D^r. In fact, even with $L(\omega) \equiv 0$ here the solution is not easy to describe. It is possible, however, to get relatively simple results under the condition that $L(\omega) \equiv 0$, for two extreme assumptions on $U(\omega)$.

Consider first the case where the upper bound of (11) is only the loose bound arising from the requirement that $|D_{sn}(\omega)|$ be bound above by $\sqrt{D_s(\omega) D_n(\omega)}$. That is, no specific information is available about $|D_{sn}(\omega)|$. This situation corresponds to a generalization of the special case mentioned in Section III-A, the signal and noise spectra now not being known but lying in classes defined by the band-model. It is therefore not surprising that one robust filter for this situation is a zero–one filter, with gain either zero or unity at any frequency ω. However, it should be noted that the least favorable matrix is *not* obtained by simply using the least-favorable spectral densities $(\hat{D}_s^r(\omega), \hat{D}_n^r(\omega))$ obtained for the band-model under an assumption that signal and noise are uncorrelated. As we had before, the least-favorable matrix D^r has components related by $D_{sn}^r(\omega) = -\min\{D_s^r(\omega), D_n^r(\omega)\}$, and (15) gives the robust filters with $D_s(\omega)$, $D_n(\omega)$ replaced with these $D_s^r(\omega)$, $D_n^r(\omega)$. We omit the statement of the complete solution, which is rather long and involves relationships between the bounding functions defining the class Δ. A detailed solution can be found in [8]. The most significant observation here is that the robust filters for these classes can be defined as zero–one filters.

Now consider the opposite situation when enough information is available about $|D_{sn}(\omega)|$ so that we have

$$U(\omega) < \min\{L_s(\omega), L_n(\omega)\}.$$

In this case we will always have, from (13), that the least-favorable matrix has cross-spectral component given by $D_{sn}^r(\omega) = -U(\omega)$. Using this in (10) we get that D^r is the matrix with cross-spectral elements $-U(\omega)$ and diagonal elements $D_s(\omega)$, $D_n(\omega)$ maximizing

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} U(\omega)\, d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\tilde{D}_s(\omega)\cdot\tilde{D}_n(\omega)}{\tilde{D}_s(\omega)+\tilde{D}_n(\omega)}\, d\omega, \tag{20}$$

where $\tilde{D}_s(\omega) = D_s(\omega) - U(\omega)$ and $\tilde{D}_n(\omega) = D_n(\omega) - U(\omega)$. To maximize (20) we need consider only the second integral involving $\tilde{D}_s(\omega)$ and $\tilde{D}_n(\omega)$. Notice that this is simply the minimum mean-squared error obtained from the optimum Wiener filter for signal and noise spectral densities $\tilde{D}_s(\omega)$ and $\tilde{D}_n(\omega)$. Also note that $\tilde{D}_s(\omega)$ and $\tilde{D}_n(\omega)$ are spectral densities lying in bounded classes defined as

$$L_s(\omega) - U(\omega) \leqslant \tilde{D}_s(\omega) \leqslant U_s(\omega) - U(\omega),$$
$$L_n(\omega) - U(\omega) \leqslant \tilde{D}_n(\omega) \leqslant U_n(\omega) - U(\omega),$$

and satisfying the power constraints

$$\int_{-\infty}^{\infty} \tilde{D}_s(\omega)\, d\omega = 2\pi\sigma_s^2 - \int_{-\infty}^{\infty} U(\omega)\, d\omega,$$
$$\int_{-\infty}^{\infty} \tilde{D}_n(\omega)\, d\omega = 2\pi\sigma_n^2 - \int_{-\infty}^{\infty} U(\omega)\, d\omega.$$

The least-favorable pair $(\tilde{D}_s^r(\omega), \tilde{D}_n^r(\omega))$ for this *uncorrelated* signal and noise situation can be found directly from previous results [1], and the desired components $D_s^r(\omega)$ and $D_n^r(\omega)$ of D^r then follow from $D_s^r(\omega) = \tilde{D}_s^r(\omega) + U(\omega)$ and $D_n^r(\omega) = \tilde{D}_n^r(\omega) + U(\omega)$. Here we find that a least-favorable D^r exists which has a corresponding well-defined optimum filter, the robust filter H_r. The significant result for this special case is that the robust filter for this situation can be obtained by modifying the original signal and noise spectral classes, obtaining the least-favorable pair for the uncorrelated-processes problem, and using this to obtain the least-favorable signal and noise spectra for the original problem. In fact, we see from the above and from (4) that the robust filter H_r is, here, the optimum filter for uncorrelated signal and noise with respective spectra $\tilde{D}_s^r(\omega)$ and $\tilde{D}_n^r(\omega)$.

This last result can be extended to apply for other power-constrained convex classes of signal spectra and noise spectra, whenever the upper bound $U(\omega)$ is lower than the minimum value attainable by either signal or noise spectra.

IV. Conclusion

We have obtained explicit solutions for robust filters for random signals in possibly correlated additive noise under spectral uncertainty classes described by upper and lower bounds. These results form an extension of earlier results which were obtained for the uncorrelated case. A situation which can occur in the correlated case is the nonuniqueness of the optimum filter for the least-favorable spectral matrix. This does not happen in the uncorrelated case, for which characterization of the robust filter in terms of a least-favorable spectral density pair is always possible.

In two special cases the results are particularly interesting. In one case very little is known about the cross spectrum $D_{sn}(\omega)$, whereas in the other case the cross spectrum is bound above by a relatively tight bound. We found that when $D_{sn}(\omega)$ is completely unspecified (or bound very loosely), a robust filter has unit gain when the least-favorable signal spectrum exceeds the least-favorable noise spectrum and zero gain otherwise; it is an ideal filter. From the last part of Section III it follows, on the other hand, that when the upper bound $U(\omega)$ on $D_{sn}(\omega)$ is lower than the minimum values of both $D_s(\omega)$ and $D_n(\omega)$, the robust filter is optimum for uncorrelated signal and noise with spectra which are least-favorable for modified classes defining uncorrelated signals and noise.

Appendix

Proof for Results in Section III-B, Case B

From (2), noting that $D_{sx}(\omega) = D_s(\omega) + D_{sn}(\omega)$ and $D_{xx}(\omega) = D_s(\omega) + D_n(\omega) + 2\,\mathrm{Re}[D_{sn}(\omega)]$, we get

$$e(D^r, H_r) - e(D, H_r) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(\omega)\, d\omega$$

where, since $H_r(\omega)$ is real,

$$P(\omega) = H_r^2(\omega)[D_n^r(\omega) - D_n(\omega)] - 2H_r(\omega)(1 - H_r(\omega))[\mathrm{Re}\{D_{sn}^r(\omega)\} - \mathrm{Re}\{D_{sn}(\omega)\}]. \tag{A1}$$

When $H_r(\omega)$ is well-defined we have either $D_{sn}^r(\omega) = -\min\{D_s(\omega), D_n^r(\omega)\}$ or $D_{sn}^r(\omega) = -U(\omega)$, from (13). In the former case $H_r(\omega)(1 - H_r(\omega))$ is zero, and in the latter case $H_r(\omega)$ and thus $H_r(\omega)(1 - H_r(\omega))$ lies in [0, 1], and $\mathrm{Re}\{D_{sn}^r(\omega)\} - \mathrm{Re}\{D_{sn}(\omega)\} \leqslant 0$. Thus when $H_r(\omega)$ is well-defined,

$$P(\omega) \geqslant H_r^2(\omega)[D_n^r(\omega) - D_n(\omega)]. \tag{A2}$$

Equation (4) does not yield a well-defined $H_r(\omega)$ when $D_s(\omega) = D_n^r(\omega) = -D_{sn}^r(\omega)$. In this case, using the inequality

$$\mathrm{Re}\{D_{sn}(\omega)\} \geqslant -\frac{D_s(\omega) + D_n(\omega)}{2}$$

(from $\mathrm{Re}\{D_{sn}(\omega)\} \geqslant -|D_{sn}(\omega)| \geqslant -\sqrt{D_s(\omega)D_n(\omega)} \geqslant -\frac{1}{2}[D_s(\omega) + D_n(\omega)]$) in (A1), we get for $0 \leqslant H_r(\omega) \leqslant 1$

$$P(\omega) \geqslant H_r(\omega)[D_n^r(\omega) - D_n(\omega)]. \tag{A3}$$

a) When $D_n^r(\omega)$ in Case B is given as $D_n^r(\omega) = f_k(\omega)$, from (12) and because $\min\{D_s(\omega), U(\omega)\} \leqslant f_k(\omega)$, we have $|D_{sn}^r(\omega)| = \min\{D_s(\omega), U(\omega)\}$.

Now note that if (4) gives a well-defined H_r, it is (using 13)

$$H_r(\omega) = \frac{D_s(\omega) - |D_{sn}^r(\omega)|}{D_s(\omega) - |D_{sn}^r(\omega)| + D_n^r(\omega) - |D_{sn}^r(\omega)|} \tag{A4}$$

which in this case becomes $H_r(\omega) = 1/(1+k)$. When (4) does not define H_r, it is taken to be $1/(1+k)^2$. Thus, using (A2) and (A3), we have

$$P(\omega) \geqslant \frac{1}{(1+k)^2}[D_n^r(\omega) - D_n(\omega)].$$

b) For values of ω where $D_n^r(\omega) = L_n(\omega)$, we again have $|D_{sn}^r(\omega)| = \min\{D_s(\omega), U(\omega)\}$, and (A4) now gives $H_r(\omega) \leqslant 1/(1+k)$, because here $D_n^r(\omega) > f_k(\omega)$. This gives, because $D_n^r(\omega) - D_n(\omega) \leqslant 0$, the result

$$P(\omega) \geqslant \frac{1}{(1+k)^2}[D_n^r(\omega) - D_n(\omega)].$$

c) When $D_n^r(\omega) = U_n(\omega)$, note that we have $D_n^r(\omega) < kD_s(\omega) + (1-k)|D_{sn}^r(\omega)|$, since here $U_n(\omega) < f_k(\omega) \leqslant D_s(\omega)$ and $|D_{sn}^r(\omega)| = \min\{U_n(\omega), \min\{D_s(\omega), U(\omega)\}\}$. Thus from (A4) $H_r(\omega) \geqslant 1/(1+k)$, and because here $D_n^r(\omega) - D_n(\omega) \geqslant 0$, we get again

$$P(\omega) \geqslant \frac{1}{(1+k)^2}[D_n^r(\omega) - D_n(\omega)].$$

For all ω we have shown that the above inequality is true; the result follows, by integrating, that

$$e(D^r, H_r) - e(D, H_r) \geqslant 0.$$

References

[1] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171–185, Oct./Nov. 1977.

[2] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 521–526, June 1980.

[3] H. V. Poor and D. P. Looze, "Minimax state estimation for linear stochastic systems with noise uncertainty," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 902–906, Aug. 1981.

[4] K. S. Vastola and H. V. Poor, "Robust linear estimation of stationary discrete-time signals," *Proc. 1981 Conf. Inform. Sci. Syst.*, The Johns Hopkins University, pp. 512–516, Mar. 1981.

[5] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753–1758, 1965.

[6] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory*, vol. IT-27, 242–247, Mar. 1981.

[7] H. V. Poor, "Minimax linear smoothing for capacities," *Ann. Prob.*, vol. 10, pp. 504–507, May 1982.

[8] G. Moustakides and S. A. Kassam, "Robust Wiener filters for correlated signals and noise," *Proc. 1980 Conf. Inform. Sci. Syst.*, (Princeton University), pp. 308–313, Mar. 1980.

[9] G. Moustakides, "Robust Wiener filters for imprecise second-order statistics," M.S.E. thesis, Department of Systems Engineering, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, PA, Aug. 1980.

A Simple Suboptimum Estimator of Prior Probability in Mixtures

REZA JALALINASAB AND JAMES A. BUCKLEW, MEMBER, IEEE

***Abstract*—A simple relative frequency type estimator of the prior probability in a mixture of two known density functions is presented. Examples are given demonstrating the ease of design and implementation of this estimator structure.**

I. Introduction

Given that an unknown density function $f(x)$ is a mixture of two known density functions $f_1(x)$ and $f_2(x)$, where the prior probability is unknown, the problem is to estimate the prior probability on the basis of N statistically independent observations. The density $f(x)$ can be written as

$$f(x) = \pi f_1(x) + (1-\pi) f_2(x), \tag{1}$$

where π the prior probability to be estimated, is assumed to be uniformly distributed on [0, 1].

This type of problem arises in pattern recognition problems where one is studying the distribution of observations belonging to individual populations and where the population mix is unknown. It also has application in biological and physical sciences. (See Choi [1], Makov and Smith [2], Davisson [3], Sakrison [4], Yakow [5], Blischke [6], and Makov [7] for examples and further references.)

The estimator is presented in Section III. It is a member of a class of estimators first presented by Boes [8]. In Sections IV and V we give three numerical examples, and we compare our results with a recursive estimation scheme due to Kazakos [9].

The calculations involved in designing Kazakos' estimator are moderately complicated. On the other hand, the estimator introduced in this paper is very simple to design. The examples in Section IV are intended to illustrate the simplicity of the estimator design. We emphasize that Kazakos' estimate always gives a lower variance, but in some applications one has a great number of samples and simplicity is more desirable.

II. Development

Let f_1, f_2, π, and N independent samples $\{x_1, \cdots, x_n\}$ be as stated in the preceding section. Consider a set of the following form:

$$A = \left\{x: \frac{f_1(x)}{f_2(x)} \geqslant T\right\}, \tag{2}$$

where T is a threshold to be described later. Define

$$P(A) = \int_A f(x)\, dx \tag{3}$$

Manuscript received January 19, 1982; revised November 11, 1982.

R. Jalalinasab was with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI.

J. A. Bucklew is with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706.

0018-9448/83/0700-0619$01.00 ©1983 IEEE