The Annals of Probability
1973, Vol. 1, No. 1, 175-179

A NOTE ON MINIMAX FILTERING[1]

BY LEO BREIMAN

A minimax procedure is found for filtering the "signal" from the "noise" in a stationary time series when it is known only that the spectral distribution function of the "signal" lies in a convex set defined by linear inequalities.

1. Summary. The problem we solve here is: Given second order stationary processes

$$\{X_n\}, \{Y_n\}, \{Z_n\}, \qquad -\infty < n < \infty$$

such that

$$X_n = Y_n + Z_n$$

where the $\{Y_n\}$ variables are uncorrelated with the $\{Z_n\}$ variables. The spectral distribution function $G(d\lambda)$ of the $\{Z_n\}$ process is assumed known. But all that is known about the spectral distribution function $F(d\lambda)$ of the $\{Y_n\}$ process is that it lies in the set $\mathscr{F}$ defined by the M inequalities

$$\int_{(-\pi,\pi]} f_m(\lambda)F(d\lambda) \leqq 1, \qquad m = 1, \cdots, M,$$

where the f_m are nonnegative and continuous. Find the linear filter

$$\hat{Y}_n = \sum a_{n-k} X_k$$

which is minimax for estimating the Y_n component in the sense that $\hat{Y}_n$ minimizes the maximum of

$$E(Y_n - \hat{Y}_n)^2$$

where the maximum is taken over all $\{Y_n\}$ processes whose spectral distributions lie in $\mathscr{F}$.

2. Background. The background for this problem is that frequently the observed process, $\{X_n\}$, $-\infty < n < \infty$, consists of a rapidly fluctuating component $\{Z_n\}$, superimposed on a slowly changing component $\{Y_n\}$. We want to smooth (filter) out the rapidly fluctuating component so as to isolate the long-term process $\{Y_n\}$.

In many particular applications, the spectral distribution of the noise can be assumed known, in the sense that it can be reasonably accurately estimated. For example, if the $\{Z_n\}$ process is assumed uncorrelated then all that is needed is a good estimate for the common variance. If $\{Y_n\}$ is slowly changing it is not difficult to get such estimates. For instance

$$\hat{\sigma}^2 = \frac{1}{2N} \sum_1^N (X_{k+1} - X_k)^2 .$$

Received February 8, 1972; revised May 2, 1972.

[1] This work supported by U.S. Department of Transportation, Federal Highway Administration Contract Number FH-11-7627 to System Development Corporation.

Key words and phrases. Stationary processes, minimax filtering.

If we also knew the covariances (or spectral distribution) of the $\{Y_n\}$ process, we could solve for the optimal linear filter in the standard way. Often all that is known about the $\{Y_n\}$ process are the observations on the $\{X_n\}$ process. We can, in principle, estimate the spectrum of the $\{Y_n\}$ process from these observations. Since $\{Y_n\}$ is assumed slow-changing, its spectrum is mainly concentrated in the low frequencies. But, estimating low frequency spectrum involves estimating covariances of widely separated variables in the $\{Y_n\}$ process. This often necessitates a larger sample size than is available.

Another procedure would be possible if one could find a frequency ω_0 such that $\{Y_n\}$ contains no frequencies higher than ω_0. Then use a low-pass filter on the data which clips off everything above ω_0. Again, with limited sample size ω_0 may be difficult to estimate. Further, the $\{Y_n\}$ process may intrinsically contain arbitrarily high frequencies.

The procedure we suggest is as follows: Do a preliminary rough smoothing of the data. Use this to get an estimate of the expected mean square values of some of the derivatives of the $\{Y_n\}$ process. For instance, letting Δ be the first difference operator, get some estimate for $E|\Delta Y_n|^2$. Suppose that estimate is β_1^2. Then the problem can be formulated as: Find the filter which is minimax against all $F(d\lambda)$ satisfying

$$E|\Delta Y_n|^2 = \int_{(-\pi,\pi]} |e^{i(n+1)\lambda} - e^{in\lambda}|^2 F(d\lambda) \leqq \beta_1^2$$

or

$$\int_{(-\pi,\pi]} |e^{i\lambda} - 1|^2 F(d\lambda) \leqq \beta_1^2 .$$

There is no necessity to work only with first differences. We might wish to estimate the mean square of the second difference operator and minimax over all processes satisfying

$$E|\Delta^2 Y_n|^2 \leqq \beta_2^2 .$$

Or perhaps we want to use both pieces of information at once and minimax over processes satisfying

$$E|\Delta Y_n|^2 \leqq \beta_1^2 , \qquad E|\Delta^2 Y_n|^2 \leqq \beta_2^2$$

simultaneously.

Notice that $E|\Delta Y_n|^2$ and $E|\Delta^2 Y_n|^2$ are simple linear functions of the variance and first two autocovariances of the $\{Y_n\}$ process. Thus, (assuming the autocovariance of the $\{Z_n\}$ process known) these expected squared differences can be estimated by using the appropriate linear combinations of the standard estimators of the autocovariances of the $\{X_n\}$ process. However, this procedure is usually very inefficient.

Another question is how many differences $E|\Delta^k Y_n|^2$ should be estimated and used in the procedure outlined above. The answer depends on two considerations: First, how many can be easily and accurately estimated? We have used first, and sometimes second differences also—but never more. Second, how large a filtering error is bearable? The more restricted the class of spectral distributions for $\{Y_n\}$, the smaller the resulting minimax mean square filtering error will be. This error can be computed from the formulas that follow, and if it is not acceptable, more restrictions can be imposed.

Assuming that the bounds defining the set $\mathscr{F}$ can be accurately estimated, the procedure we propose gives a guaranteed upper bound on the RMS filtering error. No other method we know of can give the same guarantee.

3. The minimax filter. Write

$$Y_n = \int e^{i\lambda n} Y(d\lambda)\,, \qquad Z_n = \int e^{i\lambda n} Z(d\lambda)$$

where all integrals without limits will be over the range $[-\pi, \pi)$. Then $E|Y(d\lambda)|^2 = F(d\lambda)$, and $E|Z(d\lambda)|^2 = G(d\lambda)$. Denote

$$\phi(\lambda) = \sum e^{-ik\lambda} a_k$$

so that

$$\begin{aligned} E(Y_0 - \textstyle\sum a_{-k} X_k)^2 &= E|\int Y(d\lambda) - \int \phi(\lambda) Y(d\lambda) - \int \phi(\lambda) Z(d\lambda)|^2 \\ &= E|\int (1 - \phi(\lambda)) Y(d\lambda) - \int \phi(\lambda) Z(d\lambda)|^2\,. \end{aligned}$$

Therefore, we state the problem: Find the ϕ which minimizes the maximum over $\mathscr{F}$ of

$$I(\phi, F) = [\int |1 - \phi(\lambda)|^2 F(d\lambda) + \int |\phi(\lambda)|^2 G(d\lambda)]\,.$$

Consider the class S_l of all continuous functions $\theta(\lambda) \geqq 0$ such that

$$\max_{F \in \mathscr{F}} \int \theta(\lambda) F(d\lambda) \leqq l\,.$$

To characterize this class define the set B which is the convex hull of $f_1, \cdots, f_n$. That is

$$B = \{\textstyle\sum_1^n \alpha_m f_m;\ \alpha_m \geqq 0,\ \sum_1^n a_m = 1\}\,.$$

We assert:

PROPOSITION. *S_l consists of those continuous functions $\theta(\lambda) \geqq 0$ such that there is an $f \in B$ satisfying*

$$\theta(\lambda) \leqq l f(\lambda)\,, \qquad \textit{all} \quad \lambda\,.$$

The proof of this is deferred while we point out its use. Consider all $\phi(\lambda)$ such that $|1 - \phi|^2 \in S_l$. Call this class R_l. Then for ϕ in R_l

$$I(\phi, F) \leqq l + \int |\phi|^2\, dG\,.$$

For the ϕ that minimizes $\max_{F \in \mathscr{F}} I(\phi, F)$, $|1 - \phi|^2$ is in some S_l, $l \geqq 0$. Hence

$$\min_\phi \max_{F \in \mathscr{F}} I(\phi, F) = \min_l \min_{\phi \in R_l} (l + \int |\phi|^2\, dG)\,.$$

If the minimizing value of l is l^*, then the minimax ϕ is the $\phi \in R_{l^*}$ which minimizes $\int |\phi|^2\, dG$. Put $\gamma(\lambda) = 1 - \phi(\lambda)$. We now minimize

$$\int |\phi|^2\, dG = \int |1 - \gamma(\lambda)|^2\, dG$$

over all ϕ in R_l. The way to do this is clear: Look at any $f(\lambda) \in B$. We will first minimize over all functions $\gamma(\lambda)$ satisfying $|\gamma(\lambda)| \leqq (lf(\lambda))^{\frac{1}{2}}$. It is not hard to see what the minimizing $\gamma(\lambda)$ is: If $lf(\lambda) \geqq 1$, put $\gamma(\lambda) = 1$. If $lf(\lambda) \leqq 1$, put $\gamma(\lambda) = (lf(\lambda))^{\frac{1}{2}}$. Hence

$$\min_{\phi \in R_l} \int |\phi|^2 \, dG = \min_{f \in B} \int [(1 - (lf(\lambda))^{\frac{1}{2}})^+]^2 \, dG .$$

But now the problem is reduced to an ordinary minimization problem over the $n + 1$ variables $l, \alpha_1, \cdots, \alpha_n$. That is, minimize the expression

$$l + \int [(1 - (l \sum \alpha_m f_m(\lambda))^{\frac{1}{2}})^+]^2 \, dG(\lambda)$$

over $l, \alpha_1, \cdots, \alpha_n$ subject to $\alpha_m \geqq 0$, $m = 1, \cdots, n$ and $\sum \alpha_m = 1$. At this point we consider the problem solved.

The proposition above looks as through it should surely be in the literature. But we cannot find it, so we give a proof.

LEMMA. *Let B be a compact convex set of continuous functions and g a continuous function such that there exists no $f \in B$ with $f \geqq g$ everywhere. Then there exists a positive finite measure μ such that*

$$\int g \, d\mu > \sup_{f \in B} \int f \, d\mu .$$

PROOF. This is more or less a standard result, but since we cannot locate a reference in the form cited, we give a proof. Take

$$A = \{f - g; f \in B\} .$$

Let D be the cone of nonnegative functions. Then D is closed and convex. A is closed, convex, and compact. A standard result ([1] page 417) yields the fact that if A and D are disjoint, there exists a finite measure μ and a constant α such that

$$\text{(A)} \qquad \int h \, d\mu > \alpha , \qquad \text{all} \quad h \in D$$

and

$$\text{(B)} \qquad \int (f - g) \, d\mu < \alpha , \qquad \text{all} \quad f \in B .$$

Furthermore, we must have that $\int h \, d\mu \geqq 0$, all $h \in D$. Otherwise if $\int h \, d\mu = -\delta$, then for the function $nh(\lambda)$

$$\int (nh) \, d\mu = -n\delta .$$

This implies that μ is a positive measure and $\alpha < 0$. Hence, for any $f \in B$,

$$\int f \, d\mu < \int g \, d\mu + \alpha$$

which implies

$$\sup_{f \in B} \int f \, d\mu \leqq \int g \, d\mu + \alpha .$$

COROLLARY. *Let C be a compact and convex set of continuous functions. Denote*

$$C^+ = \{\mu; \mu \geqq 0, \sup_{f \in C} \int f \, d\mu \leqq 1\} .$$

Then if g is continuous and satisfies

$$\sup_{\mu \in C^+} \int g \, d\mu \leqq 1 ,$$

there exists an $f \in C$ such that $g \leqq f$ everywhere.

PROOF. Suppose not; then by the lemma there exists a positive measure μ such that

$$\int g\, d\mu > \sup_{f \in C} \int f\, d\mu .$$

Renormalize μ so that

$$\sup_{f \in C} \int f\, d\mu = 1 .$$

Then this renormalized μ is in C^+ but we get the contradiction $\int g\, d\mu > 1$.

Now a quick comparison shows that this corollary is the proposition we wanted to prove in a thinly veiled form. The solution could also have been obtained using, instead, the fact that $I(\phi, F)$ has its maximum on one of the extreme points of $\mathscr{F}$. This approach leads to a more complicated computation.

REFERENCE

[1] DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators, Part 1.* Interscience, New York.

21540 SUMMIT TRAIL
TOPANGA, CALIFORNIA 90290