ROBUST WIENER FILTERS FOR CORRELATED SIGNALS AND NOISE*

George Moustakides
Saleem A. Kassam
Department of Systems Engineering
The Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, Pennsylvania 19104

ABSTRACT

We consider the Wiener filtering problem, when the cross spectral density matrix of the signal and noise is not exactly known. We obtain filters which are saddle point solutions for the criterion of performance (mean square error, MSE) over the classes of allowable density matrices. Solutions for various classes are given.

I. INTRODUCTION

In classical Wiener filtering, we need to assume exact knowledge of the spectral density matrix of signal and noise. In many applications this assumption of exact knowledge is unreasonable. A more realistic assumption is that our matrix belongs to a class of density matrices. This class can be defined according to our knowledge of the true density matrix. For this problem we will derive a filter that performs in an optimum way over the whole class.

Nahi and Weiss [1,2] derived the bounding filter. This filter is a Wiener filter for some density matrix D_b and if it is applied to any matrix from the class the MSE error is bounded by the minimum MSE for D_b. Because D_b does not usually belong to the class, there is no matrix in the class that can reach the bound of the MSE. This means that there is a possibility for better performance.

Kassam and Lim [3] derived the robust Wiener filter, when signal and noise are uncorrelated (density matrix diagonal). This filter sets a bound on the MSE and the bound can be reached by some matrix from the class. In [4] and [5], Poor generalized some of this work. The present paper extends the above idea to the correlated signal and noise case.

Presence of correlation is possible in many applications. An example is a multipath channel with a strong signal component, weak unwanted multipath signals and regular noise. The total "noise" is correlated with the signal.

II. ROBUST WIENER FILTERS

Let us assume that our processes are real, wide sense stationary and zero mean and the noise process is additive. We also assume that there exists a spectral density matrix D for the signal, noise processes, given by:

$$D = \begin{bmatrix} D_s(w) & D_{sn}(w) \\ D^*_{sn}(w) & D_n(w) \end{bmatrix}$$

where s is for signal, n for noise and (*) for complex conjugate. The properties that characterize such a matrix D for real processes are:

i. $D_s(w)$, $D_n(w)$ real, even, non-negative functions

ii. $|D_{sn}(w)|^2 \le D_s(w)\cdot D_n(w)$ (1)

So D is non-negative definite, and diagonal elements are even functions of w.

Given random processes s(t) and n(t) with density matrix D and a filter h(t) with Fourier transform H(w), the MSE for signal estimation using this filter is

$$e(D,H) = E\left[s(t)-\int_{-\infty}^{\infty} h(v)x(t-v)dv\right]^2$$

$$= R_{ss}(0)-2\int_{-\infty}^{\infty} h(v)R_{sx}(v)dv + \iint_{-\infty}^{\infty} h(v)h(u)R_{xx}(v-u)dvdu, \quad (2)$$

where $x(t)=s(t)+n(t)$ and $R_{ij}(\tau)$ is the cross correlation between i and j. Using Fourier transform and Parseval's theorem, (2) can be written as:

$$e(D,H)=\frac{1}{2\pi}\int_{-\infty}^{\infty} [D_s(w)-2H(w)D_{sx}(w)+|H(w)|^2 D_x(w)]dw \quad (3)$$

The optimum filter for D is given by

$$H_o = \frac{D_{sx}(w)}{D_x(w)} \quad (4)$$

If we substitute (4) into (3) we have that the optimum MSE is given by:

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D_s(w)D_x(w)-|D_{sx}(w)|^2}{D_x(w)} dw \quad (5)$$

2.1 Definition of Robust Filter

Assume that a class Δ of density matrices is given. A robust filter H_r is defined by the following properties:

a. H_r is an optimum filter for some matrix $D^r \epsilon \Delta$, so that $e(D^r,H_r) \le e(D^r,H)$ for any filter H.

b. For any $D \epsilon \Delta$ we have: $e(D,H_r) \le e_{op}(D^r)=e(D^r,H_r)$

Combining a and b we have the saddle point relation

$$e(D,H_r) \le e_{op}(D^r) \le e(D^r,H) \quad (6)$$

*This research is supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR 77-3154.

for any $D \varepsilon \Delta$ and any filter H.

2.2 Theorem 0.

Let Δ be a convex class of density matrices. Then the pair (D^r, H_r) is a saddle point solution for MSE over class Δ and the class of all linear filters (it satisfies (6)), if and only if:

$$e_{op}(D^r) = \max_{D\varepsilon\Delta} e_{op}(D)$$

The proof is given in appendix A.

Theorem 0 is the key point in our search for the robust filter, because based on it we have only to maximize $e_{op}(D)$.

2.3 Maximization of $e_{op}(D)$

From (5) if we write the error in terms of s and n we get:

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D_s(w)\cdot D_n(w) - |D_{sn}(w)|^2}{D_s(w)+D_n(w)+2\mathrm{Re}[D_{sn}(w)]}\, dw \quad (7)$$

For every w and for given $D_s(w), D_n(w), |D_{sn}(w)|$, the worst $\mathrm{Re}[D_{sn}(w)]$ is $-|D_{sn}(w)|$, because it minimizes the denominator. This gives:

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D_s(w)\cdot D_n(w) - |D_{sn}(w)|^2}{D_s(w)+D_n(w)-2|D_{sn}(w)|}\, dw \quad (8)$$

For given $D_s(w), D_n(w)$ expression (8) as a function of $|D_{sn}(w)|$ is:

a. increasing for:

$$0 \le |D_{sn}(w)| \le \min\{D_s(w), D_n(w)\}$$

b. decreasing for:

$$\min\{D_s(w), D_n(w)\} \le |D_{sn}(w)| \le \sqrt{D_s(w), D_n(w)}$$

From condition a and b we can say that, given $D_s(w)$, $D_n(w)$, the worst $|D_{sn}(w)|$ is the one that is as close as possible to $\min\{D_s(w), D_n(w)\}$. The worst $D_{sn}(w)$ is $D_{sn}(w) = -|D_{sn}(w)|$.

III. APPLICATIONS

1. $D_s(w)$, $D_n(w)$ given with upper and lower bounds on $|D_{sn}(w)|$.

Let $0 \le L(w) \le |D_{sn}(w)| \le U(w)$, with $L(w), U(w)$ given. If we define:

$$A(w) = \min\{D_s(w), D_n(w)\} \quad (9)$$

then the worst-case characteristic $|D^r_{sn}(w)|$ is given by:

$$|D^r_{sn}(w)| = \begin{cases} L(w) & \text{if } A(w) \le L(w) \\ A(w) & \text{if } L(w) \le A(w) \le U(w) \\ U(w) & \text{if } U(w) \le A(w) \end{cases}$$

where $A(w)$ defined in (9).

From section 2.2 we have that $D^r_{sn}(w) = -|D^r_{sn}(w)|$. Figure 1 illustrates this case.

2. Upper and lower bounds on $D_s(w)$ and $D_n(w)$.

We assume that bounds $L_i(w), U_i(w)$ for $D_i(w)$ are given:

$$0 \le L_i(w) \le D_i(w) \le U_i(w) \qquad i=s,n.$$

In addition we will assume knowledge of the total power of signal and noise,

$$\int_{\infty}^{\infty} D_i(w)\,dw = 2\pi\sigma_i^2 \qquad i=s,n$$

where σ_i are known.

A. If there are no bounds on $|D_{sn}(w)|$, it can reach the value $\min\{D_s(w), D_n(w)\}$. Under this condition the error is:

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\{D_s(w), D_n(w)\}\,dw \qquad (10)$$

Because of the power constraints, there are several sub-cases. First we will give some definitions.

$$A_s(w) = \min\{U_s(w),\ \max[L_s(w),\ L_n(w)]\}$$

$$A_n(w) = \min\{U_n(w),\ \max[L_s(w),\ L_n(w)]\}$$

$$B_s(w) = \min\{U_s(w),\ \max[L_s(w),\ U_n(w)]\}$$

$$B_n(w) = \min\{U_n(w),\ \max[L_n(w),\ U_s(w)]\}$$

Figure 2 illustrates the definition of $A_s(w), B_s(w)$.

A1. $\boxed{\int_{-\infty}^{\infty} A_s(w)\,dw > 2\pi\sigma_s^2,\ \int_{-\infty}^{\infty} A_n(w)\,dw > 2\pi\sigma_n^2}$ then:

$$D^r_s(w) = \begin{cases} L_s(w) & \text{if } A_s(w) = L_s(w) \\ \ell_s(w) & \text{otherwise} \end{cases}$$

$$D^r_n(w) = \begin{cases} L_n(w) & \text{if } A_n(w) = L_n(w) \\ \ell_n(w) & \text{otherwise} \end{cases}$$

$$e_{op}(D^r) = \sigma_s^2 + \sigma_n^2 - \frac{1}{2\pi}\int \{L_s(w) + L_n(w) - \min(L_n(w), L_s(w))\}\,dw$$

Where $\ell_s(w)$, $\ell_n(w)$ are arbitrary functions with

$$\ell_s(w) \le A_s(w) \text{ and } \ell_n(w) \le A_n(w)$$

but enough for $D^r_s(w)$ and $D^r_n(w)$ to fulfill the power constraints.

A2. $$\int_{-\infty}^{\infty} A_s(w)\,dw > 2\pi\sigma_s^2, \quad \int_{-\infty}^{\infty} A_n(w)\,dw < 2\pi\sigma_n^2$$

$D_s^r(w)$ is as in case A1

$$D_n^r(w) = \begin{cases} U_n(w) & \text{if } U_n(w) = A_n(w) \\ \ell_n(w) & \text{otherwise} \end{cases}$$

where $\ell_n(w)$ arbitrary function with $\ell_n(w) \geq A_n(w)$.

$$e_{op}(D^r) = \sigma_s^2 - \frac{1}{2\pi}\int_{-\infty}^{\infty} \{L_s(w) - \min(L_s(w), U_n(w))\}$$

A3. $$\int_{-\infty}^{\infty} B_s(w)\,dw > 2\pi\sigma_s^2 > \int_{-\infty}^{\infty} A_s(w)\,dw$$
$$\int_{-\infty}^{\infty} B_n(w)\,dw > 2\pi\sigma_n^2 > \int_{-\infty}^{\infty} A_n(w)\,dw$$

also assume $$2\pi\sigma_n^2 - \int_{-\infty}^{\infty} A_n(w) > 2\pi\sigma_s^2 - \int_{-\infty}^{\infty} A_s(w)\,dw$$

$$D_s^r(w) = \begin{cases} A_s(w) & \text{if } A_s(w) = B_s(w) \\ \ell_s(w) & \text{otherwise} \end{cases}$$
$$D_n^r(w) = \begin{cases} A_n(w) & \text{if } A_n(w) = B_n(w) \\ \ell_n(w) & \text{otherwise} \end{cases}$$

Where $A_s(w) \leq \ell_s(w) \leq B_s(w)$ and $D_s^r(w) \leq \ell_n(w) \leq B_n(w)$

$$e_{op}(D^r) = \sigma_s^2 - \frac{1}{2\pi}\int_{-\infty}^{\infty} \{L_s(w) - \min(U_n(w), L_s(w))\}\,dw$$

A4. $$\int_{-\infty}^{\infty} B_s(w)\,dw > 2\pi\sigma_s^2 > \int_{-\infty}^{\infty} A_s(w)\,dw, \quad 2\pi\sigma_n^2 > \int_{-\infty}^{\infty} B_n(w)\,dw$$

$D_s^r(w)$ as in case A3.

$$D_n^r(w) = \begin{cases} U_n(w) & \text{if } U_n(w) = A_n(w) \\ \ell_n(w) & \text{otherwise} \end{cases}$$

Where $\ell_n(w) > D_s^r(w)$

$$e_{op}(D^r) = \sigma_s^2 - \frac{1}{2\pi}\int_{-\infty}^{\infty} \{L_s(w) - \min(U_n(w), L_s(w))\}\,dw$$

A5. $$2\pi\sigma_s^2 > \int_{-\infty}^{\infty} B_s(w)\,dw, \quad 2\pi\sigma_n^2 > \int_{-\infty}^{\infty} B_n(w)\,dw$$

$$D_s^r(w) = \begin{cases} U_s(w) & \text{if } U_s(w) = B_s(w) \\ \ell_s & \text{otherwise} \end{cases}$$

$$D_n^r(w) = \begin{cases} U_n(w) & \text{if } U_n(w) = B_n(w) \\ \ell_n(w) & \text{otherwise} \end{cases}$$

Where $\ell_s(w) > B_s(w)$ and $\ell_n(w) > B_n(w)$.

$$e_{op}(D^r) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\{U_s(w), U_n(w)\}\,dw$$

In the cases above, interchanging the roles of s and n in the conditions gives us similar results. The proof for cases A1 and A3 can be found in Appendix B.

B. Suppose we are given a function $R(w)$ such that

$$0 \leq |D_{sn}(w)| \leq R(w) \leq \min(L_s(w), L_n(w))$$

Because of the maximization problem $|D_{sn}(w)|$ has to be as close as possible to $\min(D_s(w), D_n(w))$, so

$|D_{sn}^r(w)| = R(w)$. To find the pair $D_s^r(w)$, $D_n^r(w)$ notice that:

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D_s(w)\cdot D_n(w) - R^2(w)}{D_s(w)+D_n(w)-2R(w)}\,dw$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} R(w)\,dw + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S(w)\;N(w)}{S(w)+N(w)}\,dw \tag{11}$$

Where $S(w) = D_s(w) - R(w) \geq 0$ and $N(w) = D_n(w) - R(w) \geq 0$. To maximize (11) it is enough to maximize the second term. But this term is the expression for the minimum MSE for uncorrelated signal $S(w)$ and noise $N(w)$, with

$$L_s(w) - R(w) \leq S(w) \leq U_s(w) - R(w)$$
$$L_n(w) - R(w) \leq N(w) \leq U_n(w) - R(w)$$

and $$\int_{-\infty}^{\infty} S(w)\,dw = 2\pi\sigma_s^2 - \int_{-\infty}^{\infty} R(w)\,dw$$
$$\int_{-\infty}^{\infty} N(w)\,dw = 2\pi\sigma_n^2 - \int_{-\infty}^{\infty} R(w)\,dw.$$

This problem has been solved in [3] and gives the solution to the present case.

3. <u>Given classes for $D_s(w)$, $D_n(w)$</u>

We assume again knowledge of the total power of signal and noise. When there is no restriction on $D_{sn}(w)$ we have seen that the error is given by (10) and it can be written in the following way:

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(x)\cdot D_n(w)\cdot dw \tag{12}$$

where

$$g(x) = \begin{cases} x \text{ for } 0 \le x \le 1 \\ 1 \text{ for } \quad x > 1 \end{cases}$$

and

$$x = \frac{D_s(w)}{D_n(w)}$$

$g(x)$ is a convex function of x for $x \ge 0$. So (12) can be used as a measure of distance between $D_s(w)$ and $D_n(w)$. Now for given classes of $D_s(w)$ and $D_n(w)$ we want the pair $D_s^r(w)$, $D_n^r(w)$ that has the maximum distance. It turns out as we can see from [4,6] and [7] that this pair does not depend on the form of $g(x)$. As long as $g(x)$ is any convex function the pair is always the same. Solutions to this problem are also given in these works for various classes.

IV. NUMERICAL RESULTS

We will assume the case when signal and noise are given as in Figure 3 and there is no restriction on $|D_{sn}(w)|$. The robust filter is simply:

$$H_r(w) = \begin{cases} 1 \text{ when } D_s(w) \ge D_n(w) \\ 0 \text{ otherwise} \end{cases}$$

It turns out that this filter has the same error performance for any $D_{sn}(w)$. If we use instead of $H_r(w)$ the filter

$$H_u(w) = \frac{D_s(w)}{D_s(w)+D_n(w)}$$

assuming that signal and noise are uncorrelated, then the error is given by (3) and it becomes maximum for $D_{sn}(w) = -\sqrt{D_s(w)D_n(w)}$.

In Table 1 are given some numbers for different b. The error is e_r when we use $H_r(w)$ and e_u is the worst error when we use $H_u(w)$. In the third column we can see the percentage of performance improvement. Also shown is the optimum error e_u^* of $H_u(w)$ when $D_{sn}(w)=0$.

APPENDIX A

Before proving Theorem 0, we will prove the following lemma.

Lemma 1. Let $D', D'' \varepsilon \Delta$, define $D^\varepsilon = (1-\varepsilon)D' + \varepsilon D''$ $0 \le \varepsilon \le 1$ then the expression:

$$G(\varepsilon,w) = \frac{D_s^\varepsilon(w) \cdot D_x^\varepsilon(w) - |D_{sx}^\varepsilon(w)|^2}{D_x^\varepsilon(w)}$$

is a convex function of ε.

Proof. It is sufficient to prove that:

$$G(\varepsilon,w) \ge (1-\varepsilon)G(0,w) + \varepsilon G(1,w).$$

Subtracting each side from $D_s^\varepsilon(w)$ we have to prove that

$$\frac{|D_{sx}^\varepsilon|^2}{D_x^\varepsilon} \le (1-\varepsilon)\frac{|D'_{sx}|^2}{D'_x} + \varepsilon \frac{|D''_{sx}|^2}{D''_x} \quad \text{or}$$

$$|(1-\varepsilon)D'_{sx} + \varepsilon D''_{sx}|^2 \le \left\{(1-\varepsilon)\frac{|D'_{sx}|^2}{D'_x} + \varepsilon\frac{|D''_{sx}|^2}{D''_x}\right\} \cdot$$

$$\cdot \{(1-\varepsilon)D'_x + \varepsilon D''_x\}$$

But this is the Schwartz inequality.

Proof of Theorem 0.

The only if part is easy. From (6)

$$e(D,H_r) \le e(D^r,H_r) = e_{op}(D^r)$$

but $e(D,H_r) \ge e_{op}(D)$ so $e_{op}(D^r) \ge e_{op}(D)$.

To prove the if part, define $D^\varepsilon = (1-\varepsilon)D^r + \varepsilon D$, $0 \le \varepsilon \le 1$ where $D \varepsilon \Delta$. Because of the lemma and the fact that

$$e_{op}(D^\varepsilon) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\varepsilon,w)dw$$

$e_{op}(D^\varepsilon)$ is a convex function of ε. So:

$$e_{op}(D^\varepsilon) \ge (1-\varepsilon)e_{op}(D^r) + \varepsilon\, e_{op}(D) \text{ and}$$

$$0 \ge \frac{e_{op}(D^\varepsilon) - e_{op}(D^r)}{\varepsilon} \ge e_{op}(D) - e_{op}(D^r) \tag{13}$$

Because $\frac{e_{op}(D^\varepsilon) - e_{op}(D^r)}{\varepsilon}$ is monotonic with respect to ε and bounded as we can see from (13) its limit exists as $\varepsilon \to o^+$. But:

$$\frac{e_{op}(D^\varepsilon) - e_{op}(D^r)}{\varepsilon} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{G(\varepsilon,w)-G(o,w)}{\varepsilon}\, dw \tag{14}$$

Now $\frac{d}{d\varepsilon}G(\varepsilon,w)$ is also monotonic with respect to ε.

From (13) and (14), we have

$$0 \ge \lim_{\varepsilon \to o^+} \int_{-\infty}^{\infty} \frac{1}{2\pi} \frac{G(\varepsilon,w) - G(o,w)}{\varepsilon}\, dw =$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \lim_{\varepsilon \to o^+} \frac{G(\varepsilon,w) - G(o,w)}{\varepsilon}\, dw =$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{d}{d\varepsilon} G(\varepsilon,w)\Big|_{\varepsilon=o^+} dw$$

But the last expression after we take the derivative and make some manipulations, equals $e(D,H_r) - e_{op}(D^r)$. So $e_{op}(D^r) \ge e(D,H_r)$.

APPENDIX B

For the second application we will outline the proofs for subcases A1 and A3. In a similar way, we can prove the rest of the cases.

Lemma 2. If a,b,c non-negative numbers with $a \geq c$ then:

$$a - c \geq \min(a,b) - \min(b,c)$$

Proof. $a - c \geq 0 \geq \min(a,b) - b$

$$a - c \geq \min(a,b) - c \quad \text{so}$$

$$a - c \geq \min(a,b) - \min(b,c)$$

From (10) we have that the error is

$$e_{op}(D) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\{D_s(w), D_n(w)\} dw$$

Case A1.

$$\min\{D_s, D_n\} = \min\{L_s, D_n\} + [\min(D_s, D_n\} - \min\{L_s, D_n\}]$$

(using Lemma 2) $\leq \min\{L_s, D_n\} + D_s - L_s$

(using Lemma 2) $\leq \min\{L_s, L_n\} + D_s + D_n - L_s - L_n$

So

$$e_{op}(D) \leq \sigma_s^2 + \sigma_n^2 - \frac{1}{2\pi}\int_{-\infty}^{\infty} [L_s(w) + L_n(w) - \min\{L_s(w), L_n(w)\}] dw$$

And we have equality when $D = D^r$.

Case A3.

$$\min\{D_s, D_n\} = \min\{L_s, D_n\} + [\min\{D_s, D_n\} - \min\{L_s, D_n\}] \leq \min\{L_s, D_n\} + D_s - L_s$$

$$\leq \min\{L_s, U_n\} + D_s - L_s \quad \text{and}$$

$$e_{op}(D) \leq \sigma_s^2 - \frac{1}{2\pi}\int_{-\infty}^{\infty} [L_s(w) - \min\{L_s(w), U_n(w)\}] dw$$

with equality when $D = D^r$.

REFERENCES

1. N.E. Nahi and I.M. Weiss, "Optimum Wiener Bounding Filters in the Presence of Uncertainties," Information and Control, Vol. 28, 1975, pp. 179-191.

2. N.E. Nahi and I.M. Weiss, "Bounding Filter: A Simple Solution to Lack of Exact A Priori Statistics," Information and Control, Vol. 39, 1978, pp. 212-224.

3. S.A. Kassam and T.L. Lim, "Robust Wiener Filters," Journal of the Franklin Institute, Vol. 304, No. 415, October/November, 1977.

4. H.V. Poor, "Some Results in Robust Wiener Filtering," IEEE Decision and Control Conf., January, 1979, pp. 725-730.

5. H.V. Poor, "Further Results in Robust Wiener Filtering," IEEE Decision and Control Conf., December, 1979.

6. S.M. Ali and S.D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another," J. Royal Stat. Soc., Vol. 28, 1966, pp. 131-142.

7. S.A. Kassam, "An Extension of Huber's Results on Robust Tests," Proceedings of Midwest Symposium on Circuits and Systems, pp. 525-529, 1979.

FIGURES AND TABLES

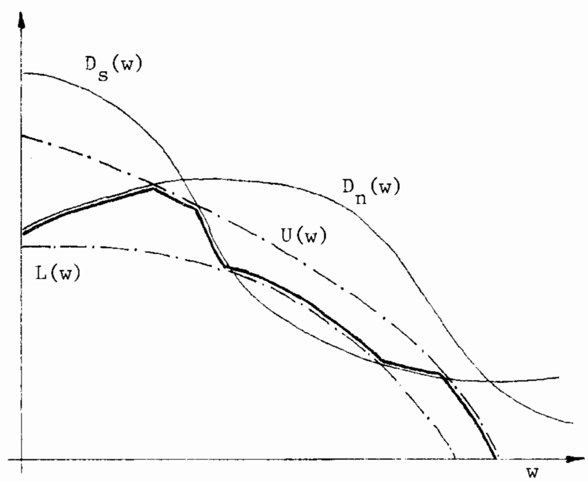

Figure 1 Worst-case $|D^r_{sn}|$ for case 1, section III

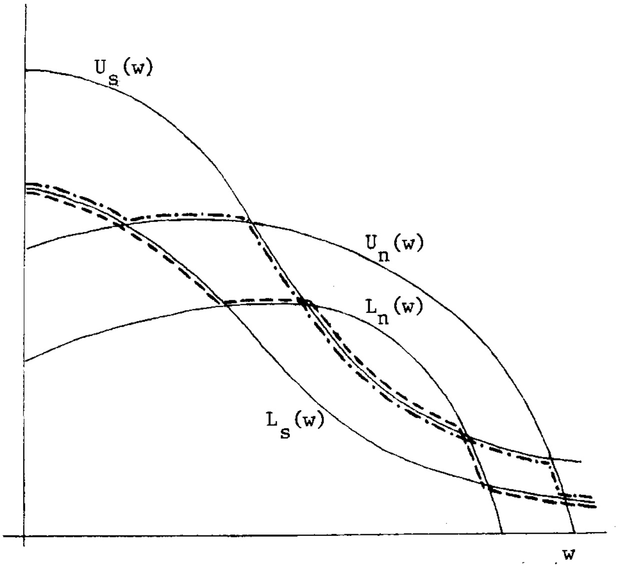

Figure 2 Illustration of functions defined in case 2A, section III

--- for $A_s(w)$

-·- for $B_s(w)$

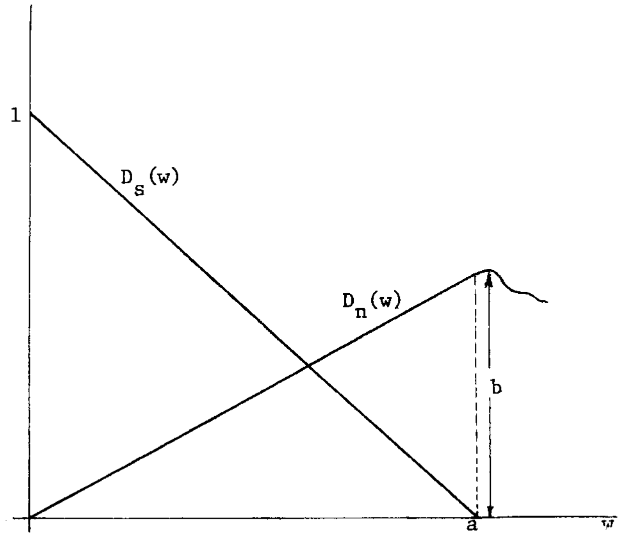

Figure 3 Example spectra of section IV

b	e_r	e_u	%	e_u^*
0.10	0.09	0.12	34.5	0.07
1.00	0.50	0.62	25.7	0.33
5.00	0.83	1.10	31.7	0.62
10.0	0.91	1.20	34.5	0.73

TABLE I Performance Comparision of Robust and Nominally Optimum Filter for Example of Figure 3. (No bound on $|D^r_{sn}(w)|$).