AD-A181 686

Scientific and
Engineering
Studies

Compiled 1983

Signal
Processing
Studies

A. H. Nuttall

NAVAL UNDERWATER SYSTEMS CENTER

Scientific and Engineering Studies

Compiled 1983

Signal Processing Studies

A. H. Nuttall

Foreword

This collection of technical reports and technical memoranda deals with the following topics: spectral analysis via combined temporal and lag weighting; programs for and performance of multi-channel linear predictive spectral estimators; direct coherence estimation via a fast algorithm; windowing with a two-parameter class of Bessel weightings, with applications to arrays in any number of dimensions; performance of robust methods of estimating signal strengths in erratic environments; exact receiver operating characteristics for a nonlinear system with quantizers, or-ing, and accumulation; and characterization of the probability distribution of measured data. Graphs of the results are presented, as well as programs that enable a user to extend the results to his particular application.

Some of the material presented here is based heavily on earlier work by the author, which can be found in the following volumes in addition to the referenced technical reports:

Performance of Detection and Communication Systems,
NUSC Scientific and Engineering Studies, 1974;

Spectral Estimation, NUSC Scientific and Engineering Studies, 1977;

Coherence Estimation, NUSC Scientific and Engineering Studies, 1979;

Receiver Performance Evaluation and Spectral Analysis,
NUSC Scientific and Engineering Studies, 1981.

Dr. William A. Von Winkle
Associate Technical Director for Technology
NAVAL UNDERWATER SYSTEMS CENTER

Compiled 1983

Table of Contents

Table of Contents (cont'd)

Spectral Analysis Via Quadratic Frequency- Smoothing of Fourier- Transformed, Overlapped, Weighted Data Segments

A. H. Nuttall

ABSTRACT

A generalized framework for spectral analysis is presented, which includes as special cases the Blackman-Tukey technique and the weighted overlapped segment-averaging FFT technique. The general method is analyzed in terms of the mean and variance of the spectral estimate, thereby revealing the fundamental dependence of its performance on the temporal weighting, lag weighting, amount of overlap, number of pieces, available data record length, and frequency resolution. To enable a fair tradeoff study and comparison between many different special cases of the technique, it is demanded that the spectral analysis technique achieve a specified frequency resolution with the given data record length. This necessitates a detailed investigation of the windowing capabilities of the temporal and lag weightings, their interaction, and the definition of an overall effective weighting and window. The possibility of using lag-reshaping to achieve desirable effective windows is considered and found to be reasonable for a wide variety of windows with good side lobe behavior and decay rates.

Results for the variance of the spectral estimate for rectangular temporal weighting indicate that if the length of the temporal weighting is selected to be somewhat larger than the length of the lag weighting, the variance is at a near minimum. Furthermore, in this situation, the possibly deleterious side lobes of the temporal weighting can be compensated by proper choice of lag weighting, resulting in low side lobes and good decay of the overall effective spectral window. For Hanning temporal weighting, the lengths of the temporal and lag weighting should be approximately equal for minimum variance of the spectral estimate.

Table of Contents

List of Illustrations

List of Tables

List of Symbols

T	available record length in seconds
B_e	effective frequency resolution in hertz (eq. 1)
t	time
$x(t)$	data record
f	frequency
$G(f)$	true (unknown) spectrum of $x(t)$
$W_0(f)$	spectral window
$H(f)$	filter voltage transfer function
$\hat{P}$	estimate of power (eq. 2)
$g(t)$	gate function (eq. 3)
Q	quality ratio (eq. 4)
$Av(\hat{P})$	average value of random variable $\hat{P}$
$Var(\hat{P})$	variance of random variable $\hat{P}$
overbar	ensemble average
τ	lag (delay) variable
$R_z(\tau)$	correlation of random process $z(t)$
$\phi_g(\tau)$	correlation of function $g(t)$ (eq. 9)
$w_1(t)$	temporal weighting (figure 2 and eq. 13)
$W_1(f)$	temporal window (Fourier transform of $w_1(t)$)
L_1	duration of $w_1(t)$; length of $\phi_1(\tau)$ (figures 2 and 4)
S	time shift of temporal weightings (eq. 13)
P	number of data pieces in first-stage spectral estimate (eqs. 13 and 14)
$y_p(t)$	p-th weighted data segment (eq. 13)
$\hat{G}_1(f)$	first-stage spectral estimate (eq. 14)
$\hat{R}_1(\tau)$	first-stage correlation estimate (eq. 16)
$W_2(f)$	lag window (eq. 18)
$w_2(\tau)$	lag weighting (eq. 20)
L_2	length of $w_2(\tau)$ (figure 3 and eq. 21)
$\hat{G}_2(f)$	second-stage spectral estimate (eq. 18)
$\hat{R}_2(\tau)$	second-stage correlation estimate (eq. 19)
Δ_t	time spacing in discrete samples of $x(t)$ (eq. 24) = time spacing of impulsive temporal weighting $w_1(t)$
N_1	number of impulses in discrete $w_1(t)$ (eq. 24)
w_{1m}	m-th sample weight in discrete $w_1(t)$ (eq. 24)
y_{pm}	m-th weighted sample of p-th data piece (eq. 26)

List of Symbols (Cont'd)

$\hat{R}_{1k}$	first-stage correlation estimate (eq. 28)
$\hat{R}_{k}^{(p)}$	k-th correlation estimate of p-th data piece (eq. 29)
Δ_F	frequency increment used for calculation of spectral estimates (eq. 32)
N	size of FFT (power of 2) (eq. 33)
$\hat{R}_{2k}$	second-stage correlation estimate (eq. 36)
Δ_f	frequency spacing in discrete samples of $\hat{G}_1(f)$ (eq. 38) = frequency spacing of impulsive lag window $W_2(f)$
N_2	number of impulses in discrete $W_2(f)$ (eq. 38)
W_{2n}	n-th sample weight in discrete $W_2(f)$ (eq. 38)
$\phi_1(\tau)$	correlation of temporal weighting $w_1(t)$ (eq. 44)
$w_e(\tau)$	effective weighting of generalized spectral analysis technique (eq. 48)
$W_e(f)$	effective window of generalized spectral analysis technique (eq. 51)
L_m	minimum of L_1, L_2 (eqs. 54 and 55)
$c\{g\}$	shape factor of function g (eq. 57)
C5,C3,C1	three weighting functions (eq. 60)
B_d	desired effective frequency resolution
$w_d(\tau)$	desired effective weighting (eq. 98)
$W_d(f)$	desired effective window (eq. 100)
$\gamma(x,y)$	window convolution function (eq. 105)
$Q_P(u)$	periodic function (eq. 106)
$\phi_3(\tau,\mu)$	third-order correlation of $w_1(t)$ (eq. 108)
NQR	normalized quality ratio (eq. 110)
$\phi_2(\tau)$	correlation of $w_1^2(t)$ (eq. 113)
$q = q\{w_1\}$	shift-fraction (eq. 139)
P_{max}	maximum value of P (eq. 141)

Spectral Analysis via Quadratic Frequency-Smoothing of Fourier-Transformed, Overlapped, Weighted Data Segments

Introduction

Spectral analysis techniques have received a great deal of attention in the past (references 1-12), ranging from the original autocorrelation approach of Blackman-Tukey (reference 2) to the more recent weighted, overlapped, segment-averaging FFT approach (references 7-12). These two apparently disparate approaches are shown here to be limiting special cases of a generalized framework for spectral analysis; thus consideration of this general technique elucidates the fundamental behavior and performance of a rather wide variety of spectral approaches and their tradeoffs. This generalized framework has already been presented in references 13-15, where a brief summary of some of the main features has been mentioned. Additionally, some of the analytical results to be presented here were alluded to there; however, none of the detailed derivations or quantitative results in this report were given at that time.

There are two fundamental parameters that critically affect the performance of any spectral estimation technique. They are the available record length, T, of the stationary random process under investigation, and the effective frequency resolution, B_e, of the technique under consideration. We would like to be able to attain fine resolution (small B_e) with short data lengths and storage (small T); however, stable results (small fluctuations) are achievable only if the product TB_e is much larger than unity. The problems we address are how to make optimum use of a given *limited* amount of data in order to realize a specified desired resolution with maximum stability, and to determine what tradeoffs are available regarding windowing and weighting at different stages of the spectral analysis procedure. It is assumed that the reader is familiar with the tradeoffs presented in reference 9 for the weighted, overlapped, segment-averaging FFT procedure.

The generalized framework for spectral analysis that is presented here is capable of a wide variety of forms in addition to the Blackman-Tukey and FFT approaches mentioned above. In order to compare these various forms with each other on a reasonable basis, it is required that each analysis technique realize the same effective resolution bandwidth, B_e, and that they all utilize the same data record length T. Without these reasonable constraints, valid conclusions about relative performances of different techniques are tentative at best. This insistence upon equal effective frequency resolution necessitates a rather detailed investigation of the effects of the weightings and windows employed in the generalized framework and their allowed durations. The desirability of an overall effective window for spectral analysis with low side lobes and good decay is achievable only through careful choice of the combined weightings. The constraint upon the effective frequency resolution naturally also shows up in the analysis of the variance of the spectral estimation technique, as well as in its average value, leading to some numerical analysis complications; nevertheless, it is believed to be the proper basis of comparison and is maintained throughout.

The two major statistical parts of this report deal with the mean and the variance of the spectral estimate. The result for the average value leads to the definition of the effective window of the generalized spectral analysis technique, in terms of the temporal and lag windows. The variance result incorporates, additionally, the amount of overlap, the number of data pieces, and the ambiguity functions of the temporal and lag windows; the complexity of the latter results debilitates easy interpretation and it has been found necessary to resort to numerical evaluation of the variance, for practical cases of interest.

Ultimate Stability Attainable From a Given Record Length

Suppose a stationary (complex) data record x(t) of length T seconds is available, and that we wish to estimate its power density spectrum* $G(f)$ with an effective frequency resolution of B_e Hz, where $W_o(f)$ is the narrowband window through which the power density spectrum is to be observed. These two frequency-domain quantities are related according to†

$$B_e = \frac{\left[\int df \ W_o(f)\right]^2}{\int df \ W_o^2(f)} \ .$$

(1)

This bandwidth measure, B_e, is called the statistical bandwidth of $W_o(f)$ in reference 5, page 265. The relation of effective bandwidth B_e to half-power bandwidth B_h is considered in appendix A; it is shown that for good windows, the ratio of the two bandwidths is relatively independent of the exact window shape. Thus it is possible to translate results to other bandwidth measures without significantly affecting the essential quantitative aspects.

If we take the original data record and pass it through a narrowband linear (complex) filter with power transfer function equal to the window, $|H(f)|^2 = W_o(f)$, and which is centered at a frequency, f_o, of interest, we will have lost no relevant information about the process in the frequency band of interest, because we have filtered out information of no use. We can now estimate the power in the narrowband filter output process and use it as a measure of the spectrum of the input process in the neighborhood of frequency f_o. See figure 1.
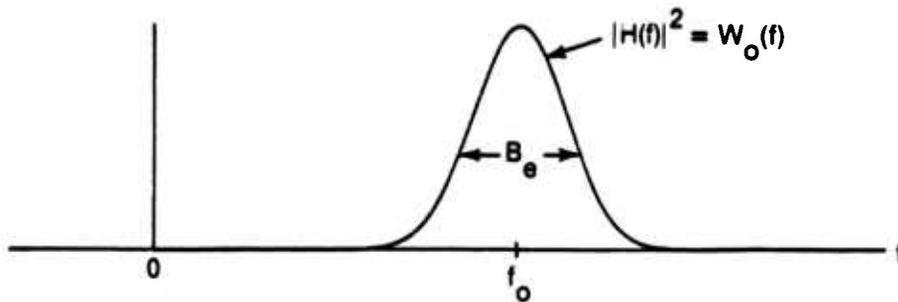
Figure 1. Power Transfer Function of Narrowband Linear Filter

Let z(t) be the complex output process from the narrowband filter when excited by the available T seconds of data x(t). If we ignore a starting transient (i.e., assuming $T \gg 1/B_e$), the filter output power estimate in the band of width B_e is

$$\hat{P} = \frac{1}{T} \int_T dt \ |z(t)|^2 = \int dt \ g(t) \ |z(t)|^2 \ ,$$

(2)

---

*For brevity, we use the term spectrum rather than autospectrum in this report.

†Integrals without limits are over the range of the nonzero integrands.

3

where gate function

$$g(t) = \begin{cases} 1/T \text{ for } t \in T \\ 0 \text{ otherwise} \end{cases} .$$

(3)

The measure of stability we adopt for this estimator, and for the others to follow, is the quality ratio defined as

$$Q \equiv \frac{\text{Var}(\hat{P})}{\text{Av}^2(\hat{P})} = \frac{\overline{\hat{P}^2} - \overline{\hat{P}}^2}{\overline{\hat{P}}^2} ,$$

(4)

where $\text{Av}(\hat{P})$ and $\text{Var}(\hat{P})$ denote the average value and variance of $\hat{P}$, respectively, and an overbar denotes an ensemble average. We have average value

$$\text{Av}(\hat{P}) = \int dt \, g(t) \, \overline{|z(t)|^2} = \overline{|z(t)|^2} = R_z(0) = \int df \, G(f) \, |H(f)|^2$$

$$\approx G(f_o) \int df \, |H(f)|^2 ,$$

(5)

assuming that filter-input spectrum $G(f)$ does not vary quickly with respect to B_e, in the neighborhood of f_o. $R_z(\tau)$ is the correlation* of filter output process $z(t)$.

Also, we have mean square value

$$\overline{\hat{P}^2} = \iint dt \, du \, g(t) \, g(u) \, \overline{|z(t)|^2 \, |z(u)|^2} .$$

(6)

Now in the interval T, filter output $z(t)$ will be approximately a stationary zero-mean, complex, analytic Gaussian process for small B_e; filter $H(f)$ has filtered out zero and all negative frequencies. Then fourth-order moment

$$\overline{z(t)z^*(t)z(u)z^*(u)} \approx R_z^2(0) + |R_z(t - u)|^2 .$$

(7)

There follows from (4) and (6),

$$\text{Var}(\hat{P}) = \iint dt \, du \, g(t)g(u) \, |R_z(t - u)|^2 = \int d\tau \, \phi_g(\tau) \, |R_z(\tau)|^2 ,$$

(8)

where gate-correlation† of function $g(t)$ is

$$\phi_g(\tau) \equiv \int dt \, g(t)g(t - \tau) .$$

(9)

---

*For brevity, we use the term correlation instead of autocorrelation in this report.
†For stationary processes, we let R denote the ensemble-average correlation, whereas for aperiodic nonrandom functions, we let ϕ denote the integral correlation; see (5) and (9).

Since the gate-correlation $\phi_g(\tau)$ extends over $\pm T$, while process correlation $R_z(\tau)$ extends only over approximately $\pm 3/B_e$, we have, via Parseval's theorem and for $TB_e \gg 1$,

$$\text{Var}(\hat{\beta}) \approx \phi_g(0) \int d\tau \left| R_z(\tau) \right|^2 = \frac{1}{T} \int df\ G_z^2(f)$$

$$= \frac{1}{T} \int df\ \left| H(f) \right|^4 G^2(f) \approx \frac{1}{T}\ G^2(f_o) \int df\ \left| H(f) \right|^4 \quad . \tag{10}$$

The quality-ratio measure of stability is then, from (4), (5), (10), and (1),

$$Q = \frac{1}{T}\ \frac{\int df \left| H(f) \right|^4}{\left[\int df \left| H(f) \right|^2 \right]^2} = \frac{1}{T}\ \frac{\int df\ W_o^2(f)}{\left[\int df\ W_o(f) \right]^2} = \frac{1}{TB_e} \quad . \tag{11}$$

This is the limiting (smallest) value of Q for specified frequency resolution B_e and available record length T when $TB_e \gg 1$. No other spectral procedures can improve on it; they can merely approximate it. As such, (11) is the benchmark against which other procedures can be compared, under the condition that T and B_e are equal to those values for the various procedures under consideration.

The normalized quality ratio is defined as $Q \cdot TB_e$. Thus the normalized quality ratio can never be smaller than unity, which value can only be approached for large TB_e through proper processing techniques.

Description of Spectral Analysis Technique

We begin by defining a temporal weighting function $w_1(t)$ of finite duration L_1; that is,

$$w_1(t) \neq 0 \quad \text{only for} \quad |t| < L_1/2 \quad . \tag{12}$$

As shown in figure 2, temporal weighting $w_1(t)$ is real, even, and peaked at the origin. Although this presentation is couched in terms of continuous functions, we shall show shortly that it includes discrete digital processing as a special case.
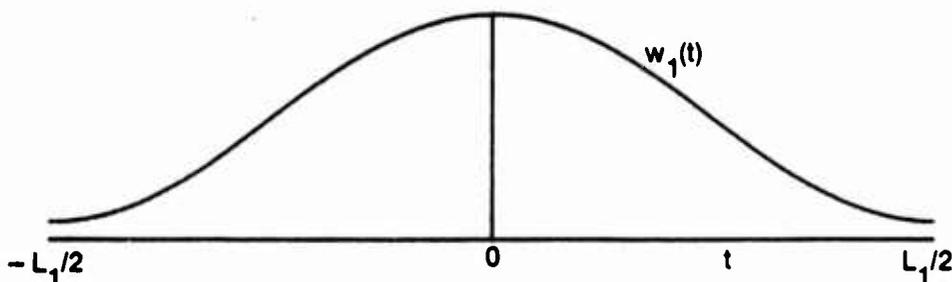
Figure 2. Temporal Weighting $w_1(t)$

The available data record is $x(t)$ for $0 < t < T$; this (complex) random process is presumed second-order stationary in that observation interval. We shift the temporal weighting by $L_1/2 + pS$ and multiply it by $x(t)$ to generate the p-th piece of weighted data:

$$y_p(t) \equiv x(t) \; w_1\left(t - \frac{L_1}{2} - pS\right) \quad \text{for} \quad 0 \leq p \leq P - 1 \quad . \tag{13}$$

Here p is an integer; if shift $S < L_1$, then $y_p(t)$ and $y_{p+1}(t)$ will overlap on the t-axis.

The first-stage power density spectral estimate at frequency f is obtained by averaging the magnitude-squared value of the Fourier transform of data piece $y_p(t)$, over a total of P pieces:

$$\hat{G}_1(f) = \frac{1}{P} \sum_{p=0}^{P-1} \left| \int dt \; \exp(-i2\pi ft) \; y_p(t) \right|^2 \quad \text{for any} \quad f \quad . \tag{14}$$

This procedure is the same as that considered in reference 7 and in reference 9, eqs. (2) and (3). Since $x(t)$ is available only for $0 < t < T$, we prevent the weighting in (13) from extending beyond that interval; mathematically this means that we must have

$$L_1 + (P - 1) S \leq T \quad . \tag{15}$$

An alternative interpretation of (14) is very illuminating. We define the inverse Fourier transform of (14) as the first-stage (auto)correlation estimate; there follows immediately at delay τ,

$$\hat{R}_1(\tau) = \int df \; \exp(i2\pi f\tau) \; \hat{G}_1(f)$$

$$= \frac{1}{P} \sum_{p=0}^{P-1} \int dt \; y_p(t) \; y_p^*(t - \tau) \quad \text{for all } \tau \; , \tag{16}$$

where we have allowed random processes $x(t)$ and $y_p(t)$ to be complex. This is recognized as the average of the sample correlations that can be formed at delay τ, from each of the P pieces of weighted data in (13). Since temporal weighting $w_1(t)$ is zero for $|t| > L_1/2$ according to (12), we see from (13) and (16) that

$$\hat{R}_1(\tau) = 0 \quad \text{for } |\tau| > L_1 \; . \tag{17}$$

The parameter, τ, is called the lag domain variable, because of the way it appears as a delayed time in (16). Equation (16) (and those to follow) is true for all τ. Both sides of (16) are zero over most of the range of τ; nevertheless, it is mathematically convenient to employ the equality of both sides of (16) for all τ in various transformations below.

The second-stage power density spectral estimate is defined as a frequency-smoothed version of the first-stage result:

$$\hat{G}_2(f) = \hat{G}_1(f) \circledast W_2(f) = \int du \; \hat{G}_1(u) W_2(f - u) \quad , \tag{18}$$

where $\circledast$ denotes convolution. This is termed quadratic smoothing since it is done in terms of power quantities rather than voltages. Equation (18) is the desired output from the generalized spectral analysis technique considered here. $W_2(f)$ is called the lag window, for reasons to be given below. The equivalent statement to (18) in the lag domain is obtained by Fourier transforming (18); the second-stage correlation is

$$\hat{R}_2(\tau) = \int df \; \exp(i2\pi f\tau) \; \hat{G}_2(f) = \hat{R}_1(\tau) \; w_2(\tau) \quad , \tag{19}$$

where we used (18) and (16) and defined the Fourier transform pair

$$w_2(\tau) = \int df \; \exp(i2\pi f\tau) \; W_2(f) \quad ,$$

$$W_2(f) = \int d\tau \; \exp(-i2\pi f\tau) \; w_2(\tau) \quad . \tag{20}$$

$w_2(\tau)$ and $W_2(f)$ are both real, even, and peaked at their origins. Since $w_2(\tau)$ appears multiplicatively in (19), it is called the lag weighting; its transform $W_2(f)$ is the lag window. The convention adopted throughout this report is that multiplication by a function in the t or τ domains is called a weighting; the counterpart to this operation in the Fourier transform domain (frequency f domain) is convolution and is called windowing.

We shall let lag weighting $w_2(\tau)$ be of duration $2L_2$; that is,

$$w_2(\tau) = 0 \quad \text{for} \quad |\tau| > L_2 \quad . \tag{21}$$

A typical plot is shown in figure 3; the reason for the apparent notational discrepancy between the lengths in figures 2 and 3 will become clear when the lag-domain counterpart of temporal weighting $w_1(t)$ is encountered later.
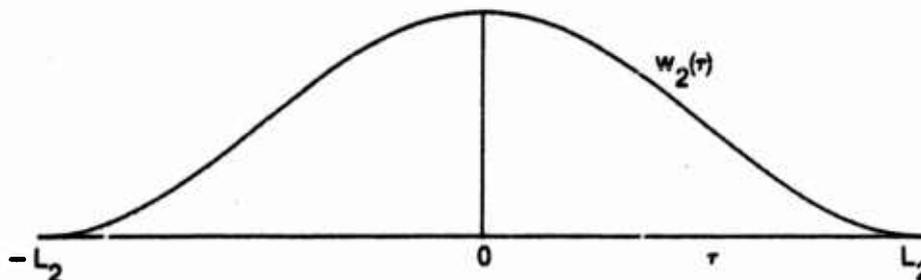
Figure 3. Lag Weighting $w_2(\tau)$

We have already observed that $\hat{R}_1(\tau)$ is zero for $|\tau| > L_1$ in (17). Therefore, it follows from (19) and (21) that

$$\hat{R}_2(\tau) = 0 \quad \text{for} \quad |\tau| > \min(L_1, L_2) \quad . \tag{22}$$

However, although we must have temporal length $L_1 < T$ (from (15) for $P = 1$), there is no restriction on L_2. We could have L_2 larger than L_1 and T; this would simply mean that we would be lag-weighting some zero estimates of $\hat{R}_1(\tau)$ in (19) for the larger values of $|\tau|$. Also there are no constraints such as realizeability on the lag weighting or window.

For example, the special case of no quadratic frequency-smoothing corresponds to

$$W_2(f) = \delta(f), \ w_2(\tau) = 1, \ L_2 = \infty \quad \text{for no smoothing,} \tag{23}$$

for which (18) yields $\hat{G}_2(f) = \hat{G}_1(f)$. Thus we have our standard first-stage spectral estimate (14) as a limiting case of the generalized spectral analysis technique. On the other hand, if lag window $W_2(f)$ were broad (small L_2), there would be a significant amount of smoothing taking place in the band about $u = f$ in (18) where window W_2 is non zero.

There is no inherent limitation on the relative sizes of L_1 and L_2 as yet: L_2 can be chosen as large as desired, while L_1 is subject to the upper bound T. However, when we specify the overall effective frequency resolution of the generalized technique, a relation between L_1 and L_2 will ensue.

Another important special case of the generalized spectral analysis technique is afforded by $P = 1$, $w_1(t) = 1$ for $|t| < L_1/2$, and $L_1 = T$. Then (16) and (13) indicate that $\hat{R}_1(\tau)$ is simply the sample correlation of the available data $x(t)$ of length T, while $\hat{R}_2(\tau)$ in (19) is a weighted version of $\hat{R}_1(\tau)$ for $|\tau| < L_2$. But this is precisely the Blackman-Tukey approach described in reference 2; the choice of lag weighting $w_2(\tau)$ and its length is fully discussed there. For example, if $w_2(\tau) = (T-|\tau|)^{-1}$ for $|\tau| < L_2 < T$, then $\hat{R}_2(\tau)$ is an unbiased estimator for $|\tau| < L_2$; see reference 2, page 11.

For $P>1$ and general temporal weighting $w_1(t)$, lag weighting $w_2(\tau)$, and overlap, a wide variety of processors is possible via the generalized framework set up above. How should the two weightings be traded off against each other? Can the deleterious effects of a poor or preselected temporal weighting be undone by proper choice of lag weighting? Recall that none of these techniques can hope to better the quality-ratio result (11), but hopefully, some can do as well, with less computational effort and storage.

A related procedure to the one presented here has been given in references 16 and 17. However, neither incorporate overlapping, and the fundamental tradeoffs between the temporal and lag weightings were not studied. Furthermore, the only frequency-smoothing case considered was a rectangular boxcar, which severely limits the potential of the technique; some advantages of the generalized technique considered here will become apparent at a later stage. For the time being, we observe that side lobe control will be realized by a mixture of temporal weighting and lag weighting (frequency smoothing), while stability will be achieved by a combination of segment averaging and frequency smoothing (lag weighting).

Discrete-Time Processing

All the functions above have been tacitly assumed no worse than discontinuous; see figures 2 and 3 for example. However, there is nothing in the above mathematics which precludes impulsive behavior. For example, suppose the temporal weighting is a sum of N_1 equispaced impulses:

$$w_1(t) = \Delta_t \sum_m w_{1m} \, \delta(t - m\Delta_t) \quad , \tag{24}$$

where $\{w_{1m}\}$ is a finite length, real sequence, symmetric about $m = 0$; this corresponds to discrete sampling of waveform $x(t)$ at time spacing Δ_t. The p th piece of weighted data is, from (13),

$$y_p(t) = x(t) \, \Delta_t \sum_m w_{1m} \, \delta\left(t - \frac{L_1}{2} - pS - m\Delta_t\right)$$

$$= \Delta_t \sum_m y_{pm} \, \delta\left(t - \frac{L_1}{2} - pS - m\Delta_t\right) \quad , \tag{25}$$

where weighted sample

$$y_{pm} \equiv w_{1m} \, x\left(\frac{L_1}{2} + pS + m\Delta_t\right) \quad . \tag{26}$$

The first-stage correlation estimate, (16), becomes

$$\hat{R}_1(\tau) = \Delta_t \sum_k \hat{R}_{1k} \, \delta(\tau - k\Delta_t) \quad , \tag{27}$$

where the area of each impulse is given by

$$\hat{R}_{1k} = \frac{1}{P} \sum_{p=0}^{P-1} \hat{R}_k^{(p)} \tag{28}$$

and

$$\hat{R}_k^{(p)} = \Delta_t \sum_m y_{p,m} \, y_{p,m-k}^* \quad . \tag{29}$$

The last quantity is the sample correlation of the p-th set of samples, and $\hat{R}_{1k}$ is their average over the total of P pieces.

The first-stage spectral estimate is the Fourier transform of (27) as usual:

$$\hat{G}_1(f) = \Delta_t \sum_k \hat{R}_{1k} \, \exp(-i2\pi f k \Delta_t) \quad , \tag{30}$$

which is finite for all f and is of period $1/\Delta_t$ in f. An alternative expression is available by substitution of (25) in (14):

$$\hat{G}_1(f) = \frac{1}{P} \sum_{p=0}^{P-1} \left| \Delta_t \sum_m y_{pm} \, \exp(-i2\pi f m \Delta_t) \right|^2 \quad . \tag{31}$$

These two expressions hold for arbitrary f; either one can be used to obtain the first-stage spectral estimate. If we restrict our calculations of interest to multiples of some frequency increment Δ_F, (31), for example, specializes to

$$\hat{G}_1(q\Delta_F) = \frac{1}{P} \sum_{p=0}^{P-1} \left| \Delta_t \sum_m y_{pm} \, \exp(-i2\pi m q \Delta_t \Delta_F) \right|^2 \quad , \tag{32}$$

where q is an integer. At this point, there needn't be any relation between Δ_t and Δ_F; we can calculate the spectral estimate at any frequencies we please. However, a favorite choice for computational purposes is to choose frequency increment

$$\Delta_F = \frac{1}{N\Delta_t} \quad , \qquad N = \text{power of 2} \quad , \tag{33}$$

to get the special digital processing result

$$\hat{G}_1\left(\frac{q}{N\Delta_t}\right) = \frac{1}{P} \sum_{p=0}^{P-1} \left| \Delta_t \sum_m y_{pm} \, \exp(-i2\pi m q/N) \right|^2 \quad , \tag{34}$$

which is recognized as the familiar power average of N-point FFTs of weighted data sequences. All the impulsive functions in (24)-(27) have dropped out of first-stage spectral estimates (30)-(34).

The temporal window associated with impulsive weighting (24) is its Fourier transform

$$W_1(f) = \Delta_t \sum_m w_{1m} \exp(-i2\pi fm\Delta_t) \quad . \tag{35}$$

Having picked an impulsive temporal weighting $w_1(t)$, we are still free to select the lag weighting or lag window as we please. For example, for any lag weighting function $w_2(\tau)$, (19) and (27) yield second-stage correlation estimate

$$\hat{R}_2(\tau) = \Delta_t \sum_k w_2(k\Delta_t) \hat{R}_{1k} \delta(\tau - k\Delta_t) \equiv \Delta_t \sum_k \hat{R}_{2k} \delta(\tau - k\Delta_t) \quad . \tag{36}$$

The corresponding second-stage spectral estimate is the Fourier transform

$$\hat{G}_2(f) = \Delta_t \sum_k w_2(k\Delta_t) \hat{R}_{1k} \exp(-i2\pi fk\Delta_t) = \Delta_t \sum_k \hat{R}_{2k} \exp(-i2\pi fk\Delta_t) \quad , \tag{37}$$

which is everywhere finite and has period $1/\Delta_t$ in f. Evaluation of (37) can therefore be confined to $|f| < (2\Delta_t)^{-1}$.

These results apply for general lag weighting. A specific choice is the lag window with N_2 equispaced nonzero impulses:

$$W_2(f) = \Delta_f \sum_n W_{2n} \delta(f - n\Delta_f) \quad . \tag{38}$$

Frequency spacing Δ_f need not be related to time spacing Δ_t in (24), nor to frequency increment Δ_F used in the frequency and FFT calculations above in (32)-(34). Also there are no relations between the real symmetric sets of numbers $\{w_{1m}\}$ in (24) and $\{W_{2n}\}$ in (38). Substitution of (38) in (18) yields for the second-stage spectral estimate

$$\hat{G}_2(f) = \Delta_f \sum_n W_{2n} \hat{G}_1(f - n\Delta_f) \quad , \tag{39}$$

which is a local average (of the first-stage estimates) in the band about the frequency, f, of interest. Equation (39) is a discrete, quadratic, frequency-smoothing operation. In fact, (39) holds for lag window (38) and *any* temporal weighting $w_1(t)$; it is not limited to the discrete-time form (24).

If we limit our calculations of $\hat{G}_2(f)$ to multiples of frequency increment Δ_F as in (32), then (39) yields

$$\hat{G}_2(q\Delta_F) = \Delta_f \sum_n W_{2n} \hat{G}_1(q\Delta_F - n\Delta_f) \quad ; \tag{40}$$

we can use (30) for $\hat{G}_1$ on the right-hand side. Finally, if we take frequency increment (33) and frequency spacing $\Delta_f = (M\Delta_t)^{-1}$, where integer M is a submultiple of N, and $M\Delta_t$ is of the order of $2L_1$, the FFT results of (34) can be employed in (40). More will be said later on the choice of frequency spacing Δ_f.

The variety of forms available at different stages of the data processing illustrates a great deal of flexibility in exactly how the available data x(t) is processed. For example, one might first evaluate $\hat{G}_1$ via FFT procedure (34). Then, since (30) can be expressed as

$$\hat{G}_1\left(\frac{q}{N\Delta_t}\right) = \Delta_t \sum_k \hat{R}_{1k} \exp(-i2\pi kq/N) \quad , \tag{41}$$

it follows that the complete nonzero portion of correlation sequence $\{\hat{R}_{1k}\}$ is recoverable from the set of numbers $\{\hat{G}_1(\frac{q}{N\Delta_t})\}_0^{N-1}$ if $N > 2N_1-1$, where N_1 is the number of nonzero weights $\{w_{1k}\}$ in (24) (see reference 18). On the other hand, for $N < 2N_1-1$, the inverse FFT of $\{\hat{G}_1(\frac{q}{N\Delta_t})\}_0^{N-1}$ would yield $\hat{R}_{1k}$ only for $|k| < N-N_1$ (reference 18, eq. (15)); thus the central values of $\hat{R}_{1k}$ are recoverable from $\hat{G}_1$. Then second-stage correlation estimate

$$\hat{R}_{2k} = w_2(k\Delta_t) \, \hat{R}_{1k} \tag{42}$$

follows from (36), and the final spectral estimate follows from (37). The lag weighting samples in (42) are arbitrary; thus this is a very general procedure for obtaining estimate $\hat{G}_2(f)$ at any f.

The relations in this subsection hold for arbitrary values of τ, f, and q. However, the functions of τ are impulsive, and are zero outside limited ranges, while the functions of f and q are periodic. These properties should be utilized in any computer processing technique employing these forms. Some further useful properties and interrelationships of the sampled lag weightings and lag windows are presented in appendix B.

Average Value of Spectral Estimate

We now return to the general situation for both the temporal and the lag weightings; that is, we do not presume discrete sampling in time or discrete smoothing in frequency. From (16) and (13), the mean value of the first-stage correlation estimate is

$$Av\{\hat{R}_1(\tau)\} = \frac{1}{P} \sum_{p=0}^{P-1} \int dt\ \overline{x(t)x^*(t-\tau)}\ w_1\left(t - \frac{L_1}{2} - pS\right)$$

$$\bullet\ w_1\left(t - \tau - \frac{L_1}{2} - pS\right) = \overline{x(t)x^*(t-\tau)}\ \phi_1(\tau) = R(\tau)\ \phi_1(\tau) \tag{43}$$

where $R(\tau)$ is the true correlation of stationary process $x(t)$, and where

$$\phi_1(\tau) = \int dt\ w_1(t)\ w_1(t-\tau) \tag{44}$$

will be called the correlation of real temporal weighting $w_1(t)$; see the footnote to (9). We have not presumed process $x(t)$ Gaussian; relation (43) holds for any stationary process $x(t)$.

Since the first-stage spectral estimate $\hat{G}_1(f)$ is a linear operation (Fourier transform) of $\hat{R}_1(\tau)$, the mean value of $\hat{G}_1(f)$ is the Fourier transform of (43); that is,

$$Av\{\hat{G}_1(f)\} = \int d\tau\ \exp(-i2\pi f\tau)\ R(\tau)\ \phi_1(\tau)$$

$$= G(f) \circledast W_1^2(f) = \int du\ G_1(u)\ W_1^2(f-u) \quad , \tag{45}$$

where $G(f)$ is the true spectrum of $x(t)$, i.e., Fourier transform of $R(\tau)$, and we have Fourier transformed (44) by interchanging integrals and using temporal window

$$W_1(f) = \int dt\ \exp(-i2\pi ft)\ w_1(t) \quad . \tag{46}$$

The convolution result in (45) is a familiar one for the standard FFT processing of weighted, overlapped data segments; see reference 9, eq. (5), for example. Window $W_1(f)$ is real and even about $f = 0$, since weighting $w_1(t)$ is real and even about $t = 0$.

The mean value of the second-stage correlation estimate follows immediately from (19) and (43):

$$Av\{\hat{R}_2(\tau)\} = R(\tau)\ \phi_1(\tau)\ w_2(\tau) = R(\tau)\ w_e(\tau) \quad , \tag{47}$$

where

$$w_e(\tau) \equiv \phi_1(\tau) \ w_2(\tau) \tag{48}$$

is called the effective (overall) weighting of the generalized spectral analysis technique. It incorporates the temporal weighting $w_1(t)$ through its correlation $\phi_1(\tau)$ defined in (44), and it involves lag weighting $w_2(\tau)$ directly. Reference to (44) and to figure 2, for a typical temporal weighting, shows that $\phi_1(\tau)$ is as depicted in figure 4; $\phi_1(\tau)$ extends over $(-L_1, L_1)$ and is zero for $|\tau| > L_1$. Since the effective weighting $w_e(\tau)$ in (48) involves $\phi_1(\tau)$ and $w_2(\tau)$, we now see the reason for the particular choices of L_1 and L_2 in figures 2-4. Specifically, $\pm L_1$ and $\pm L_2$ measure the non-zero extent, in the τ-domain, of the functions that are relevant to the effective weighting. Although L_1 measures the nonzero extent of temporal weighting $w_1(t)$ in the time domain in figure 2, and the nonzero extents of $\phi_1(\tau)$ and $w_2(\tau)$ are $2L_1$ and $2L_2$ in figures 4 and 3, respectively, we will nevertheless refer to L_1 and L_2 as the "lengths" of $\phi_1(\tau)$ and $w_2(\tau)$, respectively, in the τ-domain, for convenience.
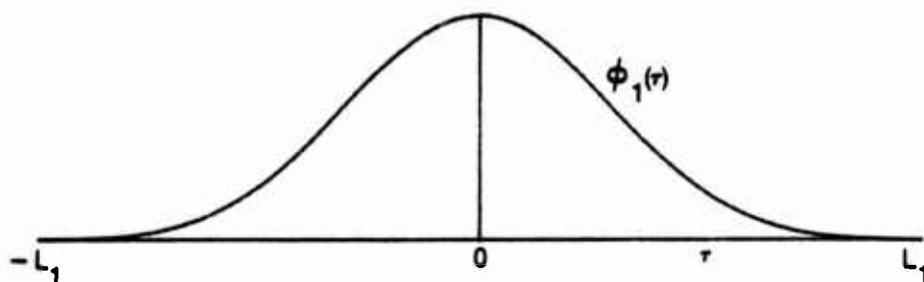
Figure 4. Correlation $\phi_1(\tau)$ of Temporal Weighting $w_1(t)$

In appendix C, $\phi_1(\tau)$ is evaluated for the class of temporal weightings*

$$w_1(t) = \sum_k \alpha_k \ \exp(i2\pi kt/L_1) \quad \text{for} \ |t| < L_1/2 \quad , \tag{49}$$

which includes a wide variety of weightings such as rectangular, Hamming, Hanning, Blackman, Harris, and the recent optimal weightings of Nuttall, reference 19. Specializations to real symmetric $\{\alpha_k\}$ and to a limited number of nonzero coefficients are also made in appendix C.

Finally, since second-stage spectral estimate $\hat{G}_2(f)$ is a Fourier transform of $\hat{R}_2(\tau)$, its mean value follows from (47) and (48) as

$$Av\{\hat{G}_2(f)\} = G(f) \circledast W_e(f) \quad , \tag{50}$$

where

$$W_e(f) = \int d\tau \ \exp(-i2\pi f\tau) \ w_e(\tau) = W_1^2(f) \circledast W_2(f) \tag{51}$$

---

*For brevity, here and later, we omit the "0 otherwise" statement that applies for $|t| > L_1/2$, as was done in (3).

is the effective (overall) window of the generalized spectral analysis technique of interest here. The result in (51) follows by Fourier transformation of the product in (48) and use of (44) (just as done in (45)). Relation (50) is a simple and informative one for the average spectral estimate; it enables ready determination of the amount of spreading caused by particular choices of temporal and lag windows. It holds for any stationary process $x(t)$ with spectrum $G(f)$; thus $x(t)$ needn't be a Gaussian process for (50) to hold true.

As a special case of (50), consider lag weighting $w_2(\tau)$ to be 1 for all τ. Then $W_2(f) = \delta(f)$ and (50) reduces to the result in (45) as expected, since we are employing no lag weighting at all in this case.

As another special case, let temporal weighting $w_1(t)$ be 1 for all $|t| < L_1/2$ and let $L_1 = T$, $L_2 \ll T$. This corresponds to Blackman-Tukey processing. Then $W_1^2(f)$ is proportional to $\text{sinc}^2(Tf)$, which is much narrower in f than $W_2(f)$, meaning that $W_e(f) \approx W_2(f)$, the lag window alone.

Interpretation of the response of the effective window, $W_e(f)$, via convolution (51) can sometimes be deceiving, and it may be helpful and necessary to resort to (48). For example, suppose $w_2(\tau)$ is 1 for $|\tau| < L_2$ and 0 otherwise, where $L_2 \geqslant L_1$. Then (51) says that we have to convolve $\text{sinc}(L_2 f)$, which has -6.63 dB side lobes, with $W_1^2(f)$. Our first impression would be that $W_e(f)$ is bound to have bad side lobes regardless of the temporal window. But recourse to (48) and figure 4 immediately reveals that $w_e(\tau) = \phi_1(\tau)$ for all τ, and that $W_2(f)$ is totally irrelevant, provided that $L_2 \geqslant L_1$. The scaling of $\phi_1(\tau)$ by a constant in (48), over the range of nonzero $\phi_1(\tau)$, obviously has no effect on the relative side lobes of $W_e(f)$. Furthermore, the actual calculation of the effective window via (51) is often tedious, whereas a Fourier transformation of the product in (48) is a reasonable approach, even if only by an FFT.

Constraint on Temporal and Lag Weighting Lengths

The effective window $W_e(f)$ was presented in (51). Its "width" is given approximately by the sum of the widths of the temporal and lag windows. As discussed earlier, we wish to constrain the effective bandwidth B_e of $W_e(f)$, so as to be able to fairly compare the performance of different spectral analysis techniques. The effective bandwidth is given by (1) and can be developed as

$$B_e = \frac{\left[\int df\ W_e(f)\right]^2}{\int df\ W_e^2(f)} = \frac{w_e^2(0)}{\int d\tau\ w_e^2(\tau)} = \frac{\phi_1^2(0)\ w_2^2(0)}{\int d\tau\ \phi_1^2(\tau)\ w_2^2(\tau)}$$

$$= \left(\int d\tau \left[\frac{\phi_1(\tau)}{\phi_1(0)}\right]^2 \left[\frac{w_2(\tau)}{w_2(0)}\right]^2\right)^{-1} \quad , \tag{52}$$

where we have used Parseval's theorem, the Fourier transform relationship in (51), and (48). Since B_e is to be considered fixed, (52) forces a relationship between lengths L_1 and L_2 of $\phi_1(\tau)$ and $w_2(\tau)$.

For example, consider rectangular temporal and lag weightings (this is not a practical case and is presented only for illustration purposes):

$$w_1(t) = 1 \quad \text{for} \quad |t| < L_1/2 \quad ,$$

$$\phi_1(\tau) = L_1 - |\tau| \quad \text{for} \quad |\tau| < L_1 \quad ,$$

$$w_2(\tau) = 1 \quad \text{for} \quad |\tau| < L_2 \quad . \tag{53}$$

Then (52) yields

$$B_e^{-1} = 2 \int_0^{L_m} d\tau \left(1 - \frac{\tau}{L_1}\right)^2 \quad , \tag{54}$$

where

$$L_m = \min(L_1,\ L_2) \quad . \tag{55}$$

Given a value of B_e, (54) can be considered as an equation for L_2 in terms of L_1, or vice versa. Here we have fixed the *shapes* of the weightings and are varying the lengths so as to realize the specified frequency resolution B_e.

Generally, the qualitative character of the interrelationship between lengths is as depicted in figure 5, for fixed weighting shapes. The larger one of the lengths is made, the smaller the other length can get and still satisfy the specified bandwidth of the effective window. For a specified pair of shapes for $w_1(t)$ and $w_2(\tau)$, a plot like figure 5 can be used in two different ways. If we pick a value for $B_e L_1$, this determines $B_e L_2$ and hence L_2/L_1. On the other hand, choice of a value for the ratio of lengths, L_2/L_1, puts a line through the origin of slope L_2/L_1, and thereby determines $B_e L_1$ and $B_e L_2$ where the line intersects the curve. We note therefore that knowledge of one of the following three quantities determines the other two: L_2/L_1, $B_e L_1$, $B_e L_2$.

The limiting parameter values on figure 5 are determined as follows: as $L_1 \rightarrow \infty$, then $L_2 \rightarrow L_2(\text{min})$, where now (from (52))

$$B_e^{-1} = \int d\tau \left[\frac{w_2(\tau)}{w_2(0)}\right]^2 = 2L_2(\text{min}) \quad c\{w_2\} \quad . \tag{56}$$

Here, $c\{ \}$ is a "shape factor" defined for any limited-duration function g as

$$c\{g\} = \frac{1}{2L_g} \int d\tau \left[\frac{g(\tau)}{g(0)}\right]^2 \quad , \tag{57}$$

where it is assumed that

$$g(\tau) = 0 \quad \text{for} \quad |\tau| > L_g \quad ; \tag{58}$$
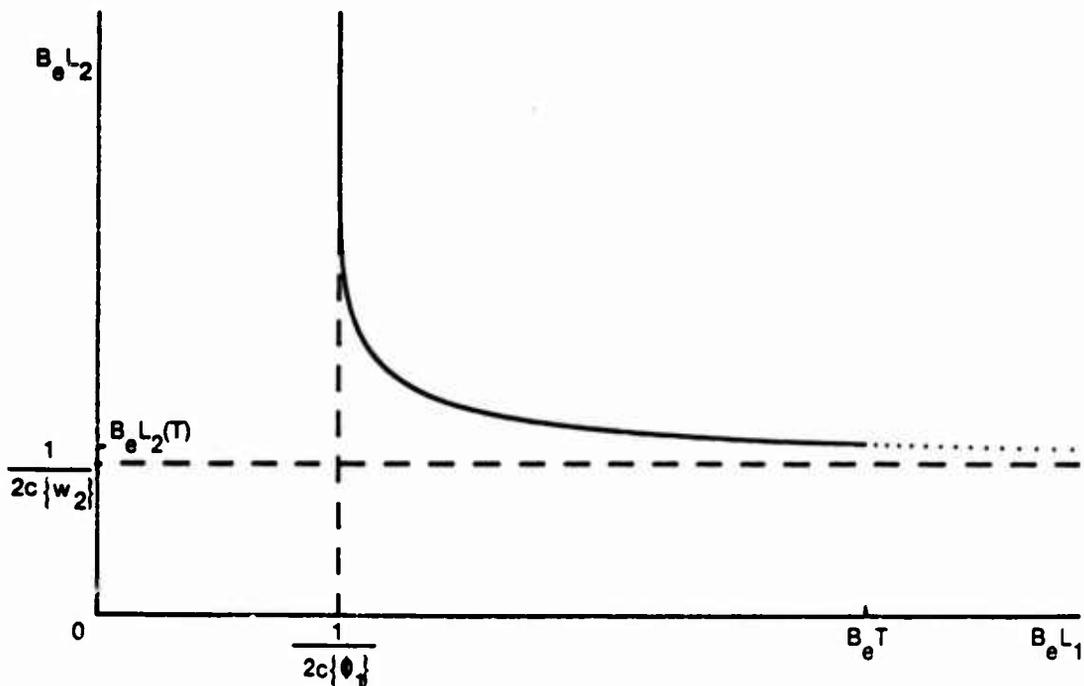
Figure 5. Interrelationship of Lengths L_1 and L_2 for Fixed Shapes of the Temporal and Lag Weightings

that is, $2L_g$ is the nonzero extent of $g(\tau)$. Shape factor $c\{g\}$ is independent of the magnitude of $g(\tau)$ and of its length on the τ-scale. Thus (56) yields $B_e L_2(min) = (2 \ c\{w_2\})^{-1}$, which is entered on the ordinate in figure 5. However, since L_1 is limited by T, the dotted portion of the curve on the far right is not attainable.

Conversely, if instead, $L_2 \rightarrow \infty$, then $L_1 \rightarrow L_1(min)$, where now (from (52))

$$B_e^{-1} = \int d\tau \left[\frac{\phi_1(\tau)}{\phi_1(0)}\right]^2 = 2L_1(min) \ c\{\phi_1\} \quad ; \tag{59}$$

that is, $B_e L_1(min) = (2 \ c\{\phi_1\})^{-1}$. This value is attainable; it corresponds to no lag weighting. The ratio, L_2/L_1, of weighting lengths can take values in the range $L_2(T)/T$ to ∞; for $B_e T \gg 1$, this constitutes the range from almost zero to infinity.

Since the shape factors $c\{w_2\}$ and $c\{\phi_1\}$ are important limits on the weighting lengths, tables of their numerical values for a number of useful weight functions are given below. The weightings listed under C5, C3, C1 are those given in reference 19, figures 10, 11, 12, respectively; the notation means

$$
\begin{aligned}
&\text{C5: } &&\text{continuous fifth derivative of weighting} \\
&\text{C3: } &&\text{continuous third derivative of weighting} \\
&\text{C1: } &&\text{continuous first derivative of weighting}
\end{aligned}
\tag{60}
$$

For the class of lag weightings given by

$$w_2(\tau) = \sum_{k \geq 0} a_k \cos(\pi k \tau / L_2) \quad \text{for } |\tau| < L_2 \quad , \tag{61}$$

the shape factor is

$$c\{w_2\} = \frac{1}{2L_2} \int d\tau \left[\frac{w_2(\tau)}{w_2(0)}\right]^2 = \frac{a_0^2 + \frac{1}{2}\left(a_1^2 + a_2^2 + \ldots\right)}{(a_0 + a_1 + a_2 + \ldots)^2} \quad . \tag{62}$$

This is evaluated for several weightings in table 1.

Table 1. Shape Factor for Lag Weighting $w_2(\tau)$

$w_2(\tau)$	$c\{w_2(\tau)\}$
Rectangular	1.000
Hanning	.3750
Hamming	.3964
Blackman	.3046
C5	.2256
C3	.2442
C1	.2558

For the class of temporal weightings given by

$$w_1(t) = \sum_{k \geq 0} a_k \cos(2\pi kt/L_1) \quad \text{for} \quad |t| < L_1/2 \quad , \tag{63}$$

the correlation $\phi_1(\tau)$ is evaluated in general in appendix C. The shape factor of $\phi_1(\tau)$,

$$c\{\phi_1\} = \frac{1}{2L_1} \int d\tau \left[\frac{\phi_1(\tau)}{\phi_1(0)}\right]^2 \quad , \tag{64}$$

can then be evaluated numerically and is given in table 2.

Table 2. Shape Factor for Correlation
$\phi_1(\tau)$ of Temporal Weighting $w_1(t)$

$w_1(t)$	$c\{\phi_1\}$
Rectangular	$.3333 = 1/3$
Hanning	$.2405 = (8\pi^2 + 35)/(48\pi^2)$
Hamming	$.2628$
Blackman	$.2073$
C5	$.1545$
C3	$.1678$
C1	$.1763$

Plots of the relationship between L_1 and L_2 dictated by (52) are given in figure 6 for various combinations of temporal and lag weightings. For a rectangular lag weighting, the curve will actually reach $B_e L_1 = (2c\{\phi_1\})^{-1}$ when $L_2 = L_1$; then the curve goes vertically up from this point for $L_2 \geq L_1$ (see figure 6A). The procedure for the evaluation of figure 6 is as described under figure 5; namely, pick a value for L_2/L_1, compute $B_e L_1$ via (52), and then compute $B_e L_2 = B_e L_1 * (L_2/L_1)$.

If the maximum segment length, L_1, is specified (as for example, when the maximum FFT size and the time-sampling spacing Δ_t are fixed), under what condition can a desired effective frequency resolution, B_d, be met? The answer to this question is available from figure 5; namely, we see that

$$B_e L_1 \geq (2c\{\phi_1\})^{-1} \quad , \quad \text{or} \quad B_e \geq \frac{1}{2c\{\phi_1\}L_1} \quad . \tag{65}$$

Thus if desired resolution B_d is greater than or equal to the right-most term of (65), there exists a choice for lag length L_2 that will yield the desired frequency resolution. The shape factor in (65) depends only on the temporal weighting $w_1(t)$.
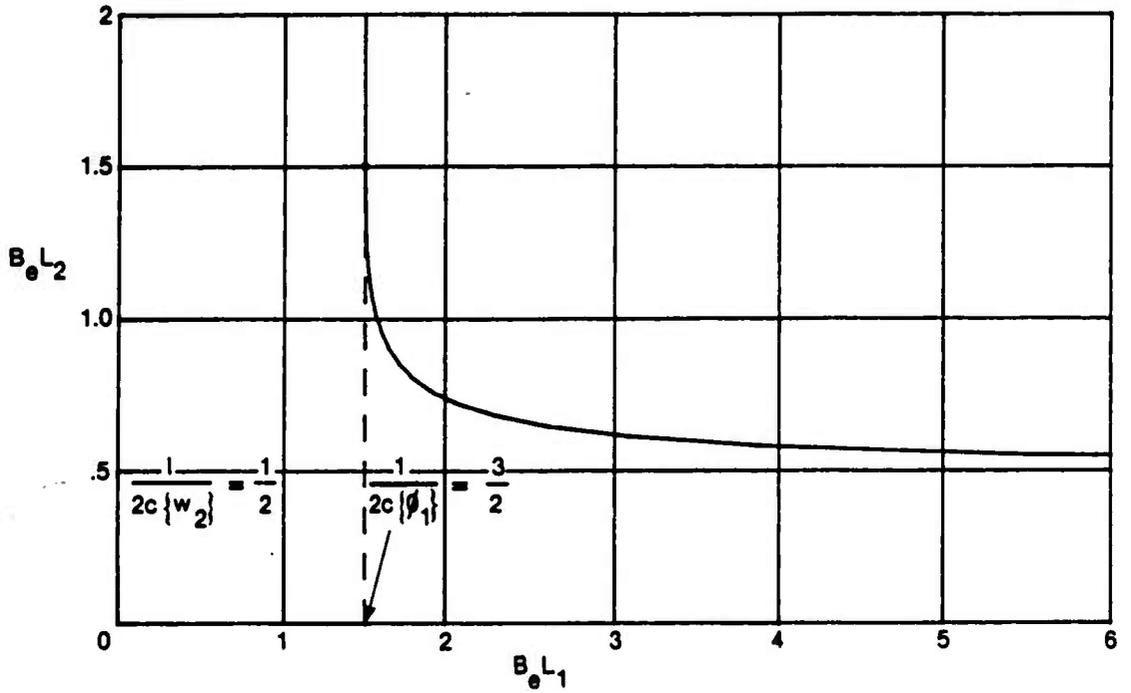
6A. Rectangular Temporal Weighting, Rectangular Lag Weighting
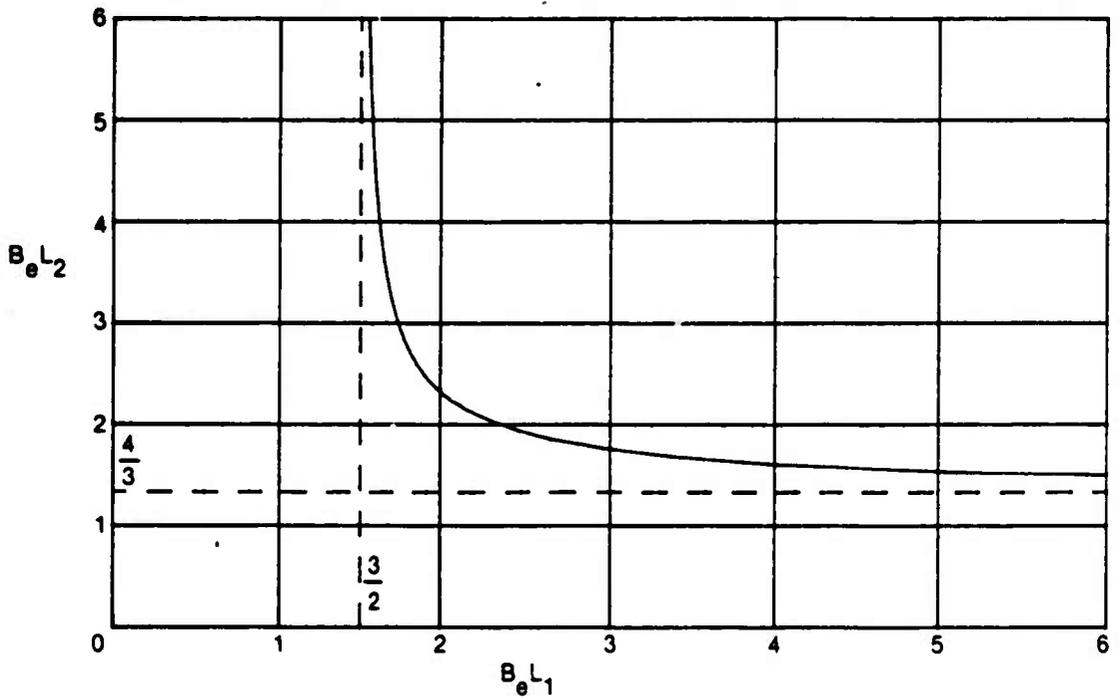
6B. Rectangular Temporal Weighting, Hanning Lag Weighting

Figure 6. Allowed Lengths of Various Temporal and Lag Weighting Pairs
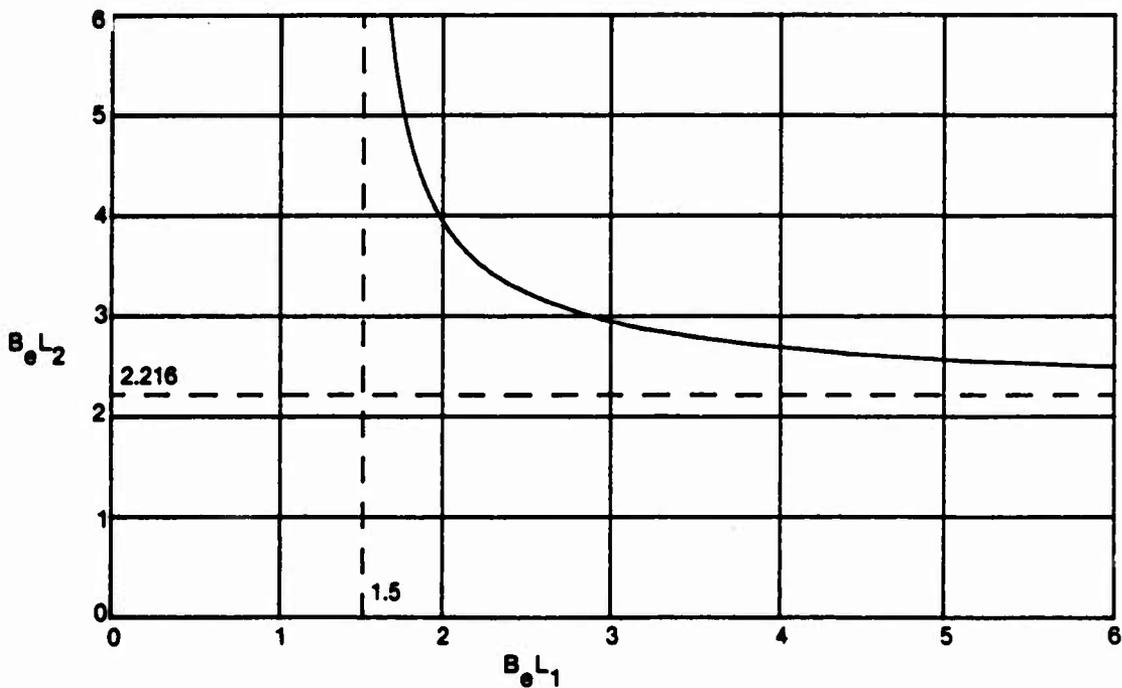
6C. Rectangular Temporal Weighting, C5 Lag Weighting
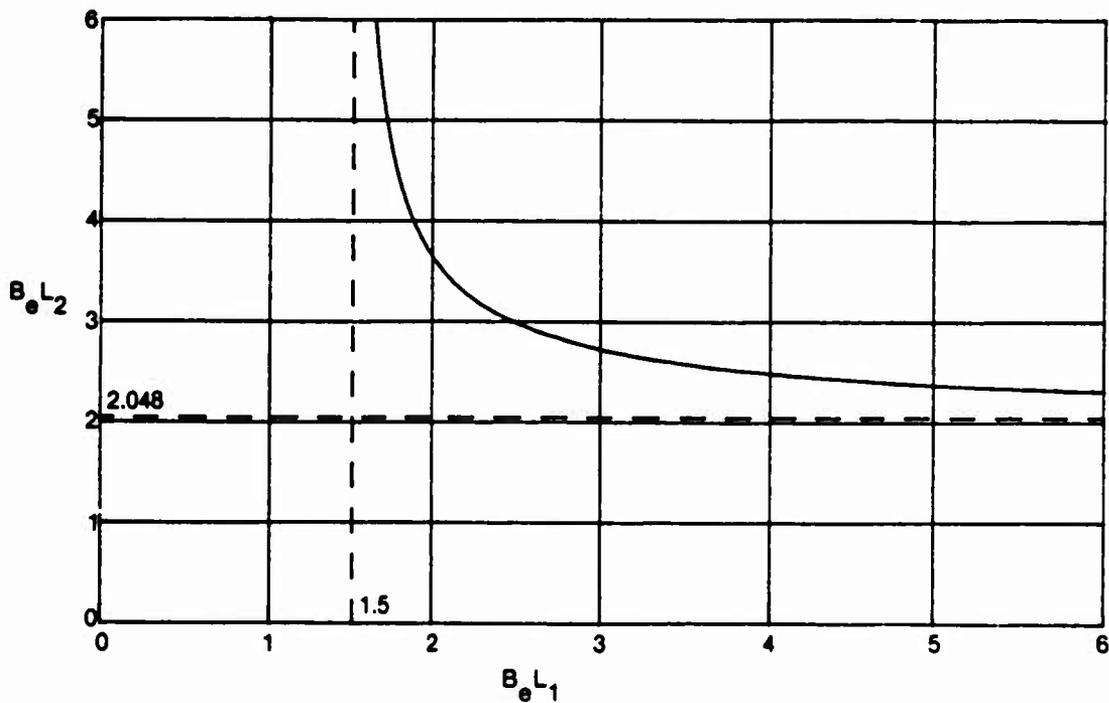
6D. Rectangular Temporal Weighting, C3 Lag Weighting

**Figure 6. Allowed Lengths of Various Temporal
and Lag Weighting Pairs (Cont'd)**
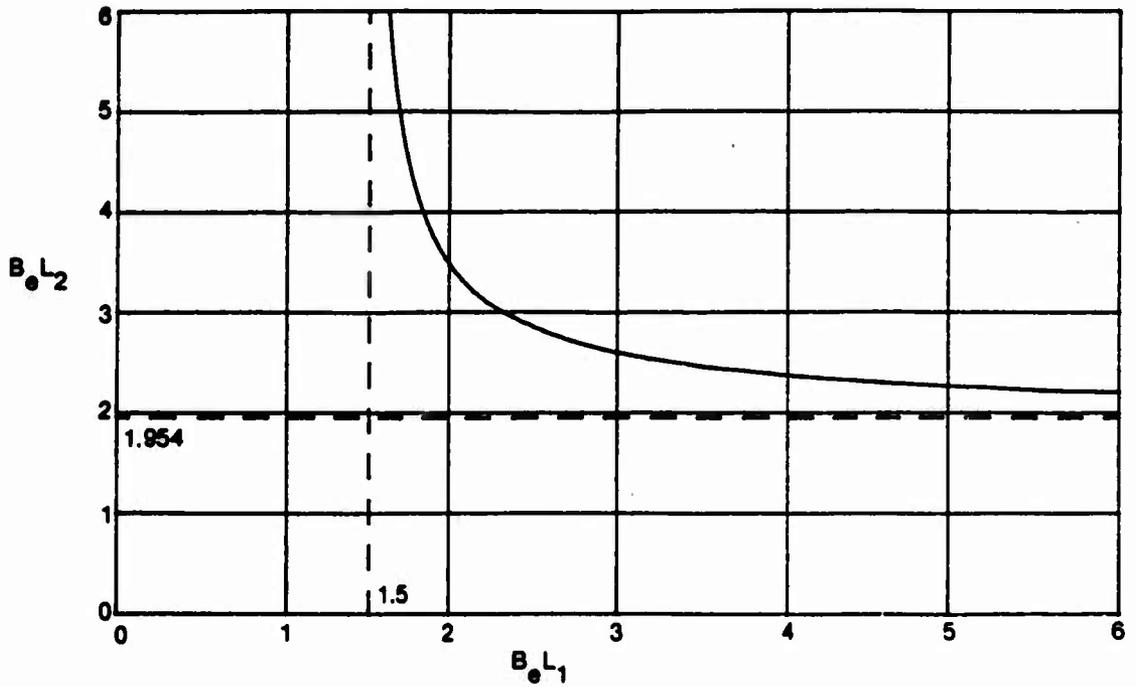
6E. Rectangular Temporal Weighting, C1 Lag Weighting
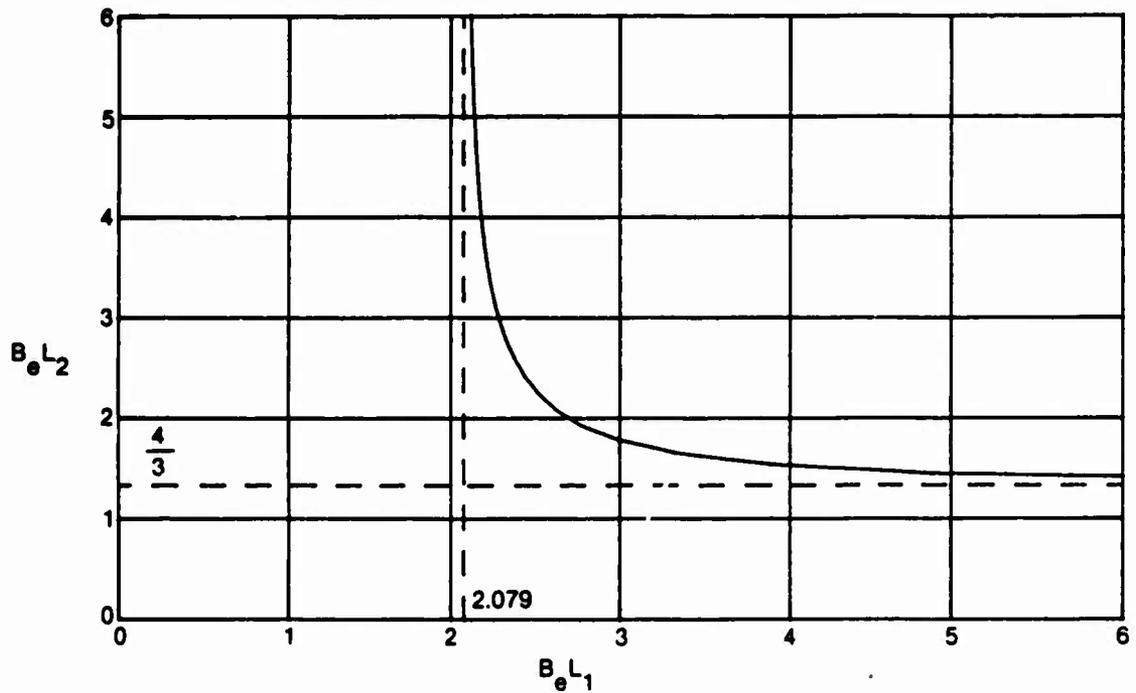
6F. Hanning Temporal Weighting, Hanning Lag Weighting

Figure 6. Allowed Lengths of Various Temporal
and Lag Weighting Pairs (Cont'd)

Normalization of Weightings

The average of the first-stage correlation estimate was given in (43). For $\tau = 0$, it yields

$$Av\{\hat{R}_1(0)\} = R(0) \, \phi_1(0) \quad . \tag{66}$$

Since $R(0)$ is the true power in the process $x(t)$ under investigation, it is convenient to normalize according to

$$1 = \phi_1(0) = \int dt \, w_1^2(t) \quad . \tag{67}$$

Then $\hat{R}_1(0)$ is an unbiased estimator of $R(0)$.

Additionally, from (47) and (48), we have, for the second-stage correlation estimate,

$$Av\{\hat{R}_2(0)\} = R(0) \, \phi_1(0) \, w_2(0) \quad . \tag{68}$$

Therefore, in addition to (67), we also set lag weighting value

$$w_2(0) = 1 \quad , \tag{69}$$

making $\hat{R}_2(0)$ an unbiased estimator of $R(0)$. There follows, for the effective weighting,

$$w_e(0) = 1 \quad . \tag{70}$$

Since there is no significant loss of generality, the normalizations in (67), (69), and (70) will be used in the rest of this report.

Discrete-Time Processing

For the impulsive temporal weighting introduced in (24)-(29), the normalization (67) must be modified somewhat, since the integral of $w_1^2(t)$ in (24) would be infinite. We resort to (28) and require that the origin value of the sample correlation satisfy the unbiased requirement that

$$Av\{\hat{R}_{10}\} = R(0) \quad . \tag{71}$$

Reference to (28), (29), (26), and (43) yields

$$Av\{\hat{R}_{10}\} = Av\{\hat{R}_o^{(p)}\} = \Delta_t \sum_m \overline{|y_{pm}|^2} = \Delta_t \sum_m w_{1m}^2 \, R(0) \quad ; \tag{72}$$

therefore the normalization is

$$\Delta_t \sum_m w_{1m}^2 = 1 \quad . \tag{73}$$

This is the discrete analog to the integral constraint in (67).

The correlation $\phi_1(\tau)$ of temporal weighting $w_1(t)$ in (24) is given by (44) as usual and is expressible as

$$\phi_1(\tau) = \Delta_t \sum_k \phi_{1k} \, \delta(\tau - k\Delta_t) \quad , \tag{74}$$

where

$$\phi_{1k} \equiv \Delta_t \sum_m w_{1,m} w_{1,m-k} \quad . \tag{75}$$

Thus we see that (73) is tantamount to $\phi_{10} = 1$, which replaces the constraint $\phi_1(0) = 1$ in (67) for the continuous temporal weighting case.

If we also require that second-stage correlation estimate (42) satisfy the unbiased requirement

$$Av\{\hat{R}_{20}\} = R(0) \quad , \tag{76}$$

then, as before, we require

$$w_2(0) = 1 \quad . \tag{77}$$

Finally, the effective weighting becomes, upon use of (74),

$$w_e(\tau) = \phi_1(\tau) \, w_2(\tau) = \Delta_t \sum_k \phi_{1k} \, w_2(k\Delta_t) \, \delta(\tau - k\Delta_t) \quad . \tag{78}$$

The normalizations adopted above make the area of the impulse at $\tau = 0$ equal to

$$\Delta_t \, \phi_{10} \, w_2(0) = \Delta_t \quad . \tag{79}$$

Examples of Effective Windows

We consider first a rectangular temporal weighting $w_1(t)$, for which the correlation is triangular,

$$\phi_1(\tau) = 1 - |\tau|/L_1 \quad \text{for} \quad |\tau| < L_1 \quad , \tag{80}$$

and the class of lag weightings as given earlier by (61):

$$w_2(\tau) = \sum_{k \geq 0} a_k \cos(\pi k \tau/L_2) \quad \text{for} \quad |\tau| < L_2 \quad . \tag{81}$$

Then constraint (52) yields

$$B_e L_2 = \left(2 \int_0^1 dx \left[\sum_{k \geq 0} a_k \cos(\pi k x) \right]^2 \left[1 - \frac{L_2 x}{L_1} \right]^2 \right)^{-1} \quad \text{for} \quad L_2 \leq L_1 \quad . \tag{82}$$

The effective weighting, $w_e(\tau)$, is given by the product of (80) and (81); its Fourier transform is the effective window

$$W_e(f) = \frac{2L_2}{\pi^2} \sum_{k \geq 0} a_k \left\{ \left(1 - \frac{L_2}{L_1} \right) 2\pi\nu \sin(2\pi\nu) \frac{(-1)^k}{4\nu^2 - k^2} \right.$$

$$\left. + \frac{L_2}{L_1} \left[1 - (-1)^k \cos(2\pi\nu) \right] \frac{4\nu^2 + k^2}{\left(4\nu^2 - k^2 \right)^2} \right\} \quad \text{for} \quad L_2 \leq L_1 \quad , \tag{83}$$

where

$$\nu \equiv B_e L_2 u \quad , \quad u \equiv f/B_e \quad . \tag{84}$$

Although (82) and (83) could be extended to the case where $L_2 > L_1$, that range is not of practical interest, as will become apparent later.

The numerical procedure for evaluation of the effective window is to first select the shape of the lag window by specifying coefficients $\{a_k\}$. Then we choose a value for L_2/L_1 and compute $B_e L_2$ from (82). We can then employ (83) and (84) to determine $W_e(f)$. Four examples are given in figure 7, where we have plotted

$$\text{dB} \equiv 10 \log \frac{W_e(f)}{W_e(0)} \quad \text{vs} \quad \frac{f}{B_e} \quad . \tag{85}$$
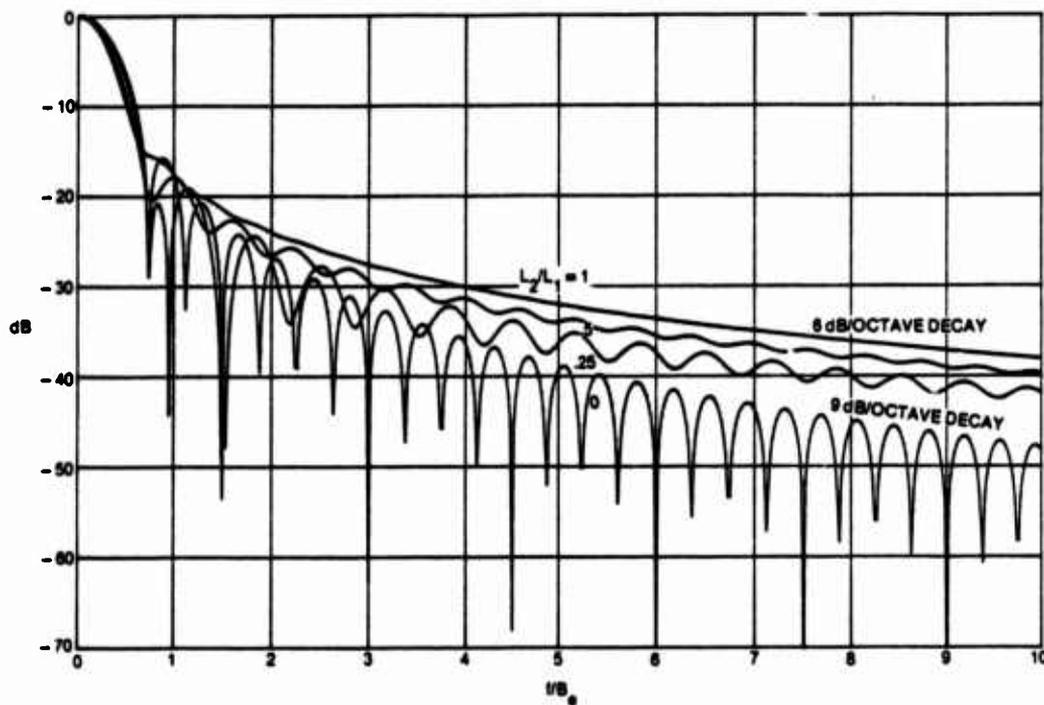
7A. Effective Window for Rectangular Temporal Weighting and Hanning Lag Weighting
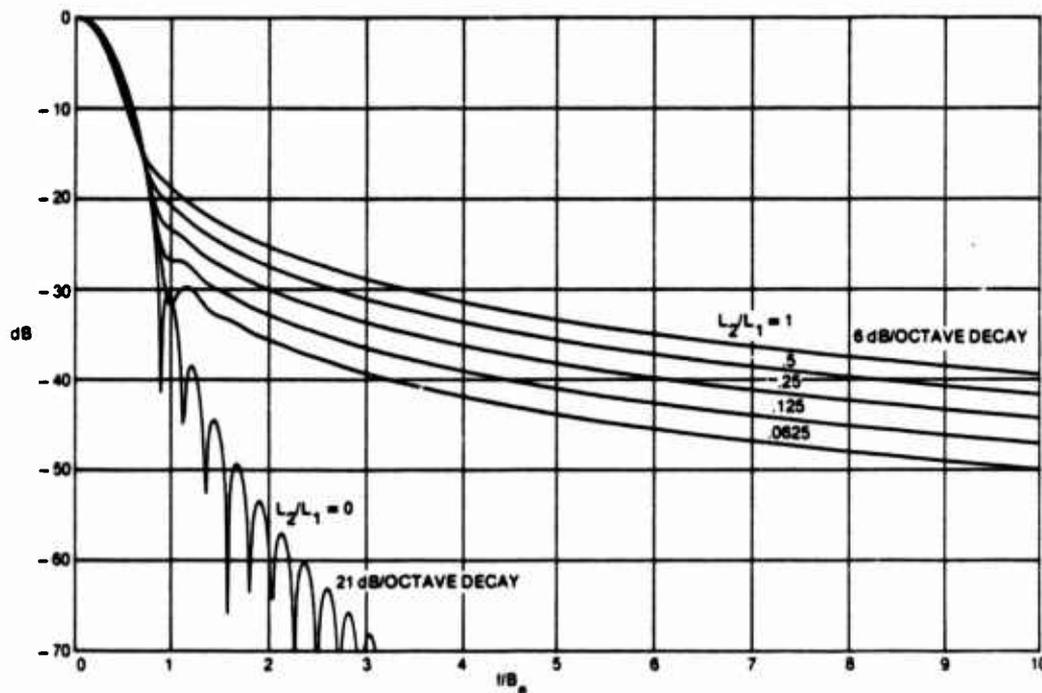
7B. Effective Window for Rectangular Temporal Weighting and C5 Lag Weighting

Figure 7. Examples of Effective Windows for Rectangular Temporal Weighting
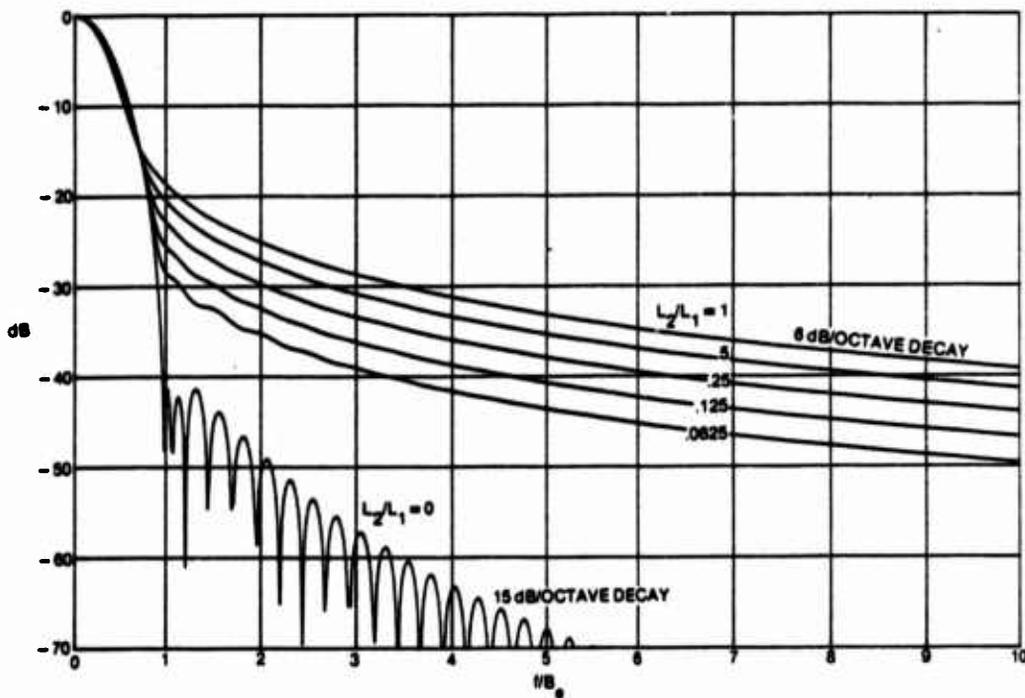
7C. Effective Window for Rectangular Temporal Weighting and C3 Lag Weighting
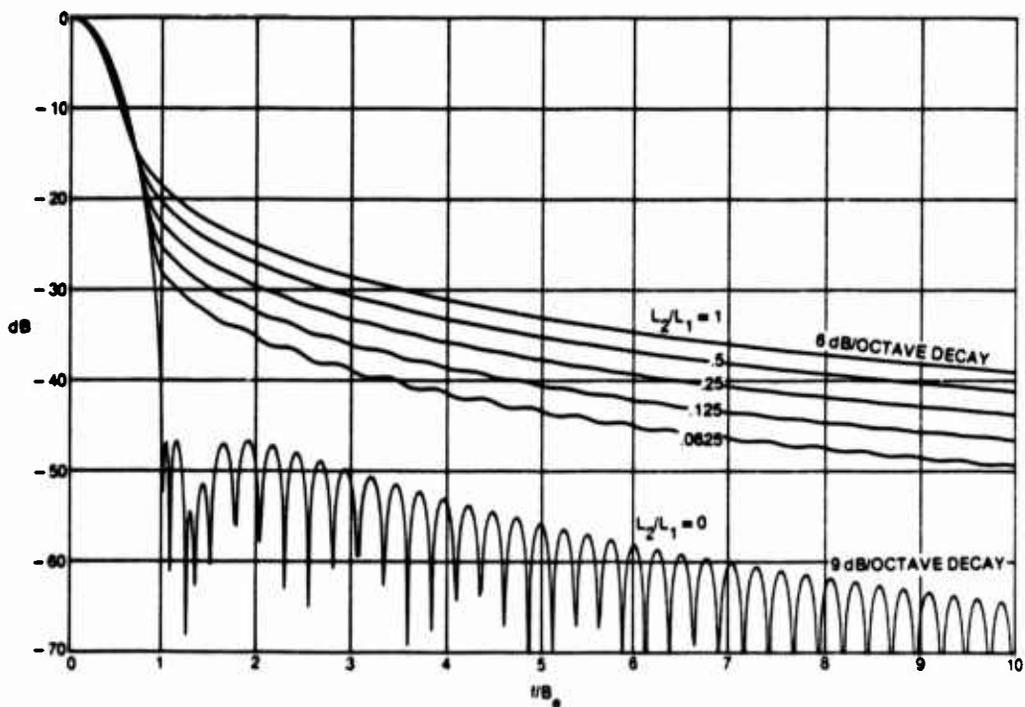
7D. Effective Window for Rectangular Temporal Weighting and C1 Lag Weighting

Figure 7. Examples of Effective Windows for Rectangular Temporal Weighting (Cont'd)

The curve corresponding to $L_2/L_1 = 0$ (i.e., $L_1 = \infty$) is that for the lag window alone.

The overriding impression of the plots in figure 7 is that the effective window has poor side lobe behavior and decay unless L_2/L_1 is chosen very small. That is, the poor side lobe behavior of temporal window $W_1^2(f)$ enters the convolution (51) for $W_e(f)$, and is difficult to suppress, even by choice of good lag windows. It would be desirable to realize the bottom-most figures in each of these plots, since these latter curves have good side lobes and decay; a procedure for accomplishing this goal is presented in the next section.

The situation is significantly improved when the temporal weighting is tapered. An example for Hanning temporal and lag weightings is given in figure 8. The bottom-most curve has an eventual 18 dB/octave decay because $\phi_1(\tau)$ has a discontinuous fifth derivative at $\tau = 0$, which is not compensated by $w_2(\tau)$. ($\phi_1(\tau)$ also has a discontinuous fifth derivative at $\tau = \pm L_1$, but this is converted to a discontinuous seventh derivative for $w_e(\tau)$ by means of $w_2(\tau)$ when $L_2 = L_1$.)
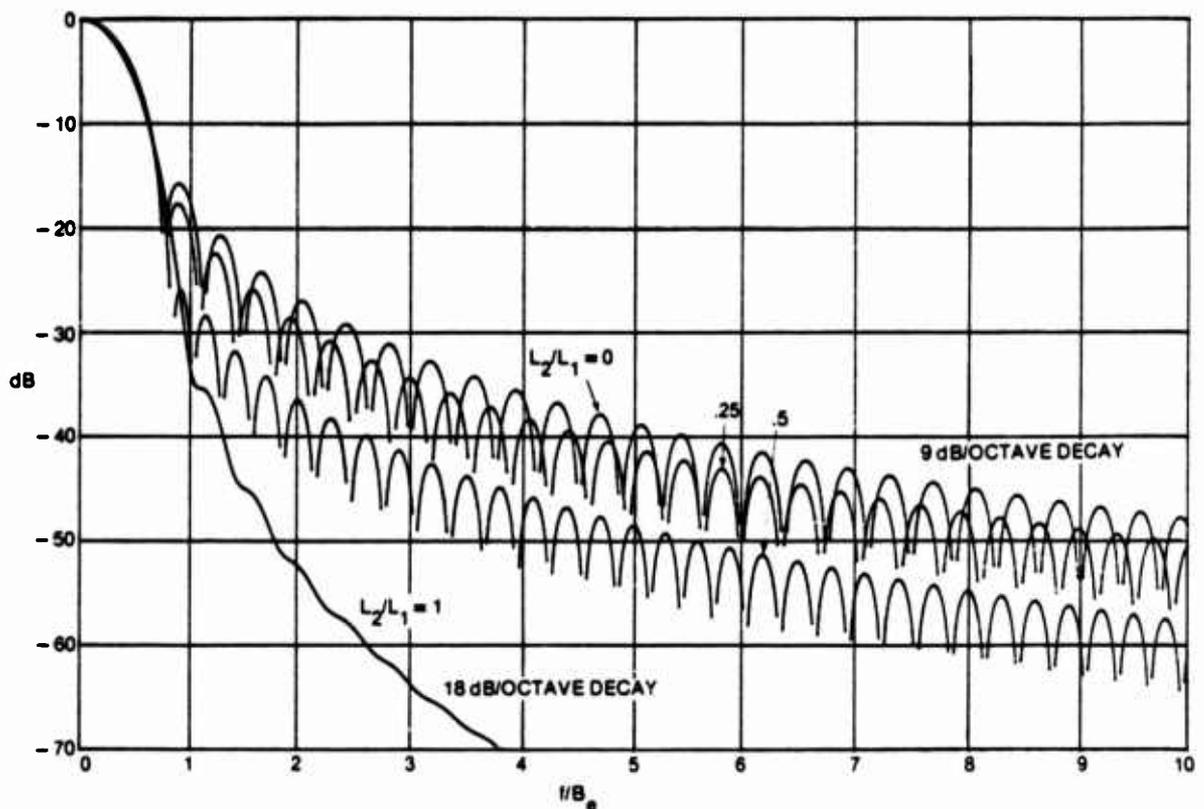
Figure 8. Effective Window for Hanning Temporal Weighting and
Hanning Lag Weighting

Discrete-Time Processing

The temporal weighting $w_1(t)$ is given by (24), and its correlation $\phi_1(\tau)$ by (74), where coefficients $\{\phi_{1k}\}$ are given by (75). The Fourier transform of (24) leads to temporal window

$$W_1(f) = \Delta_t \sum_m w_{1m} \exp(-i2\pi fm\Delta_t) \quad , \tag{86}$$

which has period $1/\Delta_t$ in f, and is real and even, since weight sequence $\{w_{1m}\}$ is real and even.

For a general lag weighting $w_2(\tau)$, the effective weighting is given by (78). The effective window $W_e(f)$ is given by (51) as the Fourier transformation of (78):

$$W_e(f) = \Delta_t \sum_k \phi_{1k} w_2(k\Delta_t) \exp(-i2\pi fk\Delta_t) \quad , \tag{87}$$

which also has period $1/\Delta_t$ in f, and is real and even, since lag weighting $w_2(\tau)$ is real and even. This result holds for *any* lag weighting $w_2(\tau)$ and is a very useful form for computing $W_e(f)$ for any value of frequency f. The convolutional form of (51) is not very useful for computing $W_e(f)$ for general $W_2(f)$.

As a special case, we can evaluate (87) at particular frequencies $f_n = n\Delta_F = n/(N\Delta_t)^{-1}$, as in (33) and (34):

$$W_e\left(\frac{n}{N\Delta_t}\right) = \Delta_t \sum_k \phi_{1k} w_2(k\Delta_t) \exp(-i2\pi nk/N) \quad , \tag{88}$$

which can be accomplished as an N-point FFT. We should choose N large in order that (88) be capable of tracing the fine detail of $W_e(f)$. This is an attractive and efficient way to evaluate the effective window.

A Special Lag Window for Discrete-Time Processing

The result in (87) applies for discrete time sampling and arbitrary lag weighting. We now specialize to the lag window given in (38):

$$W_2(f) = \Delta_f \sum_n W_{2n} \delta(f - n\Delta_f) \quad , \tag{89}$$

where sequence $\{W_{2n}\}$ is real and even, and frequency spacing Δ_f need not be related to time spacing Δ_t in (24). Then, via inverse Fourier transformation (20), the corresponding lag weighting is

$$w_2(\tau) = \Delta_f \sum_n W_{2n} \exp(i2\pi n\Delta_f \tau) \quad , \tag{90}$$

and, in particular, sample values

$$w_2(k\Delta_t) = \Delta_f \sum_n W_{2n} \exp(i2\pi nk \Delta_f \Delta_t) \quad , \tag{91}$$

which can be used in (87) to evaluate $W_e(f)$ for any frequency f whatsoever.

An alternative form to (87) for calculation of $W_e(f)$ for this special lag window (89) is afforded by substitution of (89) in the convolutional form of (51):

$$W_e(f) = \Delta_f \sum_n W_{2n} W_1^2(f - n\Delta_f) \quad . \tag{92}$$

The temporal window $W_1(f)$ is given here by (86). Equation (92) is an attractive form when the number, N_2, of nonzero coefficients $\{W_{2n}\}$ is small and $W_1(f)$ can be evaluated in closed form. In fact, (92) actually holds for *any* temporal weighting $w_1(t)$; it is not limited to the discrete-time forms (24) and (86). Equations (87) and (92) are duals in the sense that (87) applies to any $w_2(\tau)$ and an impulsive $w_1(t)$, whereas (92) applies to any $W_1(f)$ and an impulsive $W_2(f)$. Either equation can be evaluated at any f of interest.

Our first example is rectangular temporal weighting; from figure 2, (24), and (73), $w_{1m} = L_1^{-1/2}$, where $L_1 = N_1\Delta_t$. Then, from (86),

$$W_1^2(f) = L_1 \left[\frac{\sin(\pi L_1 f)}{N_1 \sin(\pi L_1 f/N_1)} \right]^2 \equiv L_1 \, Q_{N_1}(L_1 f) \quad . \tag{93}$$

For the lag window, we take impulsive form (89); then (92) and (93) give

$$W_e(f) = L_1 \Delta_f \sum_n W_{2n} Q_{N_1}(L_1 f - nL_1\Delta_f) \quad . \tag{94}$$

Two important choices yet to be made are $L_1\Delta_f$, the relative frequency spacing used in frequency smoothing, and the set of coefficients $\{W_{2n}\}$. For Hanning frequency smoothing, the latter is

$$W_{2n} = \frac{1}{\Delta_f} \begin{cases} 1/2 & \text{for } n = 0 \\ 1/4 & \text{for } n = \pm 1 \\ 0 & \text{otherwise} \end{cases} \quad , \quad N_2 = 3 \quad . \tag{95}$$

The effective windows for $L_1\Delta_f = 1/2$ and 1 are given in figure 9 for $N_1 = 32$. Window (94) is even about $f = 0$ and has period $1/\Delta_t$ in f; hence only the region $0, (2\Delta_t)^{-1}$ is plotted in f. The window in figure 9A has no deep notches since the frequency displacement (spacing) $\Delta_f = (2L_1)^{-1}$ causes the notches to be filled in; the window for $\Delta_f = L_1^{-1}$ in figure 9B reinforces the notches and has a significant shoulder near $f = 1/L_1$. Both windows have slow decay with frequency and do not have significant rejection, even near Nyquist frequency. Closer spacings than $(2L_1)^{-1}$ do not improve the decay or rejection capabilities; wider spacings than L_1^{-1} generate humps in the effective window. The bad features of rectangular temporal weighting are not undone by Hanning frequency smoothing; see also figure 7A.

9A. $L_1\Delta_f = 1/2, N_1 = 32$
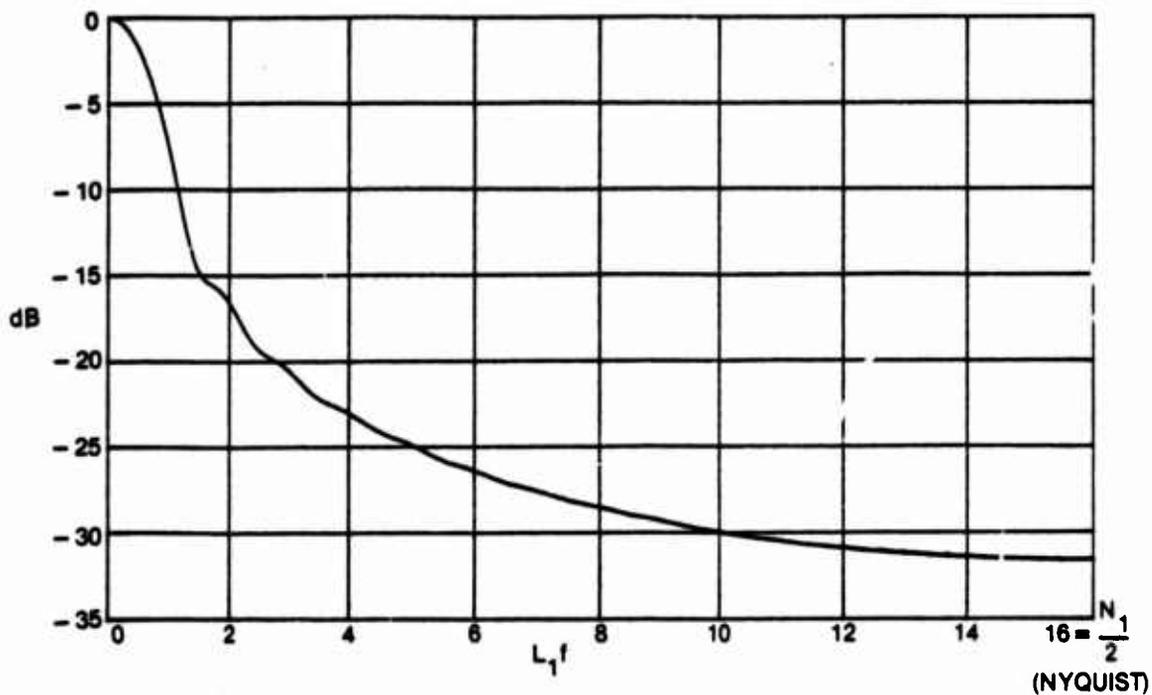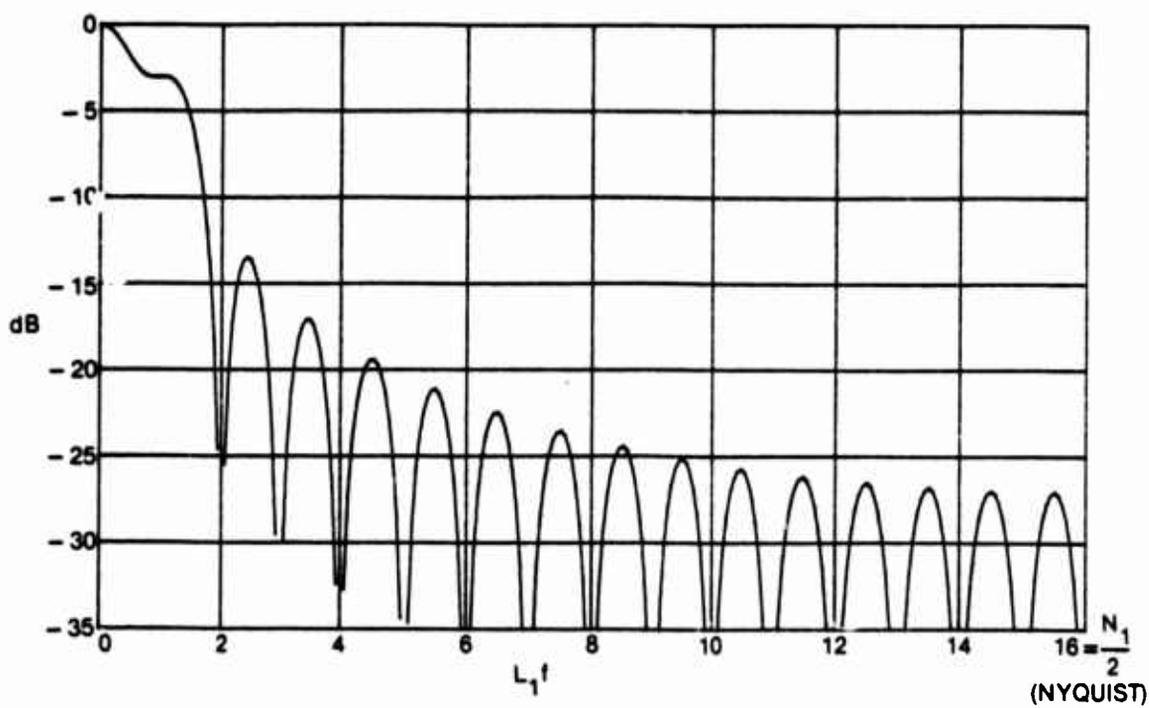
9B. $L_1\Delta_f = 1, N_1 = 32$

Figure 9. Effective Window for Discrete-Time Rectangular Temporal
Weighting and Hanning Frequency-Smoothing

The second example is rectangular temporal weighting with $N_1 = 32$ and rectangular frequency smoothing over five frequency samples; i.e.,

$$W_{2n} = \frac{1}{\Delta_f} \begin{cases} 1/5 & \text{for } |n| \leq 2 \\ \\ 0 & \text{otherwise} \end{cases} \quad , \quad N_2 = 5 \quad . \tag{96}$$

The effective windows for $L_1 \Delta_f = 1/2$ and 1 are given in figure 10. The main lobe humps in figure 10B are caused by the displacements of $W_1^2(f)$ according to (92). Both windows again have poor decay and poor rejection; however, the main lobe is more box-like in shape than previously.

The third example is identical to the previous one except that $N_2 = 11$. The plots in figure 11 reveal that the main lobe is quite box-like, but the decay and rejection are no better than previous cases. According to (92), we are merely taking the poor side lobes and decay of $W_1^2(f)$ and moving them about, but not improving them in any way.

The last example in this subsection is Hanning temporal weighting with no frequency smoothing at all. The effective window for $N_1 = 32$ is simply $W_1^2(f)$ and is plotted in figure 12. It has the familiar -31.5 dB peak side lobe, a rapid decay, and significant rejection capability.

10A. $L_1 \Delta_f = 1/2$, $N_1 = 32$
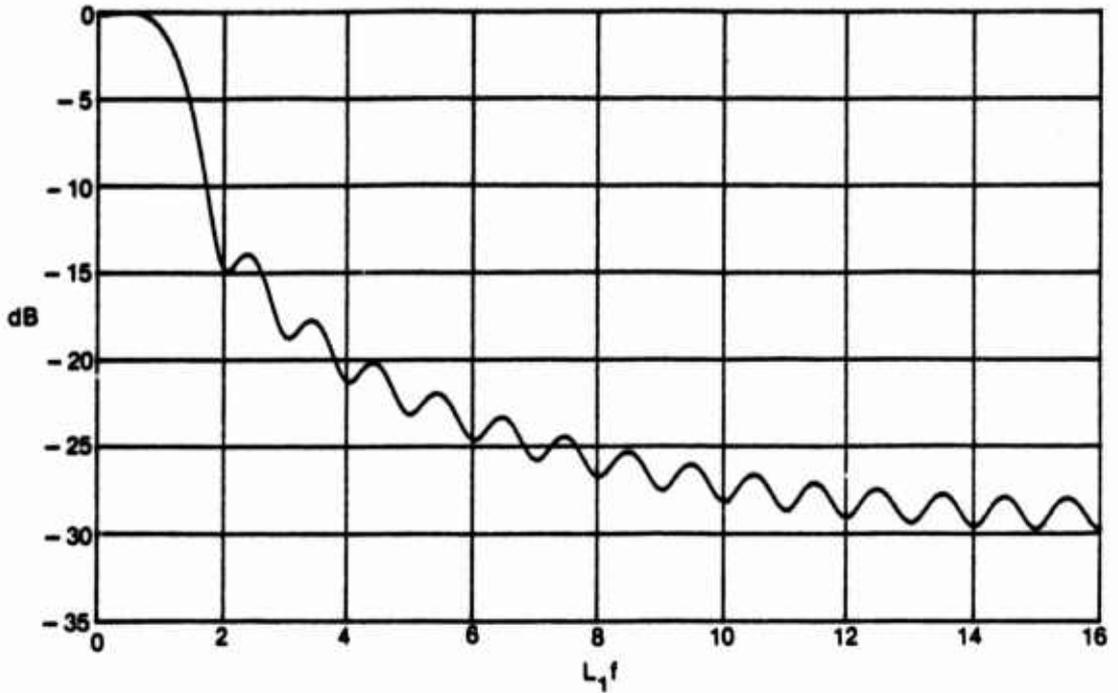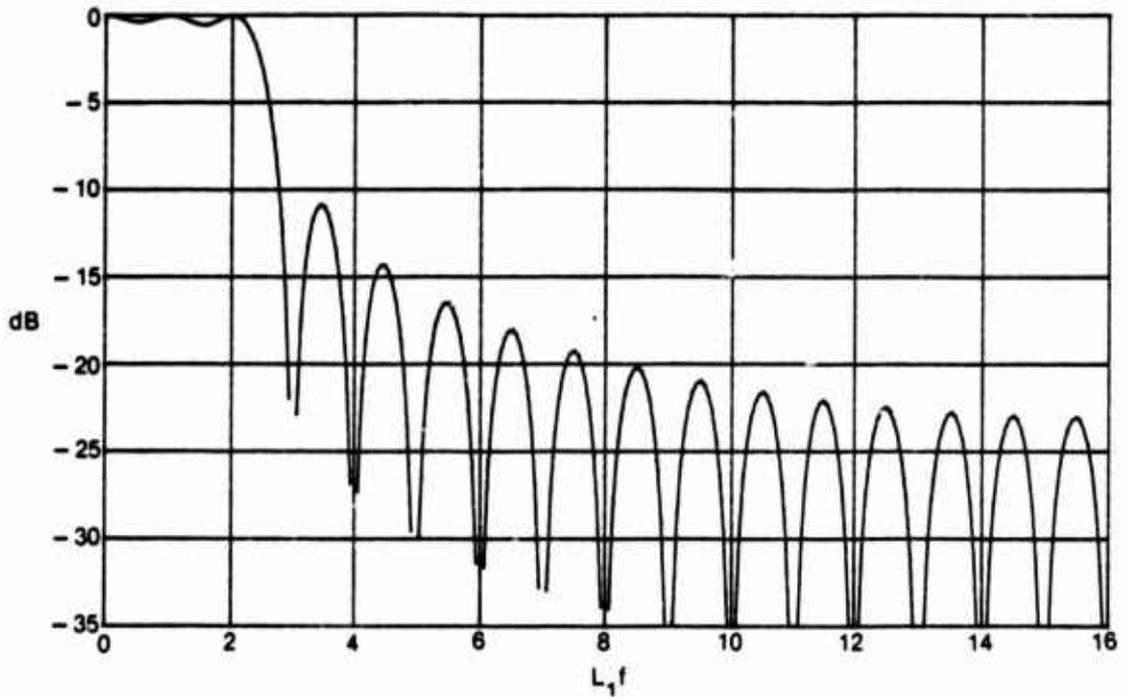
10B. $L_1 \Delta_f = 1$, $N_1 = 32$

Figure 10. Effective Window for Discrete-Time Rectangular Temporal Weighting
and Rectangular Frequency-Smoothing with $N_2 = 5$

11A. $L_1\Delta_f = 1/2, N_1 = 32$

11B. $L_1\Delta_f = 1, N_1 = 11$
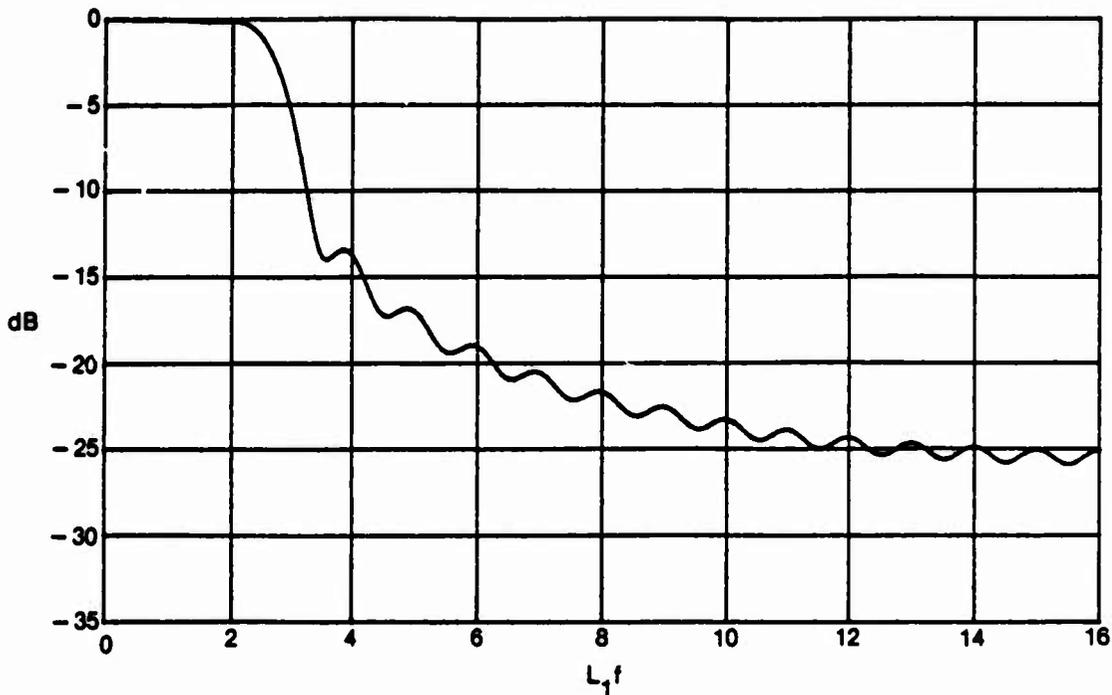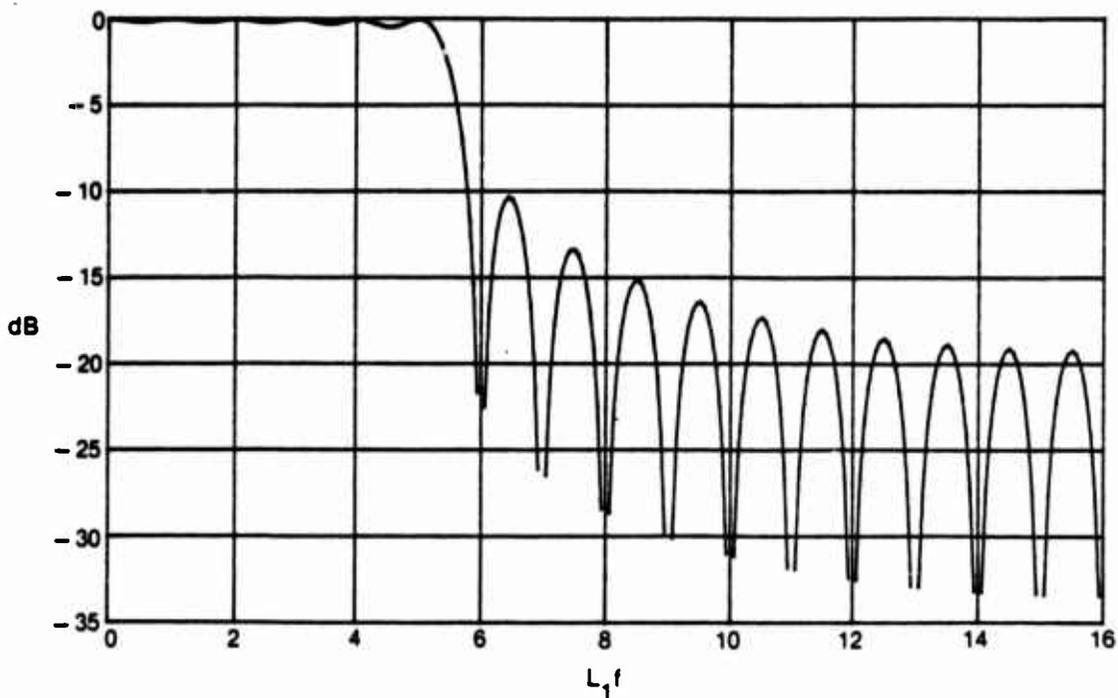
Figure 11. Effective Window for Discrete-Time Rectangular Temporal Weighting and Rectangular Frequency-Smoothing with $N_2 = 11$
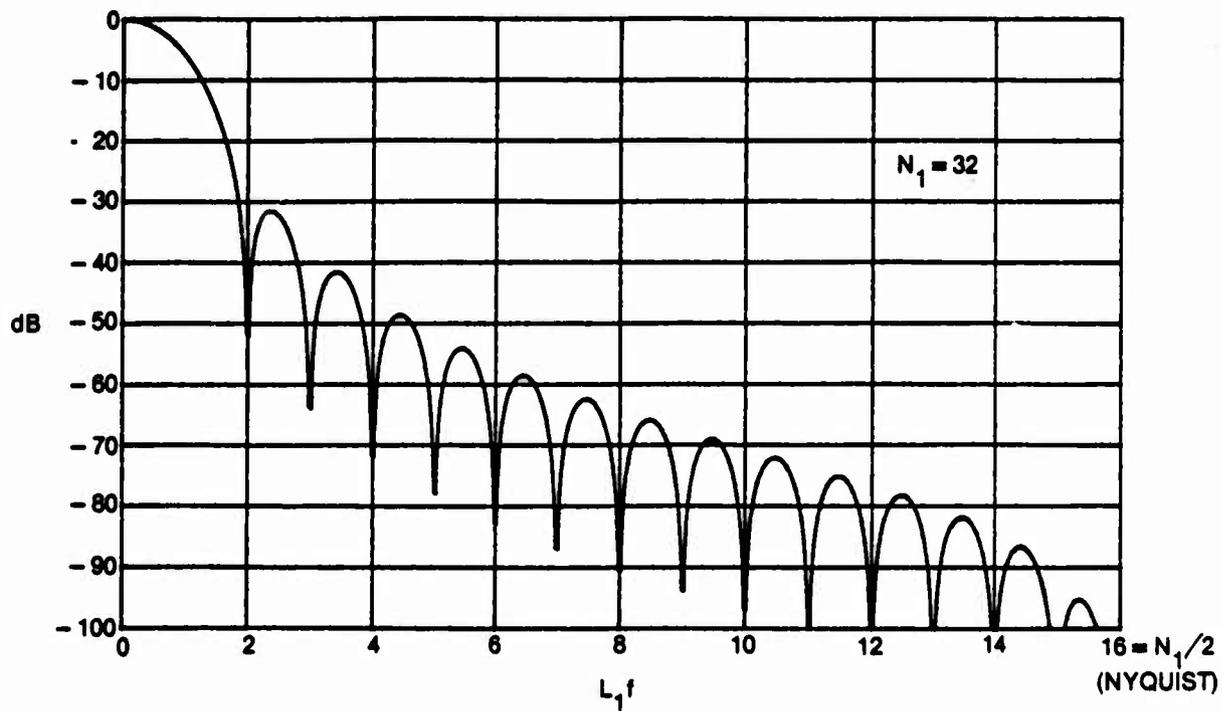
Figure 12. Effective Window for Discrete-Time Hanning Temporal Weighting

Lag Reshaping for Desired Effective Windows

From (48), the effective weighting of the generalized spectral analysis technique is given by

$$w_e(\tau) = \phi_1(\tau) \, w_2(\tau) \quad , \tag{97}$$

where $\phi_1(\tau)$ is the correlation of temporal weighting $w_1(t)$; see (44). Now suppose that for a given temporal weighting $w_1(t)$, with associated correlation $\phi_1(\tau)$, we choose lag weighting

$$w_2(\tau) = \frac{w_d(\tau)}{\phi_1(\tau)} \quad \text{for} \quad |\tau| \leq L_2 \leq L_1 \quad , \tag{98}$$

where $w_d(\tau)$ is a desirable weighting with $w_d(0) = 1$ (in keeping with (67) and (70)) and

$$w_d(\tau) = 0 \quad \text{for} \quad |\tau| > L_2 \quad . \tag{99}$$

Notice that $L_2 > L_1$ is disallowed in (98) since $\phi_1(\tau) = 0$ for $|\tau| > L_1$. Then substitution of (98) in (97) yields

$$w_e(\tau) = w_d(\tau) \quad , \quad W_e(f) = W_d(f) \quad . \tag{100}$$

That is, the effective weighting and window are equal to the desired behavior. We have "undone" the effects of bad side lobes in temporal window $W_1(f)$ by reshaping according to lag weighting $w_2(\tau)$ in (98). (The effect on the variance of the second-stage spectral estimate $\hat{G}_2(f)$ will be considered later.)

To see how much can be accomplished by this approach, some attainable effective windows that can be realized via lag reshaping, for continuous rectangular temporal weighting, are given in figure 13 for the largest possible value of L_2, namely, $L_2 = L_1$. Superposed on the window $W_1^2(f)$ for rectangular temporal weighting are the effective windows for four candidate lag reshapings, for $L_2 = L_1$. These are the narrowest possible effective windows for a *given* L_1. The first one in figure 13A corresponds to an effective Hanning weighting. The peak side lobe is only reduced from –13.3 dB to –15.7 dB, and the asymptotic decay is improved to 9 dB/octave from 6 dB/octave. The main lobe width is only slightly broadened.

Much greater improvements in side lobe behavior are possible with other lag weightings, and are illustrated in parts (B)-(D) of figure 13. They illustrate, respectively, peak side lobe levels and decays of: –30.5 dB, 21 dB/octave; –41.3 dB, 15 dB/octave; and –46.7 dB, 9 dB/octave. The deeper peak side lobe is realized at the expense of a slower asymptotic decay. They all have about the same main lobe width. The C5, C3, C1 weightings were introduced and explained in (60).
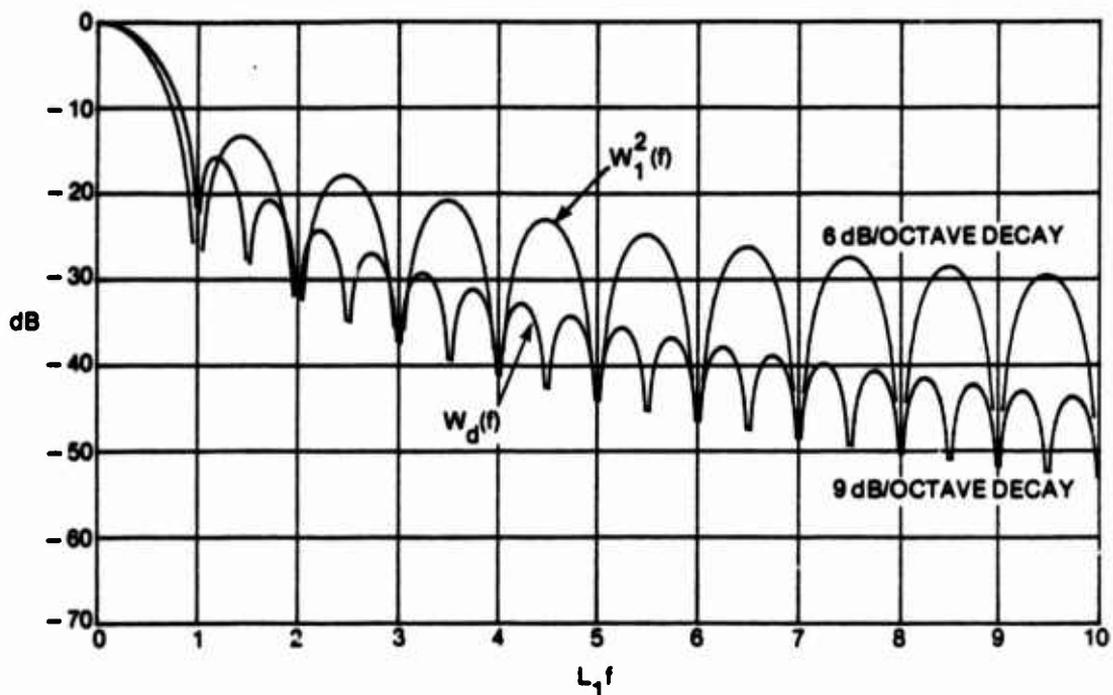
13A. Reshaping to Achieve Hanning Weighting
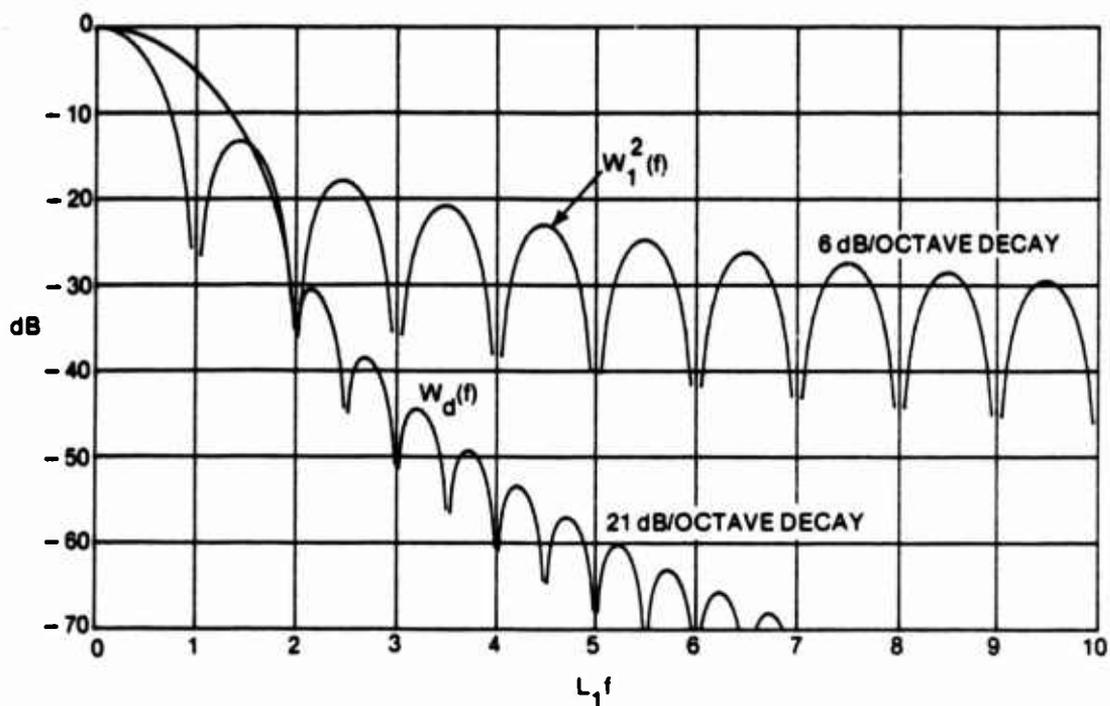
13B. Reshaping to Achieve C5 Weighting

Figure 13. Some Attainable Effective Windows via Lag Reshaping
for Rectangular Temporal Weighting and $L_2 = L_1$
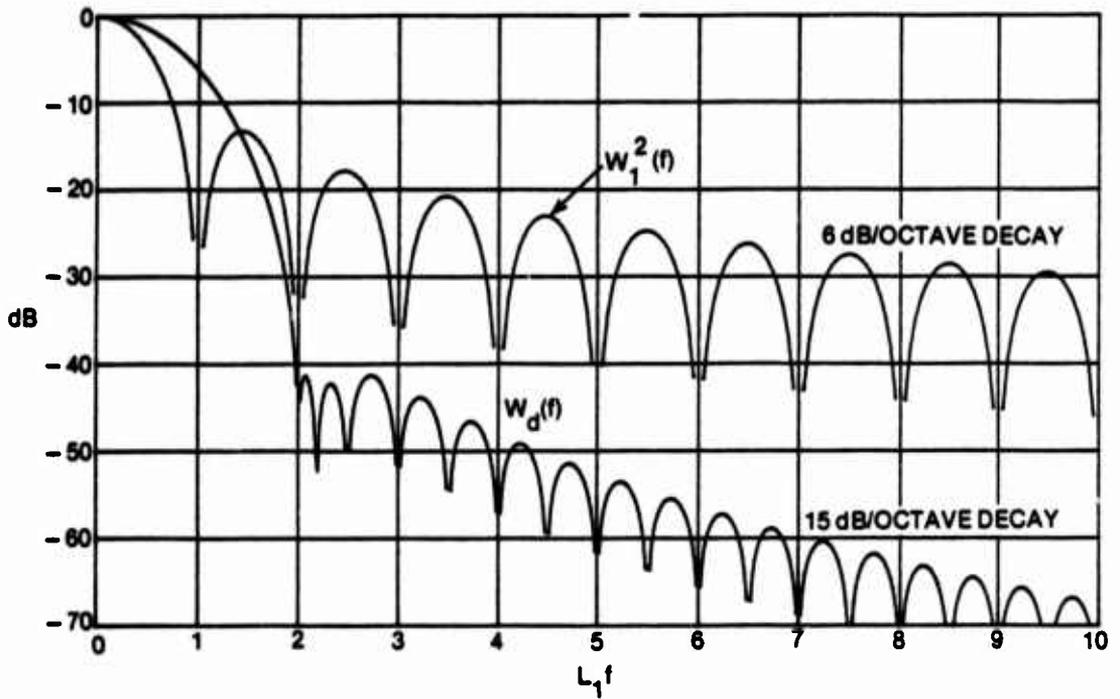
13C. Reshaping to Achieve C3 Weighting
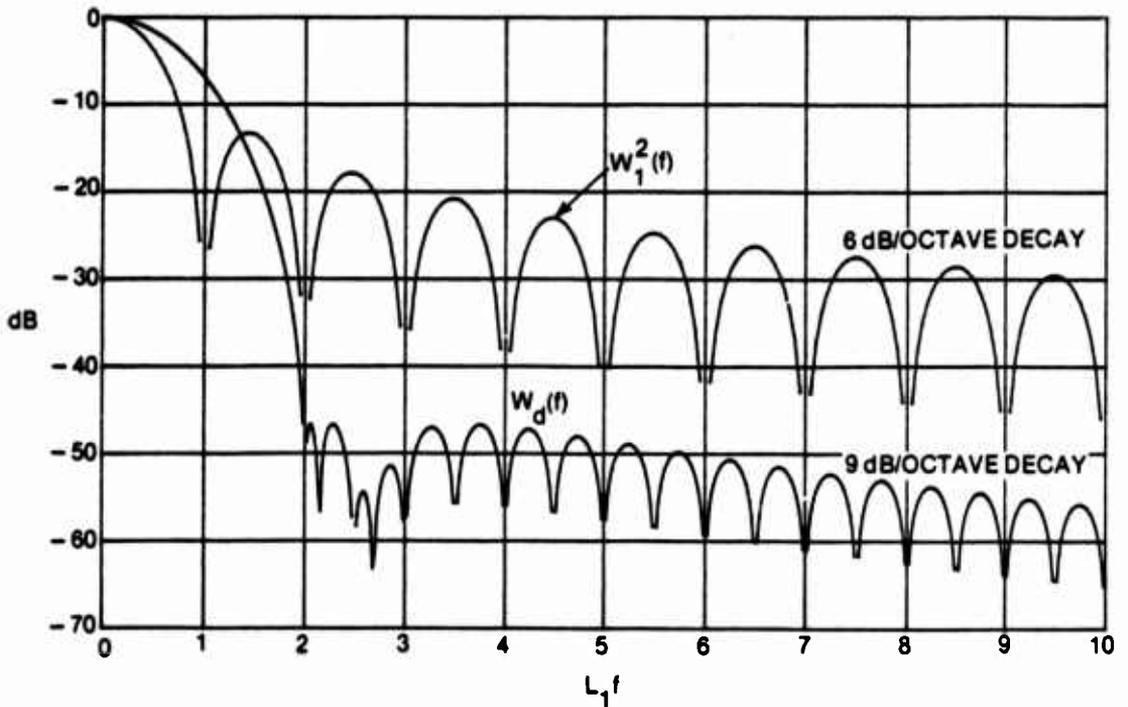
13D. Reshaping to Achieve C1 Weighting

Figure 13. Some Attainable Effective Windows via Lag Reshaping for
Rectangular Temporal Weighting and $L_2 = L_1$ (Cont'd)

Figure 13 illustrates how advantageous the reshaping technique can be in terms of peak side lobe and asymptotic decay, although the main lobe width is significantly increased. In fact, the peak side lobe at $f \cong 1.5/L_1$ for the rectangular window is really not suppressed, so much as it is smeared out; however, the other peaks of $W_1^2(f)$ for $|f| > 2/L_1$ are indeed significantly reduced. Thus reduction of leakage via lag reshaping is a very effective method, *provided* that we accept the nearest side lobe of the temporal window; this conclusion is in contrast to reference 20, page 57. These general conclusions on lag reshaping hold also for temporal weightings other than rectangular, although the exact degree of improvement will be different.

If L_2 is chosen less than L_1, the effective windows in figure 13 are simply broadened according to the ratio L_1/L_2. The peak side lobe levels and asymptotic decay are unchanged, but the main lobe width is increased. Here we are presuming L_1 fixed and decreasing L_2.

If we insist that the combination of temporal weighting $w_1(t)$ and lag reshaping $w_2(\tau)$ in (98) have effective bandwidth B_e, then use of (52), (99)-(100), and (57)-(58) yields

$$B_e = \frac{w_e^2(0)}{\int d\tau\, w_e^2(\tau)} = \frac{w_d^2(0)}{\int d\tau\, w_d^2(\tau)} = \frac{1}{2L_2\, c\{w_d\}} \quad ,$$

(101)

where $c\{w_d\}$ is the shape factor of $w_d(\tau)$ (see table 1). Thus

$$B_e L_2 = \frac{1}{2\, c\{w_d\}} \quad \text{for} \quad L_2 \leq L_1 \leq T \quad ;$$

(102A)

the limits on L_1 in (102A) follow from (98) and (15).

A plot of the interrelationship between L_1 and L_2 (introduced in figure 5) is shown in figure 14 for the case of lag reshaping. The reason that the plot is flat, in contrast to figures 5 and 6, is that the *shape* of $w_2(\tau)$ now changes as L_1 changes. This behavior is discernible from (98), since the denominator varies while the numerator remains fixed according to the selection of $w_d(\tau)$ and its associated bandwidth-length factor (102A).

If the maximum segment length, L_1, is specified (as for example when the maximum FFT size and the time-sampling spacing Δ_t are fixed), the condition under which a desired effective frequency resolution, B_d, can be met is given by figure 14. Namely, we see that

$$B_e \geq \frac{1}{2\, c\{w_d\}\, L_1} \quad .$$

(102B)

Thus if desired resolution B_d is greater than or equal to the right-side of (102B), there exists a choice for segment length L_1 that will yield the desired frequency resolution. The shape factor in (102B) depends only on the desired weighting $w_d(\tau)$. (See (65) and the accompanying discussion for the case where lag reshaping is not employed.)
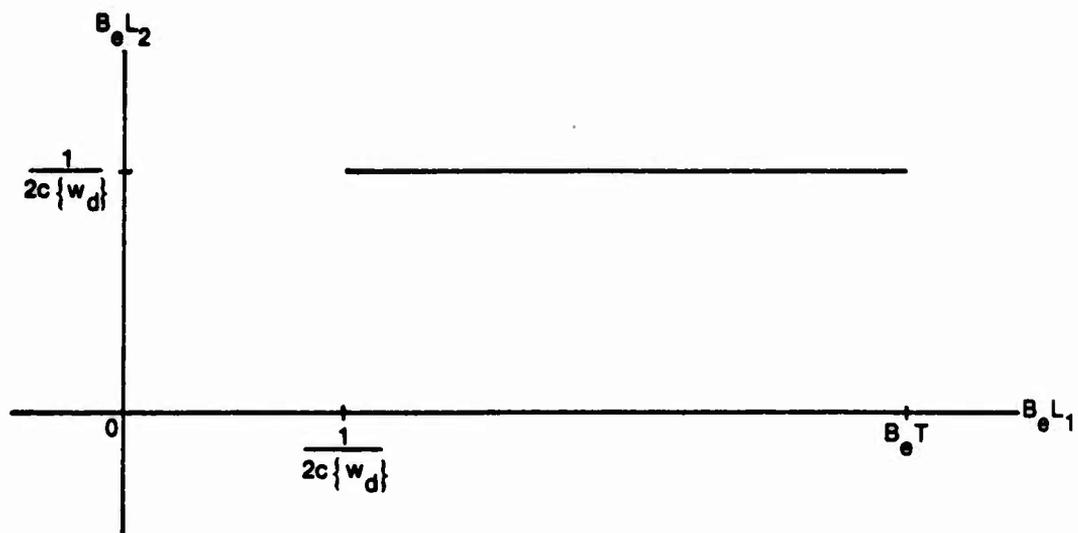
Figure 14. B_eL_2 vs B_eL_1 Plot for Lag Reshaping to Desired Weighting $w_d(\tau)$

Variance of Spectral Estimate

Up to this point, we have presumed nothing about the process x(t) except that it be (second-order) stationary during the observation interval T. Now we make a couple of assumptions about the process in order to obtain manageable expressions for the variance of the second-stage spectral estimate $\hat{G}_2(f)$, at frequency f. Our first assumption is that the true spectrum G of x(t) varies slowly in the neighborhood of the frequency of interest, f. More precisely, from (50) and (70), we obtain, for the mean spectral estimate,

$$Av\{\hat{G}_2(f)\} = \int du \; G(f - u) \; W_e(u) \approx G(f) \int du \; W_e(u) = G(f) \quad , \tag{103}$$

where we assume that spectrum G is relatively constant in the frequency band $(f-B_e/2, \; f+B_e/2)$; i.e., the only region where effective window W_e in (103) is substantially nonzero is in the range $(-B_e/2, \; B_e/2)$.

Our second assumption is that x(t) is a complex Gaussian process. The variance of $\hat{G}_2(f)$ is developed under this assumption in appendix D, culminating in the exact result in (D-13):

$$Var\{\hat{G}_2(f)\} = \iint d\alpha \; d\beta \; G(\alpha) \; G(\beta) \; |\gamma(f - \alpha, \; f - \beta)|^2 \; Q_p(S(\alpha - \beta)) \quad , \tag{104}$$

where window convolution function

$$\gamma(x, \; y) \equiv \int du \; W_2(u) \; W_1(x - u) \; W_1^*(y - u) \tag{105}$$

and periodic function

$$Q_P(u) \equiv \left[\frac{\sin(\pi P u)}{P \; \sin(\pi u)} \right]^2 \quad . \tag{106}$$

The variance result in (104) does not require that spectrum G vary slowly in the neighbor of f; the result utilizes only the Gaussian assumption on the process x(t). The temporal and lag windows contribute through the window convolution function γ, while the shift S and number of pieces P appear through the periodic function Q_P.

When the assumption regarding a slowly varying spectrum G in the neighborhood of frequency f of interest is also invoked, (104) simplifies to forms given in (D-20) and (D-24); the latter is a "weighting domain" version of the variance:

$$Var\{\hat{G}_2(f)\} = G^2(f) \; \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P} \right) \int d\tau \; w_2^2(\tau) \; \phi_3(\tau, \; pS) \quad , \tag{107}$$

where

$$\phi_3(\tau,\mu) \equiv \int dt \, w_1\left(t + \frac{\mu - \tau}{2}\right) w_1\left(t + \frac{-\mu - \tau}{2}\right) w_1\left(t + \frac{\mu + \tau}{2}\right) w_1\left(t + \frac{-\mu + \tau}{2}\right)$$

$$= \phi_3(\underline{+}\tau, \, \underline{+}\mu) \tag{108}$$

is a third-order correlation of temporal weighting $w_1(t)$. The form (107) is very useful if ϕ_3 can be evaluated in closed form. An "ambiguity domain" version of the variance is given by (D-20).

Quality Ratio

The quality ratio for spectral estimation was defined in (4). With the aid of (103) and (107), it is given by

$$Q = \frac{Var\{\hat{G}_2(f)\}}{Av^2\{\hat{G}_2(f)\}}$$

$$= \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \int d\tau \; w_2^2(\tau) \; \phi_3(\tau, pS) \quad . \tag{109}$$

Since the smallest possible value of Q is $(TB_e)^{-1}$ (see (11) et seq.), the normalized quality ratio is

$$NQR \equiv Q \cdot TB_e = \frac{TB_e}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \int d\tau \; w_2^2(\tau) \; \phi_3(\tau, pS) \quad . \tag{110}$$

This quantity can never be smaller than unity.

If we employ (52) and the normalizations (67) and (69), the convenient form

$$NQR = T \frac{\frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \int d\tau \; w_2^2(\tau) \; \phi_3(\tau, pS)}{\int d\tau \; w_2^2(\tau) \; \phi_1^2(\tau)} \tag{111}$$

for the normalized quality ratio is obtained. We are interested in the behavior of the normalized quality ratio for different choices of P, S, $w_1(t)$, and $w_2(\tau)$. The constraint of a fixed effective bandwidth B_e has been injected into the normalized quality ratio via the use of (52) in (111). The quantities ϕ_1 and ϕ_3 needed in (111) are given by (44) and (108) respectively.

Before we embark on particular cases, some general observations on overlap (shift S) are in order. For a minimum normalized quality ratio (minimum variance) with each temporal weighting $w_1(t)$, we should use approximately the optimum overlap as derived in reference 9. For example, Hanning temporal weighting should be employed with approximately 62 percent overlap, although there is only an 8 percent loss in stability if 50 percent overlap is used for convenience (reference 9, tables 5 and 6). There is no point in considering excessive or inadequate overlap, since this leads to excessive computational effort or more variance, respectively. Inadequate temporal overlap cannot be made up, in terms of variance reduction, by any amount of quadratic smoothing. This can be seen by observing that poor first-stage correlation estimates $\hat{R}_1(\tau)$ are merely multiplied by lag weighting $w_2(\tau)$, and are not improved statistically in any way for $|\tau| < L_2$; those estimates for $|\tau| > L_2$ are discarded by the lag weighting.

Some related work on the effects of windowing on stability is given in references 21 and 22. However, the present report is more thorough and detailed in its treatment of the problem and the inclusion of a bandwidth constraint.

Special Cases of Generalized Spectral Analysis Technique

This section will consider several special cases of the normalized quality ratio and show how some earlier results are obtained as limiting cases. The next section will treat the generalized spectral analysis results.

One Piece, $P = 1$

When only one piece is used in the first-stage spectral estimate $\hat{G}_1(f)$ in (14), we have a generalized version of the Blackman-Tukey approach, in that the data $x(t)$ are weighted by $w_1(t)$ prior to computing the sample correlation; see (16) and (13). Also, we allow length $L_1 < T$ (although we soon show that the best L_1 is equal to T, the available record length). From (111),

$$NQR_1 = T \frac{\int d\tau \; w_2^2(\tau) \; \phi_2(\tau)}{\int d\tau \; w_2^2(\tau) \; \phi_1^2(\tau)} \quad , \tag{112}$$

where (using (108))

$$\phi_2(\tau) \equiv \phi_3(\tau, 0) = \int dt \; w_1^2\left(t + \frac{\tau}{2}\right) w_1^2\left(t - \frac{\tau}{2}\right) = \int dt \; w_1^2(t) \; w_1^2(t - \tau) \tag{113}$$

is the correlation of the squared temporal weight function $w_1^2(t)$.

Now if $L_1 \gg L_2$, w_2 is much narrower than ϕ_1 or ϕ_2. In that case, the exact shape of w_2 is irrelevant, and (112), (113), (44), figure 2, and Schwarz's inequality yield

$$NQR_1 \approx T \frac{\phi_2(0)}{\phi_1^2(0)} = T \frac{\int dt \; w_1^4(t)}{\left[\int dt \; w_1^2(t)\right]^2} \geq \frac{T}{L_1} \quad \text{for} \quad L_1 \gg L_2 \quad . \tag{114}$$

Equality in (114) results if and only if $w_1^2(t)$ is constant for $|t| < L_1/2$; furthermore, the best value for L_1 is then its largest allowed value T (see (15)), in which case we have Blackman-Tukey processing and

$$NQR_1 (\text{rectangular } w_1) \approx 1 \quad \text{for} \quad T = L_1 \gg L_2 \quad . \tag{115}$$

This result agrees with reference 2, section B.8. It should be noted that $L_1 \gg L_2$ implies $B_e L_1 \gg B_e L_2 \geq .5/c\{w_2\} \sim 1$, according to figure 5 and table 1; thus stable estimates result in this case.

Instead of rectangular temporal weighting, consider Hanning weighting:

$$w_1(t) = \left(\frac{8}{3L_1}\right)^{\frac{1}{2}} \cos^2\left(\frac{\pi t}{L_1}\right) \quad \text{for} \quad |t| < L_1/2 \quad . \tag{116}$$

Then

$$\phi_1(0) = \int dt\ w_1^2(t) = 1 \quad , \quad \phi_2(0) = \int dt\ w_1^4(t) = \frac{35}{18L_1} \quad , \tag{117}$$

and (114) yields

$$NQR_1(\text{Hanning } w_1) \approx \frac{35}{18}\ \frac{T}{L_1} \quad \text{for} \quad L_1 \gg L_2 \quad . \tag{118}$$

The best L_1 is again T; however, the minimum value of the normalized quality ratio is then 35/18, which is twice the value in (115) for rectangular temporal weighting. This is due to the squandering of the edges of the available data record by the small values there of Hanning temporal weighting.

Now instead of assuming $L_1 \gg L_2$, let us reconsider, for general w_1, L_1, L_2, the normalized quality ratio (112). Since $w_1(t)$ is zero for $|t| > L_1/2$, we have from (44),

$$\phi_1(\tau) = \int_{a(\tau)}^{b(\tau)} dt\ w_1(t)\ w_1(t-\tau) \quad \text{for} \quad |\tau| < L_1 \quad , \tag{119}$$

where

$$\left.\begin{array}{l} a(\tau) = \max(-L_1/2,\ -L_1/2+\tau) \\[2mm] b(\tau) = \min(L_1/2,\ L_1/2+\tau) \end{array}\right\} \quad \text{for} \quad |\tau| < L_1 \quad . \tag{120}$$

Then by Schwarz's inequality, (113), and (120),

$$\phi_1^2(\tau) \le \int_{a(\tau)}^{b(\tau)} dt\ w_1^2(t)\ w_1^2(t-\tau) \int_{a(\tau)}^{b(\tau)} dt\ 1$$

$$= \phi_2(\tau)\ (L_1 - |\tau|) \quad \text{for} \quad |\tau| < L_1 \quad . \tag{121}$$

Equality is realized in (121) if and only if $w_1(t)$ is constant for $|t| < L_1/2$; that is, the best temporal weighting for maximum stability is rectangular when $P = 1$. This conclusion holds regardless of the form of lag weighting $w_2(\tau)$ or the relative sizes of L_1 and L_2.

As an example, for rectangular temporal weighting,

$$w_1(t) = L_1^{-\frac{1}{2}} \quad \text{for} \quad |t| < L_1/2 \quad ,$$

$$\phi_1(\tau) = 1 - |\tau|/L_1 \quad \text{for} \quad |\tau| < L_1 \quad ,$$

$$\phi_2(\tau) = \frac{1}{L_1}(1 - |\tau|/L_1) \quad \text{for} \quad |\tau| < L_2 \quad , \tag{122}$$

and (112) yields for general $w_2(\tau)$,

$$NQR_1(\text{rect. } w_1) = \frac{T}{L_1} \frac{\int_{-L_1}^{L_1} d\tau \, w_2^2(\tau) \left(1 - \frac{|\tau|}{L_1}\right)}{\int_{-L_1}^{L_1} d\tau \, w_2^2(\tau) \left(1 - \frac{|\tau|}{L_1}\right)^2} \quad . \tag{123}$$

The ratio of integrals is obviously greater than 1. For a monotonically decreasing lag weighting $w_2(\tau)$ of fixed shape, the ratio of integrals is minimized by choosing L_1 as large as possible. Since the leading factor also has the same behavior, the best value for the normalized quality ratio is

$$NQR_1(\text{rect. } w_1) = \frac{\int_{-T}^{T} d\tau \, w_2^2(\tau) \left(1 - \frac{|\tau|}{T}\right)}{\int_{-T}^{T} d\tau \, w_2^2(\tau) \left(1 - \frac{|\tau|}{T}\right)^2} \quad \text{for} \quad L_1 = T \quad . \tag{124}$$

We cannot give numerical values to this ratio of integrals until we select a lag weighting $w_2(\tau)$ and determine the specific value of $L_2(T)$; see figures 5 and 6. But if $TB_e \gg 1$, which is the usual case for reasonably good spectral estimates, then $L_2 \ll L_1 = T$ and

$$NQR_1(\text{rect. } w_1) = 1 \quad \text{for} \quad L_1 = T, \quad TB_e \gg 1 \quad . \tag{125}$$

This result holds independently of the exact shape of the lag weighting $w_2(\tau)$; thus we could choose $w_2(\tau)$ such that the effective weighting $w_e(\tau)$ in (48) has good side lobe behavior, as discussed in an earlier section.

No Quadratic Frequency-Smoothing

No quadratic smoothing corresponds to

$$W_2(f) = \delta(f),$$

$$w_2(\tau) = 1 \quad \text{for all } \tau \quad . \tag{126}$$

Thus $L_2 = \infty$, and (109) becomes

$$Q = \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \int d\tau \, \phi_3(\tau, pS)$$

$$= \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \phi_1^2(pS) \quad , \tag{127}$$

since

$$\int d\tau \ \phi_3(\tau,\mu) = \int d\tau \ \int du \ w_1(u) w_1(u - \mu) w_1(u + \tau) \ w_1(u + \tau - \mu)$$

$$= \int du \ w_1(u) \ w_1(u - \mu) \ \phi_1(\mu) = \phi_1^2(\mu) \tag{128}$$

by use of (108). The result in (127) is identical to reference 9, equation 8, when we recall definition (109) and normalization (67).

Non-Overlapping Segments

Let us choose time shift S in (13) equal to the segment length L_1; this leads to abutting time segments. From (15), we have

$$PL_1 = T \quad , \tag{129}$$

where we have chosen to use up all of the available data length. (This is different from the earlier subsection for $P = 1$ where we allowed $L_1 < T$.) The general normalized quality ratio in (111) reduces to

$$NQR_p = L_1 \ \frac{\int d\tau \ w_2^2(\tau) \ \phi_2(\tau)}{\int d\tau \ w_2^2(\tau) \ \phi_1^2(\tau)} \quad , \tag{130}$$

where we used (129), the fact that $w_1(t)$ is of length L_1, and (113).

Once again, we refer to bound (121) and the fact that equality is realized only for a flat weighting $w_1(t)$. Thus, from (130) and (122),

$$NQR_p(\text{rect. } w_1) = \frac{\int_0^{L_1} d\tau \ w_2^2(\tau) \ \left(1 - \frac{\tau}{L_1}\right)}{\int_0^{L_1} d\tau \ w_2^2(\tau) \ \left(1 - \frac{\tau}{L_1}\right)^2} \tag{131}$$

for any (real symmetric) lag weighting $w_2(\tau)$. The ratio of integrals is obviously always greater than unity; therefore, for a monotonically decreasing lag weighting $w_2(\tau)$ of fixed shape, values of L_1 large in comparison with L_2 are preferred. However, $L_1 \gg L_2$ means that

$$TB_e = PL_1 \ B_e \gg PL_2 \ B_e \geq \frac{P}{2c(w_2)} \sim P \quad , \tag{132}$$

according to (129) and figure 5. Thus large time-bandwidth products, TB_e, are required; also P must be kept small enough to realize $L_1 \gg L_2$. In this case, we have

$$NQR_p(\text{rect. } w_1) = 1 \quad \text{for} \quad TB_e \gg 1, \quad L_1 \gg L_2 \quad , \tag{133}$$

regardless of lag weighting $w_2(\tau)$. Qualitatively, when $L_1 \gg L_2$, the edge effects of segmenting the data x(t) are negligible, since only a small fraction of the utilized correlation values that can be calculated from a record of length T are neglected when using segments of length L_1. Stated alternatively, all the first-stage correlation estimates that are used, namely $\hat{R}_1(\tau)$ for $|\tau| < L_2$, have the same quality (stability) when $L_2 \ll L_1$.

An example of the exact normalized quality ratio for Hanning lag weighting and rectangular temporal weighting is afforded by substituting the equation

$$w_2(\tau) = \cos^2\left(\frac{\pi}{2}\frac{\tau}{L_2}\right) \quad \text{for} \quad |\tau| < L_2 \tag{134}$$

into (131) (see figure 7A for the effective window):

$$NQR_p(\text{rect. } w_1, \text{Hann. } w_2) = \frac{\int_0^y dx \, \cos^4 x \left(1 - \frac{2}{\pi}\frac{L_2}{L_1}x\right)}{\int_0^y dx \, \cos^4 x \left(1 - \frac{2}{\pi}\frac{L_2}{L_1}x\right)^2}, \tag{135}$$

where

$$y = \frac{\pi}{2}\min\left(1, \frac{L_1}{L_2}\right). \tag{136}$$

Equation (135) is plotted* as the top curve in figure 15A. As expected, the normalized quality ratio tends to 1 as L_2/L_1 tends to zero. But even for as large a value as $L_2/L_1 = .5$, the normalized quality ratio has increased only by 12 percent. Thus the penalty in increased variance, for not realizing a small ratio for L_2/L_1, is not severe.

Also plotted in figure 15A is the normalized quality ratio for the three lag weightings introduced in (60) et seq. They all lead to smaller values of the normalized quality ratio, for the same value of L_2/L_1; in fact, lag weighting C_5 incurs only a 7 percent increase in variance when $L_2/L_1 = .5$, in relation to the ideal value 1. The reason that the normalized quality ratio is lower is due to the fact that the lag weightings drop to zero faster within their length L_2.

Non-Overlapping Segments; Lag Reshaping

The possibilities of lag reshaping have been discussed earlier with regard to the mean of the spectral estimate and the effective window. We now want to see what effect lag reshaping has on the normalized quality ratio in (130). Substitution of (98) in (130) yields

$$NQR_p(\text{lag reshaping}) = L_1 \frac{\int d\tau \, w_d^2(\tau) \frac{\phi_2(\tau)}{\phi_1^2(\tau)}}{\int d\tau \, w_d^2(\tau)} \quad \text{for} \quad L_2 \leq L_1. \tag{137}$$

---

*The quantity TB_e is not involved in figure 15; some related computational considerations are discussed in appendix E.
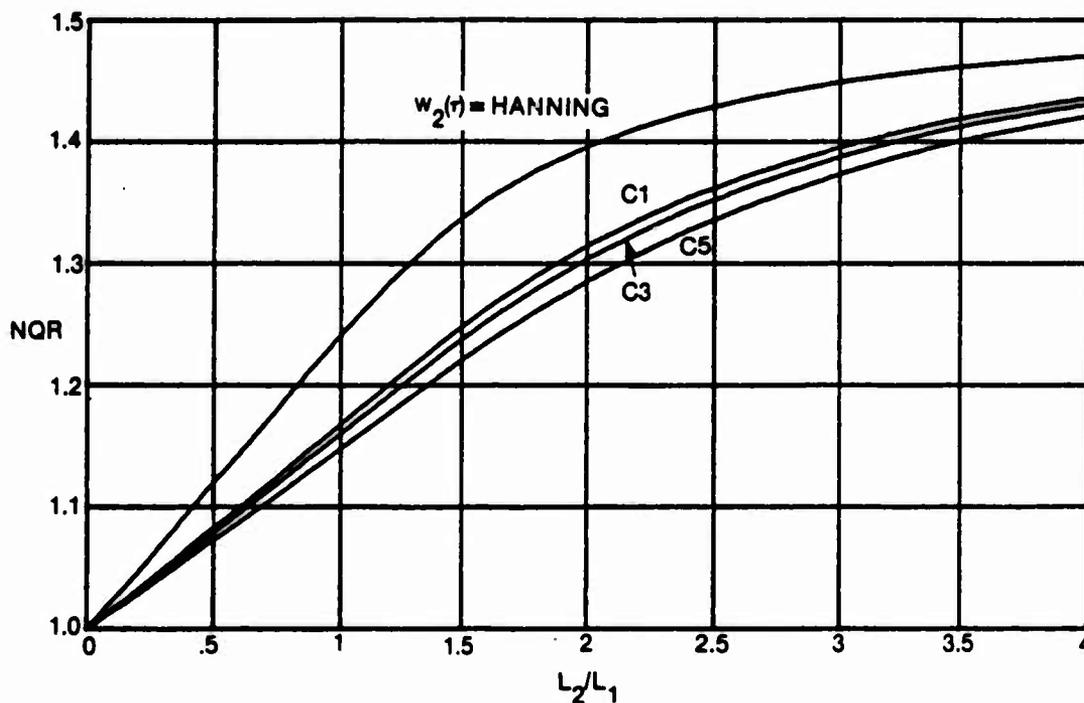
15A. Various Lag Weightings $w_2(\tau)$
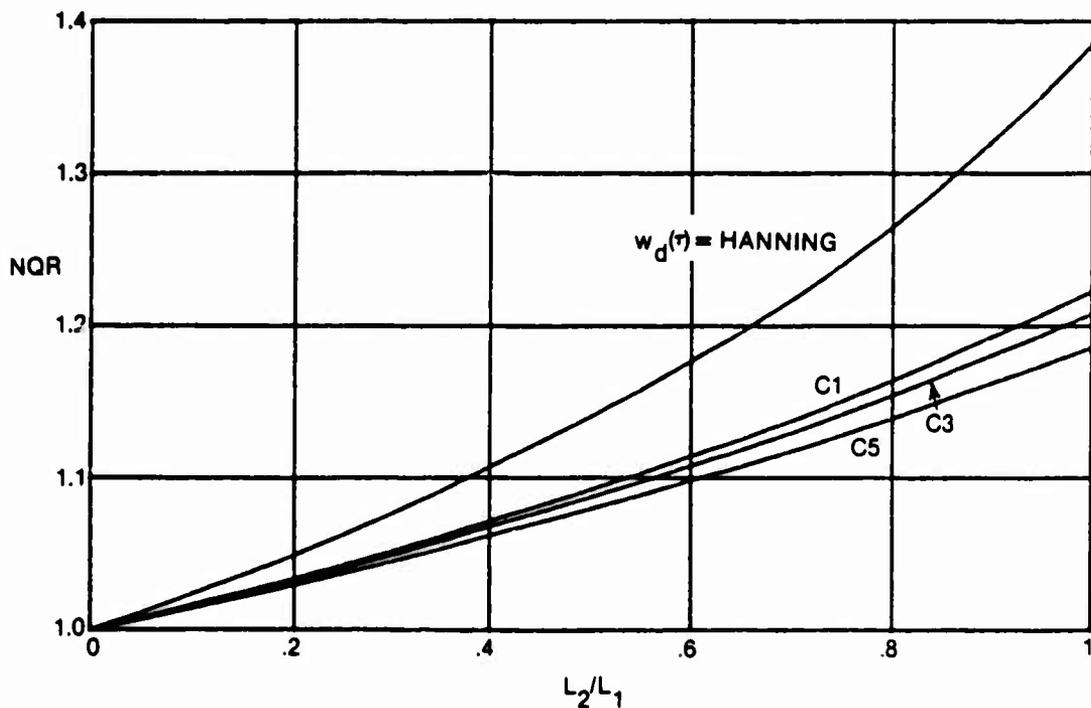
15B. Reshaping to Desired Lag Weighting $w_d(\tau)$

Figure 15. Normalized Quality-Ratio for Rectangular Temporal
Weighting and No Overlap

The special case of rectangular temporal weighting is obtained by employing (122) in (137):

$$NQR_p(\text{rect. } w_1, \text{ lag reshaping}) = \frac{\int_0^{L_2} d\tau \; w_d^2(\tau) \left(1 - \frac{\tau}{L_1}\right)^{-1}}{\int_0^{L_2} d\tau \; w_d^2(\tau)} \quad \text{for} \quad L_2 \leq L_1 \quad .$$

(138)

The division by $\phi_1(\tau)$ in reshaping (98) increases the variance (for a *specified* L_2/L_1 and for $w_d(\tau) = w_2(\tau)$) above that in (131), because we are more heavily weighting regions where the denominator in (138) is smaller.

Equation (138) is plotted in figure 15B for desired effective weightings of Hanning, C1, C3, and C5. Notice that the abscissa is now limited to $L_2/L_1 \leq 1$. As expected, the normalized quality ratio tends to 1 as L_2/L_1 tends to zero; that is, we can do lag reshaping for good side lobe behavior and lose little in terms of stability, provided that L_1 is chosen sufficiently larger than L_2. Of course, the normalized quality ratio values in figure 15B are larger than those in figure 15A, for the same value of L_2/L_1. As an example, for desired effective weighting C_1, if we take $L_1 = 2L_2$, the increase in variance over the ideal value is only 9 percent. Thus lag reshaping is an attractive procedure for spectral estimation; recall from figure 14 that L_2 is set by the specified B_e and the shape of $w_d(\tau)$.

General Results on Stability

We now return to the general normalized quality ratio in (111) and recall constraint (15). We will select time shift S according to

$$S = qL_1 \qquad . \tag{139}$$

where $q = q\{w_1\}$ is a fraction specified to be in the range $(0, 1]$ and is dependent on the particular temporal weighting $w_1(t)$ employed. The observations made in the paragraph following (111) are relevant in this regard. For example, with no quadratic frequency smoothing and with Hanning temporal weighting, $q = .39$ (61 percent overlap) is virtually optimum, although $q = 1/2$ loses only 8 percent in variance-reduction capability (reference 9, tables 4-7). We also select equality in (15) so as to make maximum use of the available record length, i.e., minimum variance of the spectral estimate. Then we have

$$L_1 = \frac{T}{1 + (P - 1)q} \qquad (P \geq 1) \qquad . \tag{140}$$

Thus for a given T and specified shift fraction q, L_1 can take on only a discrete set of values.

Arbitrarily large values of P are not allowed in (140), because this would result in such small values of L_1 that the bandwidth constraint, B_e, could not be met. From figure 5, the lower bound on $B_e L_1$ limits

$$P_{max} = 1 + \frac{2c\{\phi_1\} B_e T - 1}{q} \qquad . \tag{141}$$

(Actually, P_{max} must be the integer part of the right-hand side.) Thus P_{max} depends on the temporal weighting $w_1(t)$ directly through its shape factor $c\{\phi_1\}$ and indirectly via the selected shift fraction $q = q\{w_1\}$. For $q = 1$, no overlap, (141) reduces to (E-1).

When $P = P_{max}$, L_1 is at its minimum value, and L_2 must be greater than T; it can be ∞ (see figures 5 and 6A). In this case, there is no quadratic frequency smoothing, and we have the situation studied in detail in references 9 and 12, and mentioned earlier in (126)-(128). At the other extreme, when $P = 1$, we have Blackman-Tukey processing with the generalization that the temporal weighting need not be rectangular; this case was considered in the previous section. The range of values of L_2/L_1 is shown in (E-2) and (E-3) to be very wide when $B_e T \gg 1$, which is a usual practical case.

More generally, for P in the range [1, P_{max}], we can investigate the tradeoff between the amounts of temporal- and lag-weighting, for specified resolution B_e and for specified weighting shapes of interest. Below, we consider the two cases of rectangular temporal weighting and Hanning temporal weighting.

Rectangular Temporal Weighting

It was shown earlier in figure 15 that rectangular temporal weighting with no overlap results in small values for the normalized quality ratio, whether the lag weighting is reshaped or not, provided that L_2 is chosen somewhat smaller than L_1. Now the question arises as to whether one should use any overlap, such as 50 percent, with rectangular temporal weighting.

We presume $B_e T \gg 1$. For no overlap, the estimate of first-stage correlation $R_1(\tau)$ at $\tau \approx L_1/2$ has only half the degrees of freedom as the estimate at $\tau = 0$. But with 50 percent overlap, the degrees of freedom for estimation at $\tau \approx L_1/2$ are about the same as at $\tau = 0$. This is why 50 percent overlap for rectangular temporal weighting appears attractive.

However, for estimation of $R_1(\tau)$ at $\tau > L_1/2$, we still do not get as many degrees of freedom as for Blackman-Tukey processing, because some data points never interact. For example, although at $\tau = 3L_1/4$ we have doubled the degrees of freedom by using 50 percent overlap, we still have only about half of the number that are available at this τ value via Blackman-Tukey processing.

In order to ascertain quantitatively the merit of overlapping for rectangular temporal weighting, we have evaluated the normalized quality ratio (111) for lag reshaping to realize a desired effective weighting equal to C1 as given in (60) and (61) (reference 19, figure 12). That is, in (111), we use

$$w_2(\tau) = \frac{w_d(\tau)}{\phi_1(\tau)} \quad \text{for} \quad |\tau| < L_2 \leq L_1 \quad , \tag{142}$$

where $\phi_1(\tau)$ is given by (122). In addition, we need the third-order correlation (108), which is

$$\phi_3(\tau,\mu) = \frac{1}{L_1}\left(1 - \frac{|\tau| + |\mu|}{L_1}\right) \quad \text{for} \quad |\tau| + |\mu| < L_1 \tag{143}$$

for rectangular temporal weighting.

For the two cases of $B_e T = 100$ and $B_e T = 1000$, the normalized quality ratio has been evaluated for $q = 1$, .75, and .5, and plotted in figure 16. The explanation of the behavior of the curves is as follows: ·

$q = 1$ No Overlap

If $P = 1$, then $L_2 \ll L_1$ and it follows that for all $|\tau| < L_2$, $R_1(\tau)$ is estimated with virtually the same degrees of freedom as at $\tau = 0$, where we have the maximum degrees of freedom possible to estimate $R_1(0)$. As P increases toward P_{max}, then L_1 tends to L_2. Now $R_1(0)$ is still estimated with the full degrees of freedom, but $R_1(\tau)$ for $\tau \neq 0$ is estimated with fewer degrees of freedom. For τ near L_2, the loss in degrees of freedom in estimation of $R_1(\tau)$ is significant, and the variance increases.

16A. $B_e T = 100$
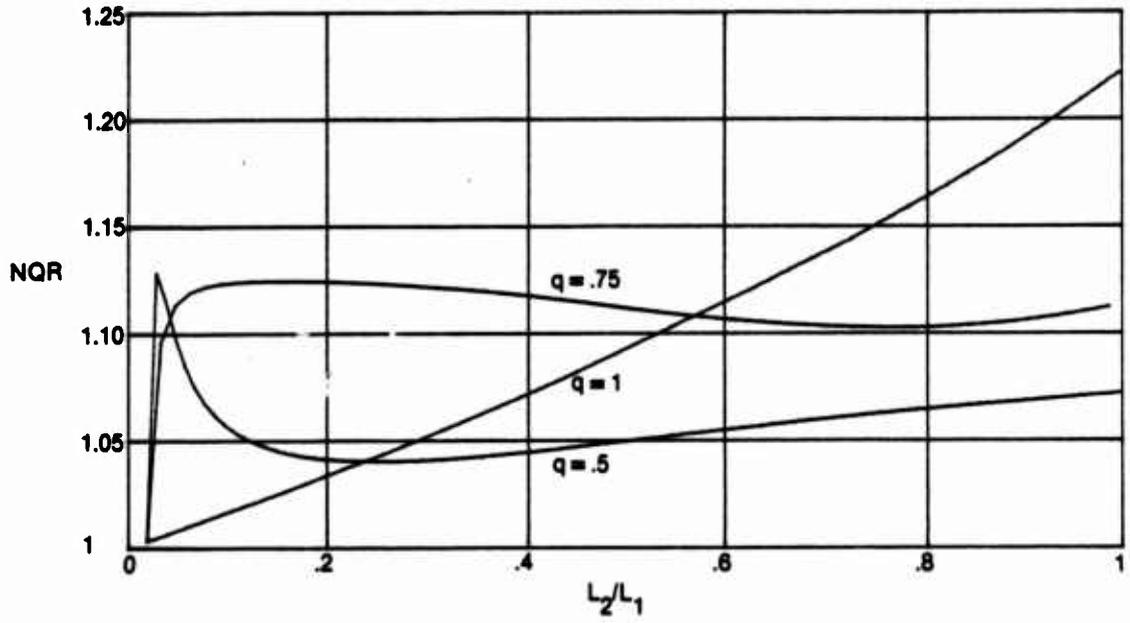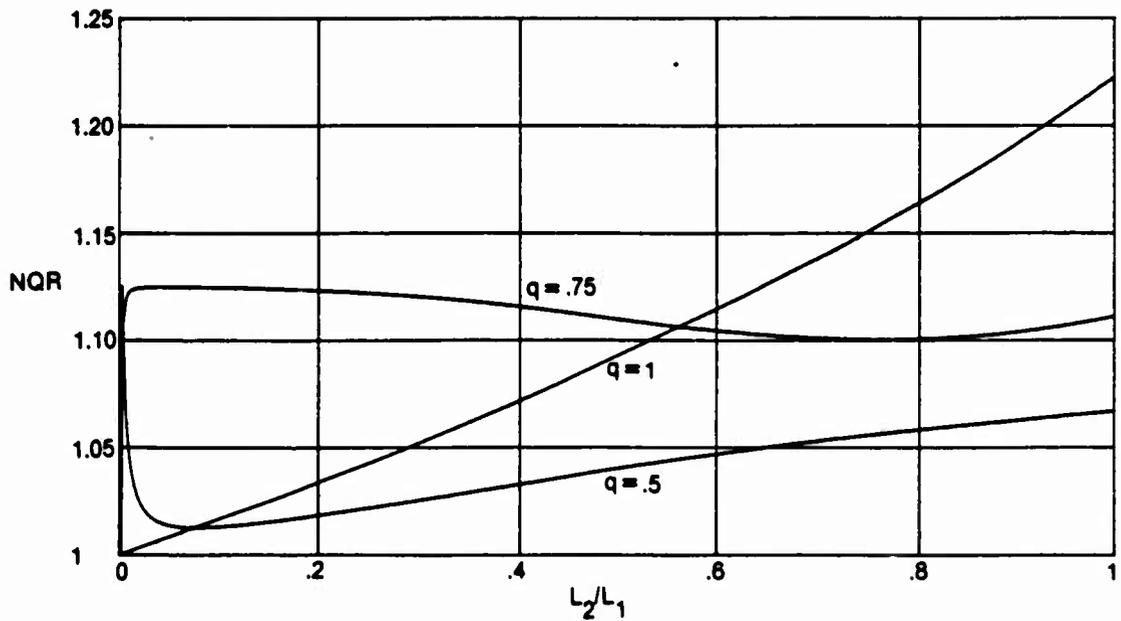
16B. $B_e T = 1000$

Figure 16. Normalized Quality-Ratio for Overlapped Rectangular
Temporal Weighting and Lag Reshaping to C1

q = .5   50 Percent Overlap

For $P = 1$, there is no overlap and conditions are identical to those described above. For $P = 2$, the sudden increase in variance can be explained as follows: from (16), the first-stage estimate is

$$\hat{R}_1(\tau) = \frac{1}{2} \left[\int dt \; y_0(t) \; y_0^*(t - \tau) + \int dt \; y_1(t) \; y_1^*(t - \tau) \right] \quad . \quad (144)$$

In particular,

$$\hat{R}_1(0) = \frac{1}{2} \left[\int dt \; |y_0(t)|^2 + \int dt \; |y_1(t)|^2 \right]$$

$$= \frac{1}{T} \int_0^T dt \; x^2(t) \; 0(t) \quad , \quad (145)$$

where the overall weighting $0(t)$ of $x^2(t)$ is depicted in figure 17. As shown, the overall weighting is very uneven, causing loss in stability. As P increases above 2, the uneveness of the overall weighting (for $q = .5$) occurs only towards the edges of the $(0, T)$ interval, yielding a decrease of variance, since more data points tend to get the same overall weighting, insofar as their effect upon the estimation of $R_1(\tau)$ is concerned. However, at the same time, the effect of fewer degrees of freedom in estimation of $R_1(\tau)$, for τ values near L_2, becomes more pronounced as P increases and L_1 decreases; this is true even for the 50 percent overlap case being considered here. Eventually, this effect dominates, and the variance increases with P.
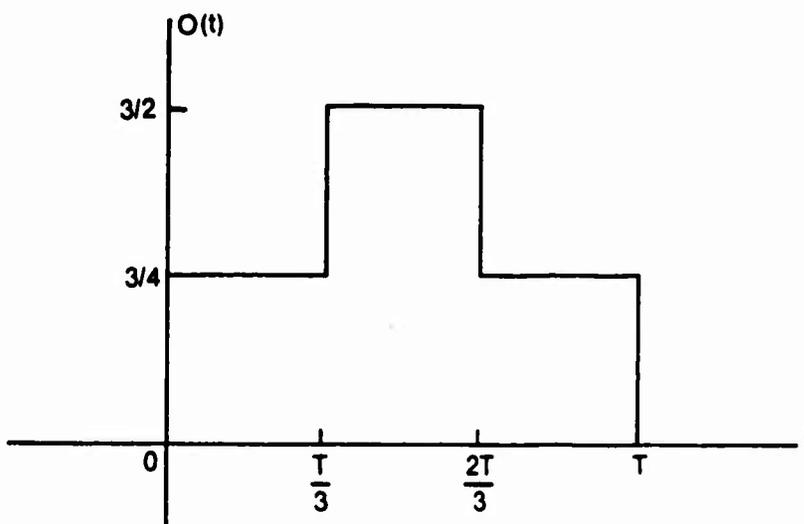
Figure 17. Overall Weighting of $x^2(t)$ for $q = .5$, $P = 2$, Rectangular $w_1(t)$

q = .75   25 Percent Overlap

For *any* value of $P > 1$, the overall weighting $O(t)$ is very jagged (as above), and the jaggedness does *not* decrease or concentrate near the edges as P increases. This is true for any overlap greater than 0 and less than 50 percent.

In summary, for rectangular temporal weighting, the smallest values for the normalized quality ratio are realized by choosing $q = 1$, no overlap, and making L_1 several times larger than L_2. This conclusion about the ratio L_2/L_1 is consistent with those reached earlier.

Hanning Temporal Weighting

The temporal weighting and associated correlation for this case are given by (116), (C-9), and (C-10):

$$w_1(t) = \left(\frac{8}{3L_1}\right)^{\frac{1}{2}} \cos^2\left(\frac{\pi t}{L_1}\right) \quad \text{for} \quad |t| < \frac{L_1}{2} \quad,$$

$$\phi_1(\tau) = \left(1 - \frac{|\tau|}{L_1}\right)\left[\frac{2}{3} + \frac{1}{3}\cos\left(\frac{2\pi\tau}{L_1}\right)\right] + \frac{1}{2\pi}\sin\left(\frac{2\pi|\tau|}{L_1}\right) \quad \text{for} \quad |\tau| < L_1 \quad.$$

$$(146)$$

Evaluation of third-order correlation $\phi_3(\tau, \mu)$ in (108) is rather tedious; the end result is given in (F-1) and (F-2). The procedure and program for the evaluation of the normalized quality ratio is given in appendix F.

The normalized quality ratio for Hanning lag weighting and $B_eT = 100$ is plotted in figure 18 for several values of the shift fraction q. When $q = 1$, no overlap, the small values of the Hanning temporal weighting at its edges cannot be compensated for, by any choice of L_2/L_1, and the variance remains at approximately twice the ideal value. For 50 percent overlap of the Hanning temporal weighting, $q = 1/2$, the situation is markedly improved, there being a value, $L_2/L_1 = .4$, at which the excess variance is only 8 percent above ideal; this is reminiscent of the variance ratio for the case of no quadratic smoothing in reference 9, tables 5-7. When q is decreased to 3/8 or 1/4 (62.5 and 75 percent overlap, respectively), virtually the ideal variance reduction can be achieved by choosing $L_2 \approx L_1$.

In figure 19, the shift fraction q is kept at 3/8, while B_eT is taken at both 100 and 32. The smaller value of B_eT leads to a slightly larger loss in performance because of more significant edge effects. However, even so, the normalized quality ratio does reach a very desirable level only 4 percent above ideal when $L_2 \sim 2L_1$.
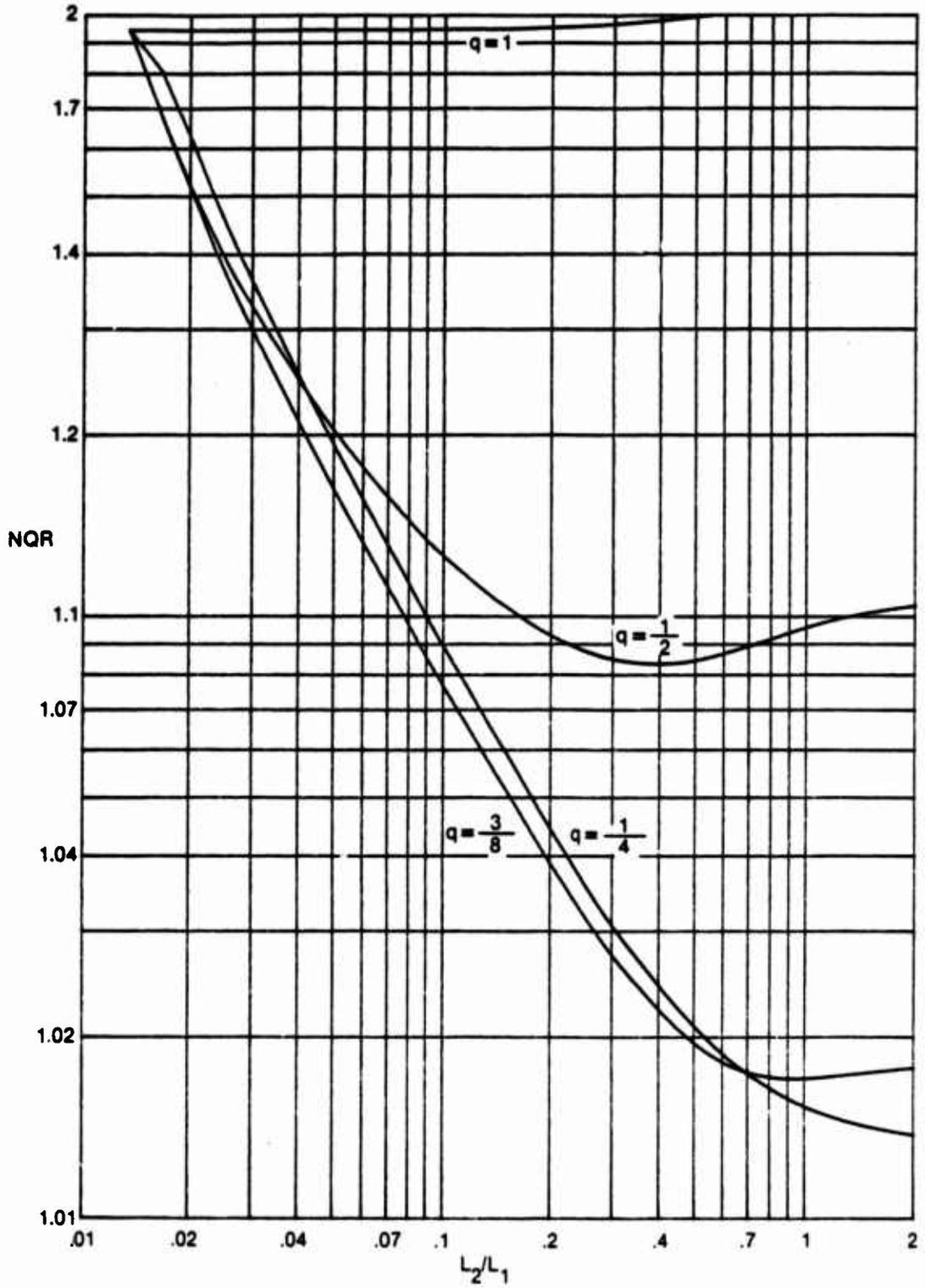
Figure 18. Normalized Quality-Ratio for Hanning Temporal Weighting
and Hanning Lag Weighting; $B_e T = 100$
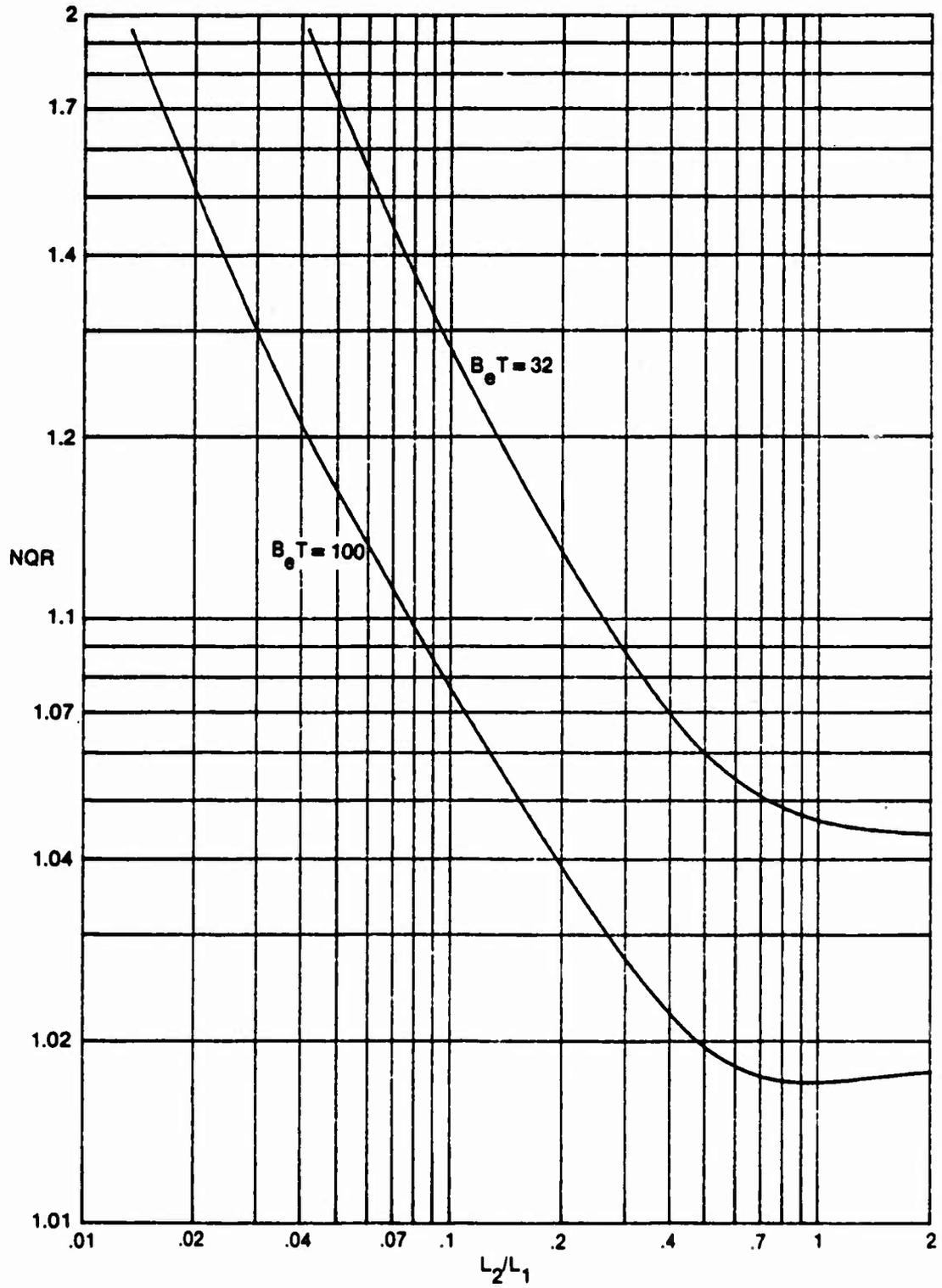
Figure 19. Normalized Quality-Ratio for Hanning Temporal Weighting
and Hanning Lag Weighting; q = 3/8

Summary

The possibilities and performance of a generalized spectral analysis technique employing temporal and lag weighting have been investigated in terms of the mean and variance of the spectral estimate. The only assumption required about the process under analysis, in so far as the mean is concerned, is that it be second-order stationary over the observation interval. We then were able to extract a simple expression for the effective window involving the temporal and lag windows.

The possibility of doing lag reshaping to achieve desirable effective windows was considered in detail and found to be reasonable for a wide variety of windows with good side lobe behavior and decay rates. In particular, if rectangular temporal weighting is employed, its inherent poor side lobe structure can be corrected via proper lag weighting, in so far as the effective window is concerned. Strictly speaking, the closest side lobe cannot be eliminated; however, all the other side lobes can be suppressed.

The effect of temporal and lag weighting on the variance of the spectral estimate was evaluated and compared with the ideal value for large $B_e T$. For rectangular temporal weighting, it was found that small values of L_2/L_1 and no overlap led to values of the normalized quality ratio virtually equal to the best attainable by any spectral analysis technique. The comparison is made under the constraint that the effective frequency resolution B_e is maintained the same for all techniques under consideration. On the other hand, if Hanning temporal weighting is employed, overlapping must be used for maximum variance reduction and the length ratio L_2/L_1 ought to be of the order of unity.

Since Hanning temporal weighting requires multiplication of each and every data segment (P pieces) and significant overlap ($\sim$ 50 percent), whereas rectangular temporal weighting requires no multiplication and no overlap, the latter approach is a strong candidate for spectral analysis, particularly since excellent effective windows (low side lobes and rapid decay) and virtually ideal variance reduction can be achieved by proper lag weighting and choice of ratio L_2/L_1. Investigation of other cases than those evaluated here can be achieved by appropriate modification of the program in appendix F. A major analytical task will be the evaluation of the third-order correlation (108), if temporal weighting $w_1(t)$ is taken other than rectangular or Hanning.

Appendix A

Comparison of Two Bandwidth Measures

The effective bandwidth of narrowband window $W_o(f)$ was defined in (1) as

$$B_e = \frac{\left[\int df \ W_o(f)\right]^2}{\int df \ W_o^2(f)} \quad .$$

(A-1)

The half-power bandwidth, B_h, is defined as the solution of

$$\frac{W_o\left(f_o \pm \frac{1}{2} B_h\right)}{W_o(f_o)} = \frac{1}{2} \quad ,$$

(A-2)

where it is assumed that window $W_o(f)$ is real, even about f_o, and peaked at f_o. We let $W_e(f) = W_o(f + f_o)$; thus $W_e(f)$ is even about $f = 0$.

The inverse Fourier transform of lowpass window $W_e(f)$ is called the weighting

$$w_e(\tau) = \int df \ \exp(i2\pi f\tau) \ W_e(f) \quad .$$

(A-3)

We consider here the class of weightings given by

$$w_e(\tau) = \sum_{k \geq 0} a_k \cos\left(\pi k\tau/L_e\right) \quad \text{for} \quad |\tau| < L_e \quad ,$$

(A-4)

and zero otherwise, where $\{a_k\}$ are real and non-negative. This class includes rectangular, Hanning, Hamming, Blackman, and the optimal windows of Nuttall, reference 19. The Fourier transform of (A-4) yields lowpass window

$$W_e(f) = \frac{2L_e}{\pi} \ 2L_e f \ \sin(2\pi L_e f) \sum_{k \geq 0} \frac{(-1)^k \ a_k}{(2L_e f)^2 - k^2} \quad ;$$

$$W_e\left(\frac{n}{2L_e}\right) = L_e \left\{ \begin{array}{ll} 2a_0 & \text{for} \quad n = 0 \\ \\ a_{|n|} & \text{for} \quad n \neq 0 \end{array} \right\} \quad .$$

(A-5)

A table of bandwidths B_e and B_h and their ratio is given below for the window in (A-5). Although these bandwidths vary significantly for the different weightings, their ratio is much more stable. In fact, for the last four weightings listed, the ratio is constant within ± 1 percent. The weightings listed under C5, C3, C1 are those given in reference 19, figures 10, 11, 12; the notation means

C5: continuous fifth derivative of weighting

C3: continuous third derivative of weighting

C1: continuous first derivative of weighting (A-6)

Table A-1. Bandwidths for Various Weightings

Weighting	B_eL_e	B_hL_e	B_e/B_h
Rectangular	0.5000	0.6034	0.8287
Hanning	1.3333	1.0000	1.3333
Hamming	1.2614	0.9109	1.3848
Blackman	1.6415	1.1494	1.4281
C5	2.2165	1.5371	1.4420
C3	2.0478	1.4139	1.4483
C1	1.9544	1.3444	1.4537

Appendix B

Some Lag Weighting and Lag Windowing Considerations For Discrete-Time Processing

It is convenient here to define an equispaced unit-impulse train by the notation

$$\delta_a(b) = \sum_n \delta(b - na) \quad , \tag{B-1}$$

where the summation on n extends over $\pm \infty$.

For discrete time sampling at spacing Δ_t, it has been observed in (30) that $\hat{G}_1(f)$ has period $1/\Delta_t$ in f. Therefore lag window $W_2(f)$ could be confined to $|f| < (2\Delta_t)^{-1}$ with no loss in generality, in so far as its effects on $\hat{G}_2(f)$ by means of (18) are concerned. In fact, for a general lag window $W_2(f)$ specified arbitrarily, the equivalent band limited lag window is

$$\tilde{W}_2(f) = \text{rect}(\Delta_t f) \left[W_2(f) \circledast \delta_{\frac{1}{\Delta_t}}(f) \right] \quad , \tag{B-2}$$

where we will utilize definitions

$$\text{rect}(x) = \begin{cases} 1 & \text{for} \quad |x| < 1/2 \\ 0 & \text{otherwise} \end{cases} \quad , \quad \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \quad , \tag{B-3}$$

and where $\circledast$ denotes convolution. That is, $W_2(f)$ is aliased into the band $|f| < (2\Delta_t)^{-1}$, and only this band-limited portion is retained for $\tilde{W}_2(f)$.

A way to demonstrate this mathematically is to note that the only values of lag weighting $w_2(\tau)$ that can affect $\hat{R}_2(\tau)$ are the samples

$$w_2(q\Delta_t) = \int df \, \exp(i2\pi f q \Delta_t) \, W_2(f) \quad . \tag{B-4}$$

The band-limited lag weighting function that passes through all these specified values, for all q, is

$$\tilde{w}_2(\tau) = \sum_q w_2(q\,\Delta_t) \, \text{sinc}\left(\frac{\tau}{\Delta_t} - q\right) \quad , \tag{B-5}$$

with corresponding Fourier transform

$$\bar{W}_2(f) = \Delta_t \ \text{rect}(\Delta_t f) \ \sum_q \ w_2(q\Delta_t) \ \exp(-i2\pi fq\Delta_t) \tag{B-6}$$

$$= \Delta_t \ \text{rect}(\Delta_t f) \ \sum_q \ \exp(-i2\pi fq\Delta_t) \int du \ \exp(i2\pi uq\Delta_t) \ W_2(u)$$

$$= \text{rect}(\Delta_t f) \ \int du \ W_2(u) \ \Delta_t \ \sum_q \ \exp\left[-i2\pi(f - u)q\Delta_t\right]$$

$$= \text{rect}(\Delta_t f) \ \int du \ W_2(u) \ \delta_{\frac{1}{\Delta_t}}(f - u)$$

$$= \text{rect}(\Delta_t f) \left[W_2(f) \ \bullet \ \delta_{\frac{1}{\Delta_t}}(f) \right] \ . \tag{B-7}$$

Relation (B-4) indicates how an arbitrarily specified $W_2(f)$ fixes the lag weights at the sample points. For the reverse problem, where sampled lag weights $\{w_2(q\Delta_t)\}$ are specified for all q, relation (B-6) gives an equivalent lag window, in particular the band-limited spectral window, which results in the same estimates $\hat{R}_2(\tau)$ and $\hat{G}_2(f)$. Notice that $W_2(f)$ is not uniquely specified by samples $\{w_2(q\Delta_t)\}$; however, the band-limited $W_2(f)$, which realizes weights $\{w_2(q\Delta_t)\}$ for all q, is unique and is given by (B-6).

As a special case of the above, consider discrete frequency smoothing with frequency spacing $\Delta_f = (M\Delta_t)^{-1}$, where $M\Delta_t$ is of the order of $2L_1$; i.e., from (38),

$$W_2(f) = \frac{1}{M\Delta_t} \ \sum_n \ W_{2n} \ \delta\left(f - \frac{n}{M\Delta_t}\right) \ , \tag{B-8}$$

where we set $W_{2n} = 0$ for $|n| > M/2$ without loss of generality, in accordance with the observation above (B-2). Then lag weights (B-4), given now by

$$w_2(q\Delta_t) = \frac{1}{M\Delta_t} \ \sum_{|n| \leq M/2} \ W_{2n} \ \exp(i2\pi nq/M) \ , \tag{B-9}$$

will equivalently accomplish the same purpose. This last relation can be accomplished by an M-point FFT, where $W_{2, \pm M/2}$ receive the same complex exponential weighting in (B-9).

It should be noted that the discrete function $w_2(q\Delta_t)$ in (B-9) has period M in q; this means that $w_2(q\Delta_t)$ in (B-9) will increase in magnitude for $M/2 < q < M$. If $\hat{R}_{1q}$ is nonzero for $|q| > M/2$, this lag weighting may cause a problem. One guaranteed way to avoid the problem is to choose M/2 larger than the nonzero extent, N_1, of $\hat{R}_{1q}$. Physically, this means that the frequency spacing $\Delta_f = (M\Delta_t)^{-1}$, used in frequency smoothing (B-8), must be small enough so as not to miss any information in $\hat{G}_1(f)$. Coarse frequency spacing gives spurious results for $\hat{G}_2(f)$. (It will also yield poor effective windows.)

Since from (24), $\hat{R}_{1q} = 0$ for $q \geq N_1$, where N_1 is the number of time samples per segment, only a finite number of the general weights $\{w_2(q\Delta_t)\}$ in (B-4) affect $\hat{R}_{2q}$. Thus in example (B-8), although (B-9) has period M in q, only the values for $|q| < N_1$ are relevant to the effect on $\hat{R}_{2q}$; more generally, the values yielded by (B-4) for a general $W_2(f)$ are relevant only for $|q| < N_1$, and only these need be evaluated and retained if we choose to process via the lag domain.

Now let us consider the reverse problem, where lag weights $\{w_2(q\Delta_t)\}$ are specified for all q, and we wish to determine some allowable lag windows $W_2(f)$ that will realize the same estimates $\hat{R}_2(\tau)$ and $\hat{G}_2(f)$, but which take advantage of the fact that only $w_2(q\Delta_t)$ for $|q| < N_1$ must be realized. One obvious candidate is the band-limited lag weighting version

$$\hat{w}_2(\tau) = \sum_{|q|<N_1} w_2(q\Delta_t) \, \text{sinc}\left(\frac{\tau}{\Delta_t} - q\right) \; ; \tag{B-10}$$

notice the limitation on q employed. The corresponding lag window is

$$\hat{W}_2(f) = \Delta_t \, \text{rect}(\Delta_t f) \sum_{|q|<N_1} w_2(q\Delta_t) \, e^{-i2\pi f q \Delta_t} \; . \tag{B-11}$$

A second candidate is

$$\hat{\hat{w}}_2(\tau) = \hat{w}_2(\tau) \; \circledast \; \delta_{M\Delta_t}(\tau) \quad , \tag{B-12}$$

provided that $M \geq 2N_1 - 1$; this provision guarantees non-overlap of the displacements of $\hat{w}_2(\tau)$. Then

$$\hat{\hat{W}}_2(f) = \hat{W}_2(f) \frac{1}{M\Delta_t} \; \delta_{\frac{1}{M\Delta_t}}(f) = \sum_n W_{2n} \, \delta\left(f - \frac{n}{M\Delta_t}\right) \tag{B-13}$$

where

$$W_{2n} \equiv \frac{1}{M\Delta_t} \hat{W}_2\left(\frac{n}{M\Delta_t}\right) = \frac{1}{M} \text{rect}\left(\frac{n}{M}\right) \sum_{|q|<N_1} w_2(q\Delta_t) \, \exp(-i2\pi nq/M) \tag{B-14}$$

from (B-11). Notice that (B-13) has the form of discrete frequency smoothing in (B-8); (B-14) gives the area of each impulse needed in (B-13). Also notice from (B-14) that all these areas are zero for $|n| > M/2$; thus we have a finite sequence to apply in the frequency domain, which is equivalent to a specified finite set of lag weights.

Equations (B-14) and (B-9) are complementary to each other. In fact, we can derive (B-9) from (B-14) as follows: from (B-14),

$$W_{2n} = \frac{1}{M} \text{rect}\left(\frac{n}{M}\right) \sum_q \overline{w}_q \, \exp(-i2\pi nq/M) \quad , \tag{B-15}$$

where

$$\overline{w}_q \equiv \begin{cases} w_2(q\Delta_t) & \text{for} \quad |q| < N_1 \\ \\ 0 & \text{otherwise} \end{cases}$$

(B-16)

Then

$$\sum_{|n| \leq M/2} W_{2n} \exp(i2\pi np/M) = \sum_{|n| \leq M/2} \exp(i2\pi np/M) \frac{1}{M} \text{rect}\left(\frac{n}{M}\right) \sum_q \overline{w}_q \exp(-i2\pi nq/M)$$

$$= \sum_q \overline{w}_q \frac{1}{M} \sum_{|n| \leq M/2} \text{rect}\left(\frac{n}{M}\right) \exp(i2\pi n(p - q)/M)$$

$$= \sum_q \overline{w}_q I_{p-q}^{(M)} = \overline{w}_p \bullet I_p^{(M)} \quad ,$$

(B-17)

where

$$I_p^{(M)} \equiv \begin{cases} 1 & \text{for} \quad p = 0, \pm M, \pm 2M, \ldots \\ \\ 0 & \text{otherwise} \end{cases}$$

(B-18)

Now if $M \geqslant 2N_1 - 1$ (as assumed above (B-13)), then

$$\sum_{|n| \leq M/2} W_{2n} \exp(i2\pi np/M) = w_2(q\Delta_t) \quad \text{for} \quad |q| < N_1 \quad .$$

(B-19)

This is (B-9) for $|q| < N_1$, which is the only range that affects $\hat{R}_{2q}$.

Appendix C

Correlation $\phi_1(\tau)$ of a General Class of Temporal Weightings

The class of temporal weightings of interest here is given by a sum of complex exponentials:

$$w_1(t) = \sum_k \alpha_k \exp(i2\pi kt/L_1) \quad \text{for} \quad |t| < L_1/2 \quad , \tag{C-1}$$

and zero otherwise. We assume that the coefficient sequence has conjugate symmetry

$$\alpha_{-k} = \alpha_k^* \quad \text{for all } k \quad ; \tag{C-2}$$

then $w_1(t)$ is real, and it follows that the (aperiodic) correlation

$$\phi_1(\tau) = \int dt \; w_1(t) \; w_1(t - \tau) \tag{C-3}$$

is also real, in addition to being even about $\tau = 0$.

Substitution of (C-1) in (C-3) yields

$$\phi_1(\tau) = \sum_{km} \alpha_k \, \alpha_m^* \int_{\tau - L_1/2}^{L_1/2} dt \, \exp\left[i2\pi kt/L_1 - i2\pi m(t - \tau)/L_1\right]$$
$$\text{for} \quad 0 \le \tau \le L_1 \quad . \tag{C-4}$$

This can be evaluated and then extended to $\tau < 0$ by the use of the even character of $\phi_1(\tau)$; there follows

$$\phi_1(\tau) = (L_1 - |\tau|) \left[\alpha_0^2 + 2 \sum_{k>0} |\alpha_k|^2 \cos(2\pi k\tau/L_1)\right]$$

$$- \frac{2L_1}{\pi} \sum_{m>k} \frac{(-1)^{m-k}}{m - k} \text{Re}\left(\alpha_k \, \alpha_m^*\right) \sin(2\pi m |\tau|/L_1)$$
$$\text{for} \quad |\tau| < L_1 \quad , \tag{C-5}$$

and zero otherwise. This is the general result for the correlation of weighting (C-1).

We now specialize (C-5) to the case of real symmetric coefficients in (C-1):

$$\alpha_k \quad \text{real}, \qquad \alpha_{-k} = \alpha_k \quad . \tag{C-6}$$

For all coefficients zero except for α_0, we have

$$w_1(t) = \alpha_0 \qquad \text{for} \quad |t| < L_1/2 \quad, \tag{C-7}$$

$$\phi_1(\tau) = (L_1 - |\tau|)\,\alpha_0^2 \quad \text{for} \quad |\tau| < L_1 \quad. \tag{C-8}$$

For all coefficients zero except for α_0, α_1, we have

$$w_1(t) = \alpha_0 + 2\alpha_1 \cos(2\pi t/L_1) \quad \text{for} \quad |t| < L_1/2 \quad, \tag{C-9}$$

$$\phi_1(\tau) = (L_1 - |\tau|)\left[\alpha_0^2 + 2\alpha_1^2 \cos(2\pi\tau/L_1)\right]$$

$$+ \frac{L_1}{\pi}\,\alpha_1(2\alpha_0 - \alpha_1)\,\sin(2\pi|\tau|/L_1) \quad \text{for} \quad |\tau| < L_1 \quad. \tag{C-10}$$

For all coefficients zero except for $\alpha_0, \alpha_1, \alpha_2$, we have

$$w_1(t) = \alpha_0 + 2\alpha_1 \cos(2\pi t/L_1) + 2\alpha_2 \cos(4\pi t/L_1) \quad \text{for} \quad |t| < L_1/2 \quad, \tag{C-11}$$

$$\phi_1(\tau) = (L_1 - |\tau|)\left[\alpha_0^2 + 2\alpha_1^2 \cos(2\pi\tau/L_1) + 2\alpha_2^2 \cos(4\pi\tau/L_1)\right]$$

$$+ \frac{L_1}{6\pi}\left[2\alpha_1(6\alpha_0 - 3\alpha_1 - 4\alpha_2)\,\sin(2\pi|\tau|/L_1)\right.$$

$$\left. - \alpha_2(6\alpha_0 - 16\alpha_1 + 3\alpha_2)\,\sin(4\pi|\tau|/L_1)\right]$$

$$\text{for} \quad |\tau| < L_1 \quad. \tag{C-12}$$

For all coefficients zero except for $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, we have

$$w_1(t) = \alpha_0 + 2\alpha_1 \cos(2\pi t/L_1) + 2\alpha_2 \cos(4\pi t/L_1) + 2\alpha_3 \cos(6\pi t/L_1)$$

$$\text{for} \quad |t| < L_1/2 \quad, \tag{C-13}$$

$$\phi_1(\tau) = (L_1 - |\tau|)$$

$$\cdot \left[\alpha_0^2 + 2\alpha_1^2 \cos(2\pi\tau/L_1) + 2\alpha_2^2 \cos(4\pi\tau/L_1) + 2\alpha_3^2 \cos(6\pi\tau/L_1)\right]$$

$$+ \frac{L_1}{30\pi} \left[5\alpha_1(12\alpha_0 - 6\alpha_1 - 8\alpha_2 + 3\alpha_3) \sin(2\pi|\tau|/L_1)\right.$$

$$- \alpha_2(30\alpha_0 - 80\alpha_1 + 15\alpha_2 + 48\alpha_3) \sin(4\pi|\tau|/L_1)$$

$$\left. + \alpha_3(20\alpha_0 - 45\alpha_1 + 72\alpha_2 - 10\alpha_3) \sin(6\pi|\tau|/L_1)\right]$$

$$\text{for} \quad |\tau| < L_1 \quad . \qquad \text{(C-14)}$$

This last case includes all the weightings considered in reference 19, with the identification of coefficients as

$$\alpha_0 = a_0 \quad , \qquad \alpha_k = \frac{1}{2}a_k \quad \text{for} \quad k > 0 \quad . \qquad \text{(C-15)}$$

Then we can express the temporal weighting as

$$w_1(t) = \sum_{k \geq 0} a_k \cos(2\pi k t/L_1) \quad \text{for} \quad |t| < L_1/2 \quad . \qquad \text{(C-16)}$$

Appendix D

Derivation of Variance of Spectral Estimate

Our starting point is (14). The integral on t is a Fourier transform of the product in (13), and can therefore be expressed as a convolution:

$$Y_p(f) = X(f) \circledast \left[W_1(f) \exp\left(-i2\pi f\left(\frac{L_1}{2} + pS\right)\right) \right]$$

$$= \int du\, X(f - u)\, W_1(u)\, \exp\left(-i2\pi u\left(\frac{L_1}{2} + pS\right)\right) , \qquad \text{(D-1)}$$

where we used (46) and defined

$$X(f) = \int dt\, x(t)\, \exp(-i2\pi ft). \qquad \text{(D-2)}$$

Although the relations to follow could be derived in the time domain, it is more convenient to develop them in the frequency domain because of the frequent and useful occurrence of delta functions.

$X(f)$ is complex Gaussian for all f, since (D-2) is a linear transformation and we have assumed $x(t)$ to be a complex Gaussian process, for the variance calculation to follow. Furthermore covariance

$$Av\{X(f_1)\, X^*(f_2)\} = \iint dt_1\, dt_2\, \overline{x(t_1)\, x^*(t_2)}\, \exp(-i2\pi f_1 t_1 + i2\pi f_2 t_2)$$

$$= \iint dt_1\, dt_2\, R(t_1 - t_2)\, \exp(-i2\pi f_1 t_1 + i2\pi f_2 t_2)$$

$$= \iint du\, dt_2\, R(u)\, \exp(-i2\pi f_1(u + t_2) + i2\pi f_2 t_2)$$

$$= G(f_1)\, \delta(f_1 - f_2) , \qquad \text{(D-3)}$$

upon use of (43) and (45). When $x(t)$ is a single-sided (analytic) complex process, there then follows for the fourth-order average, which will be needed later (reference 23),

$$Av\{X(f_1)\, X^*(f_2)\, X(f_3)\, X^*(f_4)\}$$

$$= G(f_1)\, G(f_3)\left[\delta(f_1 - f_2)\, \delta(f_3 - f_4) + \delta(f_1 - f_4)\, \delta(f_2 - f_3)\right] . \qquad \text{(D-4)}$$

(When x(t) is a real process, there would be a third term in addition. However, even then, this term contributes only near $f = 0$; see reference 9, equations (A-4) - (A-8).)

Upon use of (D-1), (14) becomes*

$$\hat{G}_1(f) = \frac{1}{P} \sum_{p=0}^{P-1} \iint du\ dv\ X(f - u)\ X^*(f - v)$$

$$\cdot\ W_1(u)\ W_1^*(v)\ \exp\left(-i2\pi(u - v)\left(\frac{L_1}{2} + pS\right)\right)\ . \tag{D-5}$$

Then the average of the product of the first-stage spectral estimates is

$$Av\{\hat{G}_1(f_1)\ \hat{G}_1(f_2)\} = \frac{1}{P^2} \sum_{p,q=0}^{P-1} \iiiint du\ dv\ d\mu\ d\nu$$

$$\cdot\ \overline{X(f_1 - u)\ X^*(f_1 - v)\ X(f_2 - \mu)\ X^*(f_2 - \nu)\ W_1(u)\ W_1^*(v)\ W_1(\mu)\ W_1^*(\nu)}$$

$$\cdot\ \exp\left(-i2\pi(u - v)\left(\frac{L_1}{2} + pS\right)\right) \exp\left(-i2\pi(\mu - \nu)\left(\frac{L_1}{2} + qS\right)\right)\ . \tag{D-6}$$

Reference to (D-4) enables us to express the fourth-order average as

$$G(f_1 - u)\ G(f_2 - \mu)$$

$$\cdot\ \left[\delta(u - v)\ \delta(\mu - \nu) + \delta(f_1 - f_2 + \nu - \mu)\ \delta(f_1 - f_2 + \mu - v)\right]\ . \tag{D-7}$$

Use of the first term of (D-7) in (D-6) yields

$$\frac{1}{P^2} \sum_{p,q=0}^{P-1} \iint du\ d\mu\ G(f_1 - u)\ G(f_2 - \mu)\ |W_1(u)|^2\ |W_1(\mu)|^2$$

$$= \int du\ G(f_1 - u)\ |W_1(u)|^2 \cdot \int d\mu\ G(f_2 - \mu)\ |W_1(\mu)|^2$$

$$= Av\{\hat{G}_1(f_1)\} \cdot Av\{\hat{G}_1(f_2)\}\ , \tag{D-8}$$

where we employed (45) in the last line. Moving this term to the left side of (D-6), and using the second term of (D-7), we obtain, for the covariance of the first-stage spectral estimates,

---

*For more generality, we allow temporal weight w_1 and window W_1 to be complex in this appendix.

$$\text{Cov}\{\hat{G}_1(f_1), \hat{G}_1(f_2)\} = \frac{1}{P^2} \sum_{p,q=0}^{P-1} \iint du \; dv \; G(f_1 - u) \; G(f_1 - v)$$

$$\cdot W_1(u) \; W_1^*(v) \; W_1(v + f_2 - f_1) \; W_1^*(u + f_2 - f_1) \; \exp(-i2\pi(u - v)(p - q)S)$$

$$= \frac{1}{P^2} \sum_{p,q=0}^{P-1} \iint d\alpha \; d\beta \; G(\alpha) \; G(\beta) \; W_1(f_1 - \alpha) \; W_1^*(f_1 - \beta)$$

$$\cdot W_1(f_2 - \beta) \; W_1^*(f_2 - \alpha) \; \exp(-i2\pi(\beta - \alpha)(p - q)S)$$

$$= \frac{1}{P^2} \sum_{p,q=0}^{P-1} \left| \int d\alpha \; G(\alpha) \; W_1(f_1 - \alpha) \; W_1^*(f_2 - \alpha) \; \exp(i2\pi\alpha(p - q)S) \right|^2$$

$$= \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \left| \int d\alpha \; G(\alpha) \; W_1(f_1 - \alpha) \; W_1^*(f_2 - \alpha) \; \exp(i2\pi\alpha pS) \right|^2$$

$$= \iint d\alpha \; d\beta \; G(\alpha) \; G(\beta) \; W_1(f_1 - \alpha) \; W_1^*(f_1 - \beta) \; W_1(f_2 - \beta) \; W_1^*(f_2 - \alpha)$$

$$\cdot \left[\frac{\sin(\pi PS(\alpha - \beta))}{P \sin(\pi S(\alpha - \beta))} \right]^2 \; . \tag{D-9}$$

Here we have used the identities

$$\frac{1}{P^2} \sum_{p,q=0}^{P-1} \exp(i2\pi(p - q)u) = \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \exp(i2\pi pu)$$

$$= \left[\frac{\sin(\pi Pu)}{P \sin(\pi u)} \right]^2 \equiv Q_p(u) \; . \tag{D-10}$$

For $f_1 = f_2$, (D-9) checks with reference 9, equation (A-9); more generally, it is equation D-2 of reference 9. We observe that if $|f_2 - f_1|$ is greater than the effective bandwidth of temporal window W_1, (D-9) will be small since $W_1(f_1 - \alpha)$ and $W_1^*(f_2 - \alpha)$ will then not overlap significantly on the α-scale. Also notice that spectrum G is still left under the integral sign; i.e., there are no assumptions yet on the character of the spectrum.

We are now prepared to consider the second-stage spectral estimate $\hat{G}_2(f)$ as given by (18):

$$\hat{G}_2(f) = \int df_1 \; \hat{G}_1(f_1) \; W_2(f - f_1) \; . \tag{D-11}$$

Then

$$\hat{G}_2(f) - Av\left\{\hat{G}_2(f)\right\} = \int df_1 \left[\hat{G}_1(f_1) - Av\left\{\hat{G}_1(f_1)\right\}\right] W_2(f - f_1) \quad , \tag{D-12}$$

and therefore (recall that W_2 is real) the variance of the second-stage spectral estimate is

$$Var\left\{\hat{G}_2(f)\right\} = \iint df_1 \, df_2 \, Cov\left\{\hat{G}_1(f_1), \, \hat{G}_1(f_2)\right\} W_2(f - f_1) \, W_2(f - f_2)$$

$$= \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \iint d\alpha \, d\beta \, G(\alpha) \, G(\beta) \, \exp(i2\pi(\alpha - \beta)pS) \, |\gamma(f - \alpha, \, f - \beta)|^2$$

$$= \iint d\alpha \, d\beta \, G(\alpha) \, G(\beta) \cdot |\gamma(f - \alpha, \, f - \beta)|^2 \, Q_P(S(\alpha - \beta)) \, , \tag{D-13}$$

where we used (D-9) and (D-10), interchanged integrals, and defined window convolution function

$$\gamma(x,y) \equiv \int du \, W_2(u) \, W_1(x - u) \, W_1^*(y - u) \quad . \tag{D-14}$$

Relation (D-13) is exact; it makes no presumption about the relative widths of the spectrum G and the windows W_1, W_2. The compact expression (D-13) involves the windows W_1, W_2 through the convolution function γ, and involves the shift S and number of pieces P through the periodic function Q_P defined in (D-10).

The window convolution function γ in (D-14) realizes its peak value at $x = 0$, $y = 0$, and is rather small everywhere else, since the windows are virtually unimodal and rather narrow. In fact, a special case is the diagonal slice

$$\gamma(x,x) = \int du \, W_2(u) \, |W_1(x - u)|^2 = |W_1(x)|^2 \circledast W_2(x) = W_e(x) \quad , \tag{D-15}$$

by reference to (51). Generally, $\gamma(x, y)$ is substantially nonzero only in the region B_e, B_e at the origin of x, y space.

We now employ the assumption discussed in connection with approximation (103), namely, that spectrum G is relatively constant in the band of width B_e about the frequency f of interest. Then the major contribution to the variance, (D-13), comes from the region near $\alpha = \beta = f$ in α, β space. There follows the approximation for the variance of the second-stage spectral estimate,

$$\text{Var}\{\hat{G}_2(f)\} \simeq G^2(f) \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right)$$

$$\cdot \iint d\alpha \ d\beta \ \exp(i2\pi(\alpha - \beta)pS)|\gamma(f - \alpha, \ f - \beta)|^2$$

$$= G^2(f) \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \iint dx \ dy \ \exp(i2\pi(y - x)pS)|\gamma(x, \ y)|^2$$

$$= G^2(f) \iint dx \ dy \ |\gamma(x, \ y)|^2 \ Q_p(S(y - x)) \ , \qquad \text{(D-16)}$$

where we used (D-10) again.

We now simplify the double integral in (D-16); from the second line of (D-16) and from (D-14),

$$D \equiv \iint dx \ dy \ \exp(i2\pi(y - x)pS) \left| \int du \ W_2(u) \ W_1(x - u) \ W_1^*(y - u) \right|^2$$

$$= \iint du \ dv \ W_2(u) \ W_2(v) \left| \int dx \ \exp(-i2\pi xpS) \ W_1(x - u) \ W_1^*(x - v) \right|^2$$

$$= \iint du \ dv \ W_2(u) \ W_2(v) \ |\chi_1(pS, \ u - v)|^2 \ , \qquad \text{(D-17)}$$

where the complex ambiguity function of the temporal weight and window is defined as

$$\chi_1(\tau, \ \nu) = \int df \ \exp(i2\pi f\tau) \ W_1\left(f + \frac{\nu}{2}\right) \ W_1^*\left(f - \frac{\nu}{2}\right)$$

$$= \int dt \ \exp(-i2\pi\nu t) \ w_1\left(t + \frac{\tau}{2}\right) \ w_1^*\left(t - \frac{\tau}{2}\right) \ . \qquad \text{(D-18)}$$

Now let $\nu = u - v$ in (D-17), and obtain a single integral for D:

$$D = \int d\nu |\chi_1(pS, \ \nu)|^2 \ \int du \ W_2(u) \ W_2(u - \nu)$$

$$= \int d\nu |\chi_1(pS, \ \nu)|^2 \ \chi_2(0, \ \nu) \ , \qquad \text{(D-19)}$$

in terms of the ambiguity function χ_2 of the lag window and weight. Substitution of (D-19) in the second line of (D-16) yields for the variance,

$$\text{Var}\{\hat{G}_2(f)\} = G^2(f) \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \int d\nu |\chi_1(pS, \ \nu)|^2 \ \chi_2(0, \ \nu) \ . \qquad \text{(D-20)}$$

An alternative, and perhaps more useful form, to (D-20) is attained as follows; the integral in (D-20) is expressible (by use of the definition (D-18) of the ambiguity function) as

$$\int d\nu |\chi_1(pS, \nu)|^2 \int dt \exp(-i2\pi\nu t) w_2^2(t)$$

$$= \int dt \, w_2^2(t) \int d\nu \exp(-i2\pi\nu t) |\chi_1(pS, \nu)|^2 \quad . \tag{D-21}$$

Now we have the general result that

$$\int d\nu \exp(-i2\pi\nu t) |\chi_1(\tau, \nu)|^2$$

$$= \int d\nu \exp(-i2\pi\nu t) \iint dt_1 \, dt_2 \exp(-i2\pi\nu(t_1 - t_2))$$

$$\cdot w_1\left(t_1 + \frac{\tau}{2}\right) w_1^*\left(t_1 - \frac{\tau}{2}\right) w_1^*\left(t_2 + \frac{\tau}{2}\right) w_1\left(t_2 - \frac{\tau}{2}\right)$$

$$= \iint dt_1 \, dt_2 \, w_1\left(t_1 + \frac{\tau}{2}\right) w_1^*\left(t_1 - \frac{\tau}{2}\right) w_1^*\left(t_2 + \frac{\tau}{2}\right) w_1\left(t_2 - \frac{\tau}{2}\right) \delta(t + t_1 - t_2)$$

$$= \int dt_1 \, w_1\left(t_1 + \frac{\tau}{2}\right) w_1^*\left(t_1 - \frac{\tau}{2}\right) w_1^*\left(t + t_1 + \frac{\tau}{2}\right) w_1\left(t + t_1 - \frac{\tau}{2}\right)$$

$$= \int du \, w_1\left(u + \frac{\tau - t}{2}\right) w_1^*\left(u + \frac{-\tau - t}{2}\right) w_1^*\left(u + \frac{\tau + t}{2}\right) w_1\left(u + \frac{-\tau + t}{2}\right)$$

$$\equiv \phi_3(t, \tau) \quad . \tag{D-22}$$

This is a third-order correlation of temporal weighting w_1; see the footnote to (9). Thus (D-21) becomes

$$\int dt \, w_2^2(t) \, \phi_3(t, pS) \quad , \tag{D-23}$$

and the variance in (D-20) becomes

$$\text{Var}\{\hat{G}_2(f)\} = G^2(f) \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P}\right) \int dt \, w_2^2(t) \, \phi_3(t, pS) \tag{D-24}$$

This "weighting-domain" version of the variance is very useful if third-order correlation ϕ_3 can be evaluated in closed form.

Appendix E

Computational Considerations for Non-Overlapping Segments

The curves in figure 15 for non-overlapping time segments are drawn over a continuum of values of L_2/L_1. However, if we were given a value of TB_e, all values of L_2/L_1 may not be allowed. To develop this point, suppose that we pick an integer value for the number of pieces, P, and solve for $L_1B_e = TB_e/P$ according to (129). From figures 5 and 6, this dictates the value for L_2B_e, and hence a discrete value for L_2/L_1 is specified for each value of P. The number of pieces, P, can range from 1 to P_{max}, where

$$B_e L_1 (min) = \frac{B_e T}{P_{max}} = \frac{1}{2c\{\phi_1\}} \quad , \quad P_{max} = 2c\{\phi_1\} B_e T \quad , \qquad (E-1)$$

from figure 5. For $B_e T \gg 1$, we have, for

$$P = 1, \quad B_e L_1 = B_e T, \quad B_e L_2 = \frac{1}{2c\{w_2\}} , \quad \frac{L_2}{L_1} = \frac{1}{2c\{w_2\}B_e T} \ll 1 \quad , \qquad (E-2)$$

and for

$$P = P_{max}, \quad B_e L_1 = \frac{1}{2c\{\phi_1\}} , \quad B_e L_2 \geq B_e T, \quad \frac{L_2}{L_1} \geq 2c\{\phi_1\}B_e T \gg 1 \quad . \qquad (E-3)$$

Thus a very wide range of discrete values of L_2/L_1 is allowed when $B_e T \gg 1$.

The problem with this approach is that when $B_e L_1$ is calculated, $B_e L_2$ must be solved for from the integral relation (52) (or approximately from figure 6). This tedious procedure can be circumvented by specifying L_2/L_1 instead; if desired, we could then use (52) to determine $B_e L_1$, and solve for $P = B_e T/(B_e L_1)$. However, P will not necessarily turn out to be an integer for a given fixed T; thus only a set of discrete values of L_2/L_1 are strictly legal. But if NQR_p in (130) does not vary radically with L_2/L_1, this is not a significant limitation. And since it is simpler, we adopt it. In Appendix F, we cannot avoid the calculation of $B_e L_2$ from a given $B_e L_1$.

Appendix F

Computational Considerations for Overlapping With Hanning Temporal Weighting

The temporal weighting is given by (146). Substitution in (108) and evaluation of the integral yields

$$\phi_3(\tau, \mu) = \frac{4}{9\pi L_1} \left\{ V_0(\pi - \alpha - \beta) + \sum_{k=1}^{4} (-1)^{k-1} V_k \sin(k(\alpha + \beta)) \right\}$$
$$\text{for} \quad \alpha + \beta \leq \pi \quad , \qquad \text{(F-1)}$$

where

$$\alpha = \frac{\pi}{L_1}|\tau| \quad , \qquad \beta = \frac{\pi}{L_1}|\mu| \quad ,$$

$$C_\alpha = \cos(\alpha) = \cos\left(\frac{\pi\tau}{L_1}\right) \quad , \quad C_\beta = \cos(\beta) = \cos\left(\frac{\pi\mu}{L_1}\right) \quad ,$$

$$V_0 = \frac{3}{8} - C_\alpha^2 - C_\beta^2 + 4C_\alpha^2 C_\beta^2 + C_\alpha^4 + C_\beta^4 \quad ,$$

$$V_1 = C_\alpha C_\beta \left(4C_\alpha^2 + 4C_\beta^2 - 1 \right) \quad ,$$

$$V_2 = -\frac{1}{4} + \frac{1}{2}\left(C_\alpha^2 + C_\beta^2 \right) + C_\alpha^2 C_\beta^2 \quad ,$$

$$V_3 = \frac{1}{3} C_\alpha C_\beta, \qquad V_4 = \frac{1}{32} \quad . \qquad \text{(F-2)}$$

The procedure for the evaluation of the normalized quality ratio follows. We specify a value for $B_e T$ and select a temporal weighting $w_1(t)$ and a lag weighting $w_2(\tau)$. We then evaluate shape factor $c\{\phi_1\}$ from (64) or table 2, and select a shift fraction $q = q\{w_1\}$ according to (139). We then solve (141) for P_{max}, and allow P to take integer values in the range $1 \leq P \leq P_{max}$. $B_e L_1$ can then be evaluated from (140) as

$$B_e L_1 = \frac{B_e T}{1 + (P - 1)q} \qquad \text{(F-3)}$$

for each integer value of P. We then solve (52) in the form

$$\frac{1}{B_e L_1} = \frac{1}{L_1} \int d\tau \, w_2^2(\tau) \, \phi_1^2(\tau) = \int dx \, w_2^2(L_1 x) \, \phi_1^2(L_1 x) \quad , \qquad \text{(F-4)}$$

for the ratio L_2/L_1. $w_2(L_1 x)$ is a function of L_2/L_1, while $\phi_1(L_1 x)$ is independent of L_1 and L_2. Next we compute quality ratio (109) in the form

$$Q = \frac{1}{P} \sum_{p=1-P}^{P-1} \left(1 - \frac{|p|}{P} \right) \int dx \ w_2^2(L_1 x) \ L_1 \phi_3(L_1 x, \ pq \ L_1) \qquad \text{(F-5)}$$

where we let $\tau = L_1 x$ and used (139). The quantity $w_2(L_1 x)$ is a function of only L_2/L_1, while the remaining quantity in the integrand of (F-5) is independent of L_1 and L_2. Finally, we multiply (F-5) by $B_e T$ according to (110) in order to determine the normalized quality ratio.

To reduce computation time, we take advantage of various properties of the functions involved. First, since $\phi_3(\tau, \mu)$ is even in τ and μ (see (108)), and w_2 is even, we express (F-5) as

$$Q = \frac{4}{P} \sum_{k=0}^{P-1} \epsilon_k \left(1 - \frac{k}{P} \right) \int_0^\infty dx \ w_2^2(L_1 x) \ L_1 \phi_3(L_1 x, \ qk \ L_1) \quad , \qquad \text{(F-6)}$$

where

$$\epsilon_k \equiv \begin{cases} 1/2 & \text{for} \quad k = 0 \\ 1 & \text{for} \quad k \geq 1 \end{cases} \quad . \qquad \text{(F-7)}$$

Also, from (F-1) and (F-2), we have normalized form

$$L_1 \phi_3(L_1 \ a, \ L_1 b) = \frac{4}{9\pi} \left\{ V_0 \ \pi(1 - a - b) + \sum_{k=1}^{4} (-1)^{k-1} V_k \ \sin\left[k\pi(a + b) \right] \right\}$$

$$\text{for} \ 0 \leq a, \quad 0 \leq b, \quad a + b \leq 1 \quad , \qquad \text{(F-8)}$$

where now

$$C_\alpha = \cos(\pi a), \quad C_\beta = \cos(\pi b) \quad , \qquad \text{(F-9)}$$

and $\{V_k\}_0^4$ are still as given in (F-2).

Since $L_1 \phi_3(L_1 x, \ pq L_1)$ in (F-6) is zero if $x + qk > 1$, we can limit the sum on k in (F-6) to $k_m = \min (P - 1, 1/q)$, and we need to evaluate the integral on x in (F-6) only up to $X_m = \min (L_2/L_1, 1 - qk)$. The number of x intervals needed in (F-6) is about 16 with the Trapezoidal rule for integration. These features are incorporated in the program listing below.

```
1 ! NUSC TECHNICAL REPORT 6459, 29 MAY 1981, A. H. NUTTALL
10     Bet=100                          ! Be T
20     C1=1/6+35/(48*PI^2)              ! HANNING TEMPORAL-WEIGHTING W1; TABLE 2
30     Q1=3/8                           ! 62.5% OVERLAP FOR W1; 0<Q1<=1
40     Pm=1+(2*C1*Bet-1)/Q1             ! Pmax; EQ. 141
50     PRINT "Be T =";Bet,"q = q(w1) =";Q1
60     PRINT
70     COM T1,T2,T3,T4
80     T2=2*PI
90     T3=1/3
100    T4=1/T2
110    FOR P=1 TO Pm
120    Bell=Bet/(1+(P-1)*Q1)            ! Be L1; EQ. 140
130    L211=FNL211(Bell,C1)             ! L2/L1; SOLUTION OF EQ. 52
140    T5=.5*PI/L211
150    Km=MIN(P-1,1/Q1)
160    S=.5*FNInt(0,L211,T5)
170    FOR K=1 TO Km
180    S=S+(1-K/P)*FNInt(Q1*K,L211,T5)
190    NEXT K
200    Q=4/P*S
210    PRINT P,L211,Q*Bet
220    NEXT P
230    END
240    !
250    DEF FNL211(Bell,C1)              ! SOLVE EQ. 52 FOR L2/L1
260    Eps=Bell-.5/C1
270    X1=4/(3*Bell)
280    IF Bell<10 THEN X1=.603246/SQR(Eps)-Eps/(6+12*Eps)
290    X2=X1*1.037
300    F1=FNF(X1,Bell)
310    F2=FNF(X2,Bell)
320    IF ABS(F2-F1)<1E-6 THEN 380
330    T=X2
340    X2=(F2*X1-F1*X2)/(F2-F1)
350    X1=T
360    F1=F2
370    GOTO 310
380    RETURN X2
390    FNEND
400    !
```

```
410    DEF FNF(L211,Bel1)     ! RIGHT SIDE - LEFT SIDE OF EQ. F-4
420    COM T1
430    Xm=MIN(1,L211)
440    T1=.5*PI/L211
450    N=16
460    Del=Xm/N
470    F=.5
480    FOR K=1 TO N-1
490    F=F+FNG(K*Del)
500    NEXT K
510    RETURN 2*Del*F-1/Bel1
520    FNEND
530    !
540    DEF FNG(X)                        ! (W2(L1*X)*Phi1(L1*X))^2
550    COM T1,T2,T3,T4
560    P=T2*X
570    G=(1-X)*(2+COS(P))*T3+T4*SIN(P)
580    W2=COS(T1*X)^2                    ! HANNING LAG-WEIGHTING W2
590    RETURN (W2*G)^2
600    FNEND
610    !
620    DEF FNInt(Qk,L211,T5)     ! INTEGRAL OF EQ. F-5
630    Xm=MIN(L211,1-Qk)
640    Del=Xm/16
650    S=.5*FNPhi3(0,Qk)
660    FOR J=1 TO 15
670    X=Del*J
680    S=S+COS(T5*X)^4*FNPhi3(X,Qk)
690    NEXT J
700    RETURN Del*S
710    FNEND
720    !
730    DEF FNPhi3(A,B)  ! L1*PHI3(L1*A,L1*B) for A>=0, B>=0; EQ. F-8
740    IF A+B>=1 THEN RETURN 0
750    Ab=PI*(A+B)
760    Ca=COS(PI*A)
770    Cb=COS(PI*B)
780    Ca2=Ca^2
790    Cb2=Cb^2
800    S=Ca2+Cb2
810    V0=.375-S+4*Ca2*Cb2+Ca2^2+Cb2^2
820    V1=Ca*Cb*(4*S-1)
830    V2=-.25+.5*S+Ca2*Cb2
840    V3=Ca*Cb/3
850    S=V1*SIN(Ab)-V2*SIN(2*Ab)+V3*SIN(3*Ab)-.03125*SIN(4*Ab)
860    L1phi3=4/(9*PI)*(V0*(PI-Ab)+S)
870    RETURN L1phi3
880    FNEND
```

References

1. M. S. Bartlett, *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications*, Cambridge University Press, Cambridge, 1953.

2. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover Publications, Inc., New York, 1959.

3. E. Parzen, "Mathematical Considerations in the Estimation of Spectra," *Technometrics*, vol. 3, 1961.

4. R. B. Blackman, *Data Smoothing and Prediction*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1965.

5. J. S. Bendat and A. G. Piersol, *Measurement and Analysis of Random Data*, J. Wiley & Sons, Inc., New York, 1966.

6. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day Company, San Francisco, 1968.

7. P. D. Welch, "The Use of FFT for the Estimation of Power Spectra: a Method Based on Time Averaging over Short Modified Periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol AU-15, no. 2, June 1967, pp. 70-73.

8. C. Bingham, M. D. Godfrey, and J. W. Tukey, "Modern Techniques of Power Spectrum Estimation," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-15, no. 2, June 1967.

9. A. H. Nuttall, "Spectral Estimation by Means of Overlapped Fast Fourier Transform Processing of Windowed Data," NUSC Report No. 4169, 13 October 1971.

10. A. H. Nuttall, "Estimation of Cross-Spectra via Overlapped Fast Fourier Transform Processing," NUSC Technical Report 4169-S, 11 July 1975.

11. A. H. Nuttall, "Minimum-Bias Windows for Spectral Estimation by Means of Overlapped Fast Fourier Transform Processing," NUSC Technical Report 4513, 11 April 1973.

12. A. H. Nuttall, "Probability Distribution of Spectral Estimates Obtained via Overlapped FFT Processing of Windowed Data," NUSC Technical Report 5529, 3 December 1976.

13. A. H. Nuttall and G. C. Carter, "A Generalized Framework for Power Spectral Estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 3, pp. 334-335, June 1980.

14. G. C. Carter and A. H. Nuttall, "A Brief Summary of a Generalized Framework for Spectral Estimation," *Signal Processing*, vol. 2, no. 4, pp. 387-390, October 1980.

15. G. C. Carter and A. H. Nuttall, "On the Weighted Overlapped Segment-Averaging Method for Power Spectral Estimation," *Proceedings IEEE*, vol. 68, no. 10, pp. 1352-1354, October 1980.

List of References (Cont'd)

16. J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, J. Wiley & Sons Inc., N.Y., 1971. See section 9.6, especially subsection 9.6.2.

17. R. K. Otnes and L. Enochson, *Applied Time Series Analysis*, J. Wiley and Sons Inc., N.Y., 1978. See section 8.8, esp. pp. 354-355.

18. A. H. Nuttall, "Reconstruction of Power Spectral Estimates," NUSC Technical Memorandum 781051, 6 March 1978.

19. A. H. Nuttall, "Some Windows with Very Good Sidelobe Behavior," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol, ASSP-29, no. 1, pp. 84-91, February 1981.

20. C. Bingham, M.D. Godfrey, and J. W. Tukey, "Modern Techniques of Power Spectrum Estimation," *IEEE Trans. on Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 56-66, June 1967.

21. E. A. Sloane, "Comparison of Linearly and Quadratically Modified Spectral Estimates of Gaussian Signals," *IEEE Trans. on Audio and Electroacoustics*, vol. AU-17, no. 2, pp. 133-137, June 1969.

22. G. H. Robertson, "Influence of Data Window Shape on Detectability of Small CW Signals in White Noise," *J. Acoust. Soc. Am.*, vol. 67, no. 4, pp. 1274-1276, April 1980.

23. A. H. Nuttall, "High-Order Covariance Functions for Complex Gaussian Processes," *IEEE Trans. on Information Theory*, vol. IT-8, no. 3, pp. 255-256, April 1962.

Two-Channel Linear-Predictive Spectral Analysis; Program For the HP 9845 Desk Calculator

A. H. Nuttall

ABSTRACT

A program for linear-p,edictive spectral analysis of two channels of data, including estimation of the auto-spectra, cross-spectrum, magnitude-squared-coherence, and the argument of the coherence, is presented in BASIC for the HP 9845 desk calculator. Timing results for the major subroutines are included, and their dependence on the fundamental parameters of the data, filter, and desired spectral resolution is indicated. These techniques and program for spectral analysis are very appropriate for short data segments. In particular, a positive-definite spectral matrix estimate is guaranteed.

Applications to examples including a strong tonal interference in one channel are made, and a possible shortcoming of the technique is pointed out. A suggested remedy is proposed and a philosophy for multichannel spectral analysis is suggested for further consideration.

Table of Contents

List of Illustrations

List of Illustrations (Cont'd)

List of Tables

List of Symbols

N	Number of data points in each process
Pmax	Maximum order of predictive filter to be considered
Nfft	Size of FFT to be used in spectral analysis
AIC	Akaike Information Criterion
Pbest	Best order of predictive filter to use, based on AIC
Δ	Time sampling increment
f	Frequency
f_N	Nyquist frequency, $(2\Delta)^{-1}$
$x_1(k)$	k-th sample of process 1
$x_2(k)$	k-th sample of process 2
$w_n(k)$	Additive white noise $(n = 1, 2)$
P	Predictive filter order
$A_P^{(P)}$	Partial forward correlation coefficient of order P
R_o	Covariance matrix at zero delay

Two-Channel Linear-Predictive Spectral Analysis; Program for the HP 9845 Desk Calculator

Introduction

Spectral analysis of short data segments by the standard FFT procedure is not a viable approach; unstable and/or coarse estimates of the spectra result. An attractive technique in this case is linear-predictive spectral analysis, both for the single-channel as well as the multiple-channel cases. See references 1-9, particularly references 7-9 which derive and give Fortran programs for a multiple-channel linear-predictive spectral analysis technique that is a generalization of Burg's technique for the single-channel case (reference 1).

The purpose of this report is twofold: first, we translate the Fortran program in reference 9 into Basic for use on the Hewlett-Packard HP 9845 Desk Calculator, and in the process, also make some minor improvements and modifications to the format and printout statements. We also limit consideration to the two-channel case and thereby take advantage of some simplifications in computing possible for this special case. Second, we apply the program to a pair of stationary processes, one of which has pure tones that are not present in the other process. In this manner, we point out a possibly deleterious effect on the auto-spectral estimates and the coherence estimate, and indicate a method for circumventing some of the difficulty. As a byproduct, a philosophy for multichannel spectral analysis is suggested for further consideration.

Use of Program For Spectral Analysis

In appendix A, the listing for the two-channel linear-predictive spectral analysis technique is presented. Inputs required of the user are the following:

$$
\begin{array}{lll}
\text{N} & \text{Number of data points in each process} & \\
\text{Pmax} & \text{Maximum order of predictive filter to be considered} & \\
\text{Nfft} & \text{Size of FFT to be used in spectral computation.} & (1)
\end{array}
$$

In addition to these integer inputs, the user must modify the subroutine SUB Data (N,X(*)) to accommodate and read in his particular two-channel data. All data are presumed real.

The program computes the (sample) means of each of the two processes and subtracts the means from the data. (Some possible ramifications of this procedure are considered in reference 6, appendix B; in addition, the effect of choosing too small an FFT size, Nfft, is discussed in reference 6.) Next, the covariance matrix (at zero delay) of the input data is computed, and the Akaike Information Criterion (AIC, reference 8, pages 42-44) is evaluated and used to select the integer

$$
\text{Pbest} \qquad \text{Best order of predictive filter to use.} \qquad (2)
$$

The forward and backward partial correlation coefficients (references 7-9) are evaluated through order Pmax, as well as the forward predictive filter coefficients for Pbest. The normalized correlation matrices are computed through Pmax (extrapolated values beyond Pbest) and the spectral density matrix is computed (via an FFT) from zero to Nyquist frequency, $f_N = (2\Delta)^{-1}$, where Δ is the time-sampling increment of the processes. A partial check on the adequacy of the FFT size, Nfft, is afforded by a printout of the areas under the spectral estimates and comparison with the (sample) covariances of the input data. Finally, the inverse FFT of the spectral estimate gives the aliased normalized correlation matrices; the motivation and equations for this approach are given in reference 9.

A sample printout for a short data sequence (20 data points in each process) is given after the program listing in appendix A, as a test or check case on a user-written program. Also, plots of the corresponding auto-spectral estimates and the coherence estimates are given there for completeness, although this example has no real physical significance.

Timing Results

Execution times for the five major subroutines,

$$
\begin{array}{lll}
\text{Pcc} & \text{Partial correlation coefficients} & \\
\text{Pfc} & \text{Predictive filter coefficients} & \\
\text{Peftf} & \text{Predictive error filter transfer function} & \\
\text{Sdm} & \text{Spectral density matrix} & \\
\text{Acm} & \text{Aliased correlation matrices,} & (3)
\end{array}
$$

are given in tables 1-5 below, for the HP 9845B Desk Calculator equipped with the Fast Processor Upgrade Kit. Only those variables utilized in each subroutine are considered in these tables, since execution time is independent of the other variables; for example, the execution time of subroutine Pcc does not depend on Pbest.

Table 1. Execution Times for Subroutine Pcc

N	Pmax	Seconds
20	6	1.9
50	10	5.4
100	5	4.7
100	10	9.5
100	15	14.1
1000	47	404.2

Table 2. Execution Times for Subroutine Pfc

Pmax	Pbest	Seconds
5	1	.09
10	1	.15
6	4	.24
15	5	.62
15	11	1.41
47	12	4.06

Table 3. Execution Times for Subroutine Peftf

Pbest	Nfft	Seconds
4	256	17.5
11	256	17.8
1	512	32.0
5	512	32.8
1	1024	63.9
11	1024	66.6

Table 4. Execution Times for Subroutine Sdm

Nfft	Seconds
256	8.9
512	17.7
1024	35.3

Table 5. Execution Times for Subroutine Acm

Nfft	Seconds
256	9.8
512	18.6
1024	37.7

From these tables, we are able to extract the following fairly accurate rules of thumb: the execution time of

Pcc is linearly dependent on N and Pmax
Pfc is linearly dependent on Pmax and Pbest
Peftf is linearly dependent on Nfft, but is essentially independent of Pbest
Sdm is linearly dependent on Nfft
Acm is linearly dependent on Nfft.   (4)

These rules allow extrapolation to other cases of interest to the user. The execution times of the FFT itself are given in table 6.

Table 6. Execution Times for Subroutine Fft10

Nfft	Seconds
128	2.6
256	4.5
512	8.4
1024	17.1

If the user is interested only in obtaining the predictive filter coefficients (for example, to do time domain prediction and signal processing), these results are available immediately after execution of subroutine Pfc. There is then no need to resort to the frequency domain routines that follow Pfc; in this manner, execution time and storage can be significantly reduced. An additional reduction in execution time is available by declaring all the loop counters in a subroutine to be INTEGER.

Application to Processes With Tones

Our first example is the two-channel case given numerically by the sample values in reference 7, page 17; reference 8, page K-12; and in reference 9, page D-18. The analytic expression for the autoregression is

$$x_1(k) = .85\, x_1(k-1) - .75\, x_2(k-1) + w_1(k)$$
$$x_2(k) = .65\, x_1(k-1) + .55\, x_2(k-1) + w_2(k) \tag{5}$$

where $\{w_1(k)\}$ and $\{w_2(k)\}$ are uniformly distributed, independent white noise processes with zero means and variances 1/12. General filter and spectral relations for moving-average and autoregressive processes are given in appendix B; these general relations are then specialized to this particular numerical example. It is shown in (B-31) et seq. that the auto spectrum of process $\{x_1(k)\}$ has *four poles and three zeros* in the finite z-plane, even though the two-channel recursion, (5), is only first-order regressive.

Generally, for a two-channel P-th order regression and independent white excitations (i.e., $E_k = 0$ for $k > P$, $F_k = I\,\delta_{ko}$, and $Q(z) = \Delta I$ in (B-18)), the auto- and cross-spectra of the processes each possess 4P poles and 3P zeros in the finite z-plane (of which P zeros occur at the origin). This is in contrast to the single-channel case, where 2P poles (and a P-th order zero only at the origin) can occur. This increased generality can be anticipated by the observation that whereas a single-channel approximation requires estimation of only P parameters, an M-channel approximation requires estimation of M^2P parameters (4P for the two-channel case $M = 2$). Of course, for a fixed number, N, of data points from each process, the estimation of an increased number of parameters can only be done with increased variance; this is a manifestation of the tradeoff between resolution and stability that accompanies all spectral analysis techniques.

The first-order forward partial correlation coefficient for two-channel process (5) is

$$A_1^{(1)}(\text{true}) = \begin{bmatrix} .85 & -.75 \\ .65 & .55 \end{bmatrix}, \tag{6}$$

and all other higher-order coefficients are zero. The exact auto spectrum of the first process, $\{x_1(k)\}$, is shown (in dB) in figure 1A; the auto spectrum of the second process, $\{x_2(k)\}$, is shown in figure 1B; the magnitude-squared coherence is displayed in figure 1C; and the argument of the complex coherence or cross spectrum is depicted in figure 1D. There is seen to be a strong narrowband component at approximately one-fourth of the Nyquist frequency $f_N = (2\Delta)^{-1}$, where Δ is the time-sampling increment for the two-channel process (5). This leads to a peak magnitude-squared coherence value of .999013 at $2f\Delta = .2459$.

The results of applying the two-channel spectral analysis program in appendix A to the numerical data cited above, with $N = 100$, are shown in figure 2, where the
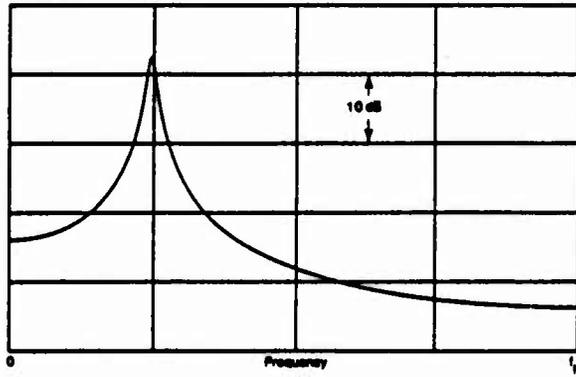
1A. Auto Spectrum of First Process
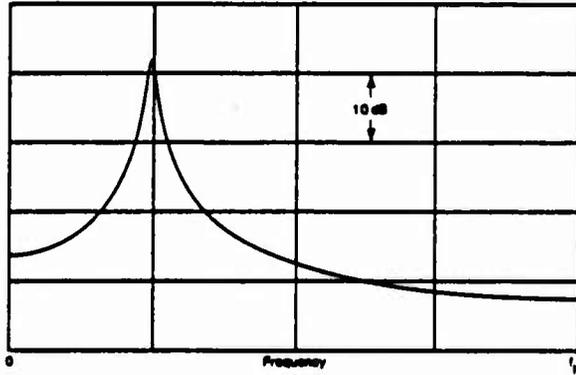
1B. Auto Spectrum of Second Process
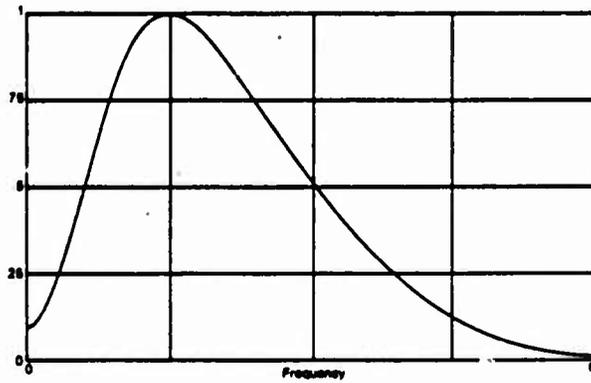
1C. Magnitude-Squared Coherence

1D. Argument of Complex Coherence

Figure 1. True Spectral Characteristics With No Tone

2A. Auto Spectrum of First Process

2B. Auto Spectrum of Second Process
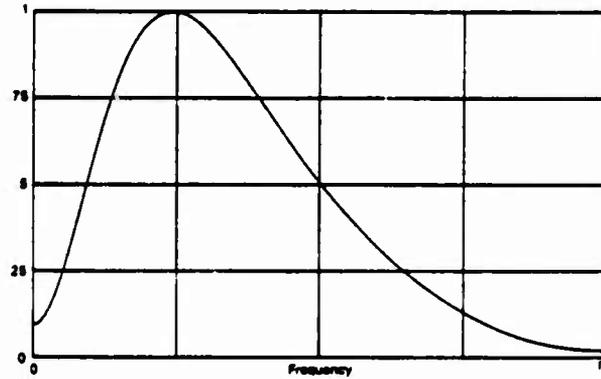
2C. Magnitude-Squared Coherence
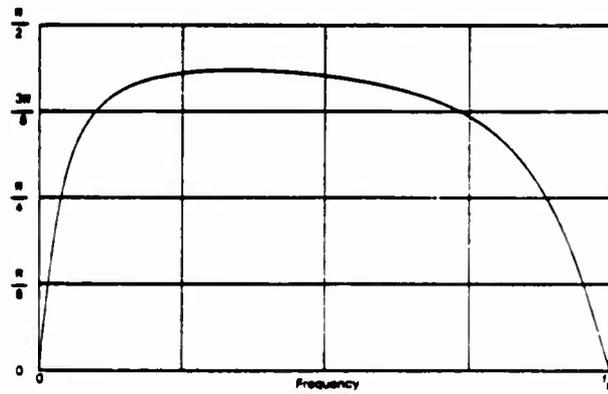
2D. Argument of Complex Coherence

Figure 2. Spectral Estimates for N = 100, Pmax = 15, Pbest = 1, No Tone

four parts of this figure correspond directly to those of figure 1. Pbest turns out to be equal to the correct value 1, and the spectral estimates are all quite good. In fact, the estimated magnitude-squared coherence reaches a peak value of .999745 versus the true value of .999013.

The covariance matrix of the process generated by (5) is (reference 8, page 18, after scaling by variance $1/12$)

$$R_o(true) = \begin{bmatrix} 2.095 & 0.405 \\ 0.405 & 1.804 \end{bmatrix} . \tag{7}$$

The corresponding matrix estimate yielded by the program here, based on the particular $N = 100$ data values cited above, is

$$R_o = \begin{bmatrix} 4.62 & .916 \\ .916 & 3.80 \end{bmatrix} \text{ for } N = 100; \tag{8}$$

these values are approximately 2.2 times larger than (7), due to the fact that (5) is a narrowband process and the particular 100 pairs of samples used in the spectral estimates happen to lie on a local peak of the instantaneous waveforms. Although the local estimates of the absolute power levels are off considerably, the estimate of the forward partial correlation coefficient is very good; we find, instead of (6),

$$A_1^{(1)} = \begin{bmatrix} .872 & -.770 \\ .634 & .560 \end{bmatrix} \text{ for } N = 100. \tag{9}$$

Next we add a pure tone only to the second process $\{x_2(k)\}$ at a frequency equal to 0.6 of the Nyquist frequency, i.e., at $0.6f_N$. The power in the tone is $1/512$, i.e., 32.9 dB below the average power, 3.80, in this particular segment of autoregressive process $\{x_2(k)\}$; see (8). The resultant spectral estimates are shown in figure 3; they are virtually identical to figure 2. The only inadequacy of figure 3 is that the autospectral estimate in figure 3B gives no indication of the added tone; of course, there should ideally be no indication of the tone in figure 3A for the auto spectrum of $\{x_1(k)\}$. The value of Pbest was again 1, as determined by the AIC.

When the tonal power in the second process is increased to -26.9 dB, Pbest increases to 4 (see figure 4) and there are humps in both auto-spectral estimates near the tone frequency $0.6f_N$. The coherence estimates (magnitude and argument) are significantly perturbed in a considerable neighborhood of $0.6f_N$; this broad frequency-perturbation width is due to a small value of Pbest having been selected by the AIC.

Increasing the tonal power to -20.8 dB results in the estimates depicted in figure 5. Now there is a considerable indication of the tonal power in figure 5B; however, there is also an *undesirable* indication in figure 5A at frequency $0.6f_N$ in the auto-spectral estimate for process $\{x_1(k)\}$. This "feed-across" is due to the fact that we are working with only $N = 100$ data samples of each process; with this small a data set, the "best" two-channel linear-prediction is misled into an erroneous indication. It is important to observe at this point that *any* auto-spectral estimate based on
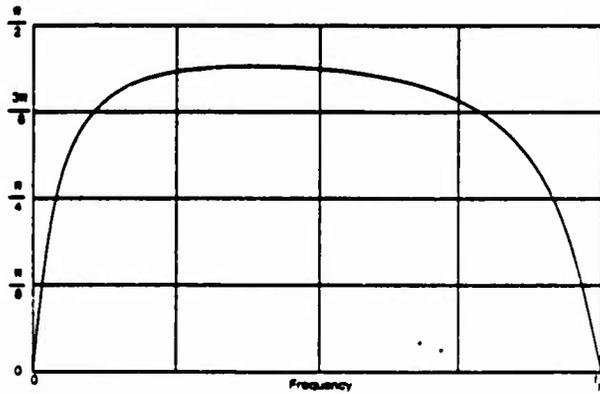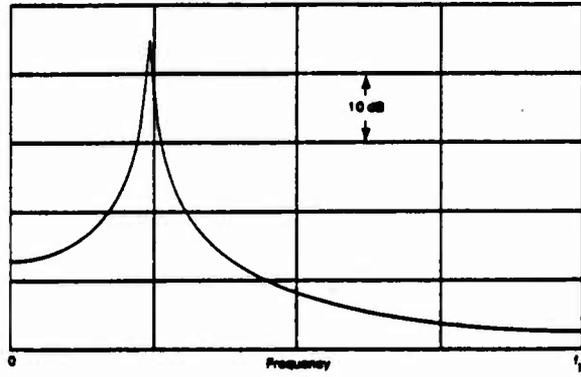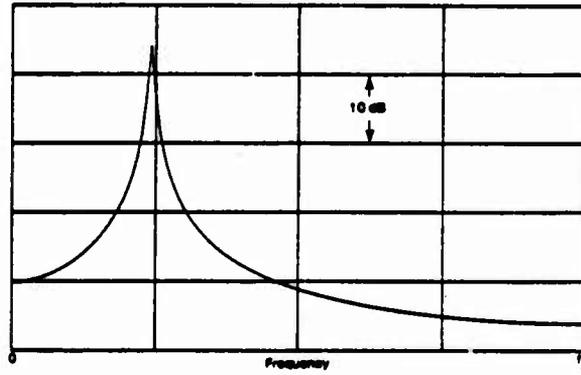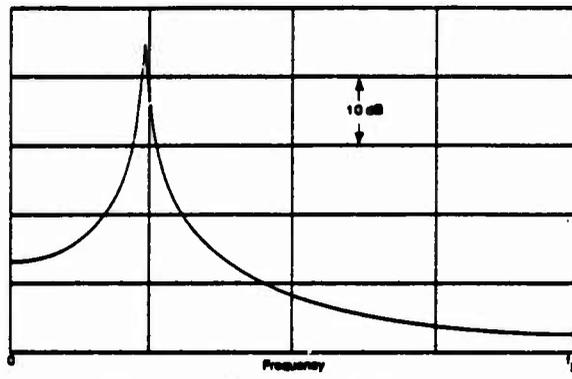
3A. Auto Spectrum of First Process
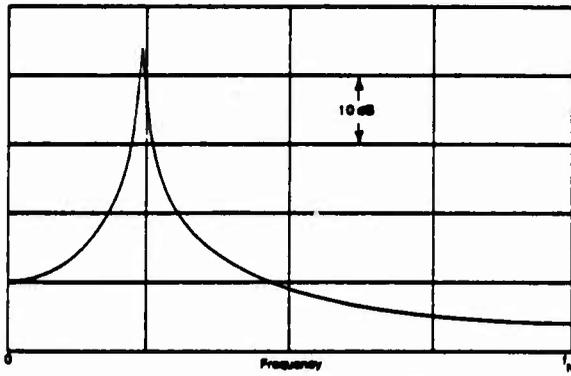
3B. Auto Spectrum of Second Process
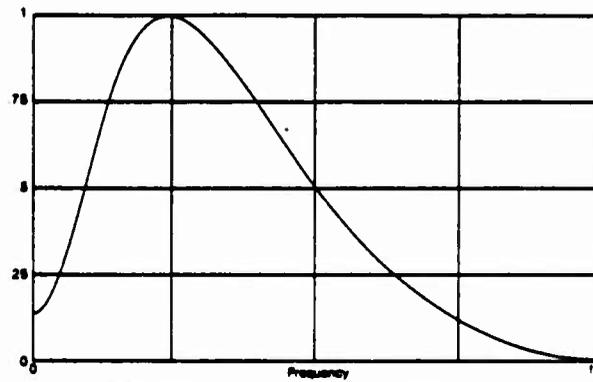
3C. Magnitude-Squared Coherence
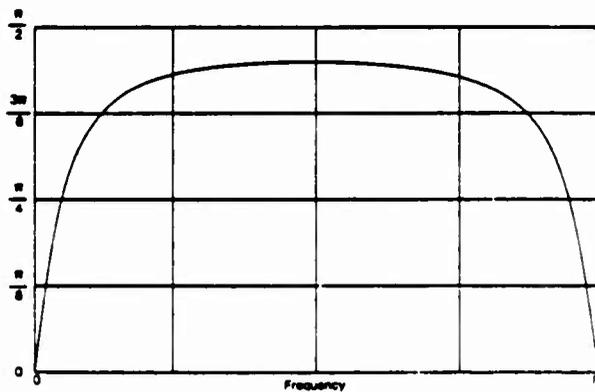
3D. Argument of Complex Coherence

Figure 3. Spectral Estimates for N = 100, Pmax = 15, Pbest = 1, Tone Power = −32.9 dB

4A. Auto Spectrum of First Process

4B. Auto Spectrum of Second Process

4C. Magnitude-Squared Coherence

4D. Argument of Complex Coherence

Figure 4. Spectral Estimates for $N = 100$, Pmax = 15, Pbest = 4, Tone Power = –26.9 dB

5A. Auto Spectrum of First Process

5B. Auto Spectrum of Second Process

5C. Magnitude-Squared Coherence

5D. Argument of Complex Coherence

Figure 5. Spectral Estimates for N = 100, Pmax = 15, Pbest = 8, Tone Power = -20.8 dB

samples of process $\{x_1(k)\}$ *alone* would not give this tonal indication, since the tone is not present in this process.

The coherence estimates in figure 5 fare no better, even though Pbest = 8 now. A large magnitude-squared coherence value of 0.85 is yielded at frequency $0.6f_N$. The progression towards poorer behavior is also present in figure 6, which employs a tonal power of –14.8 dB relative to the sample power in $\{x_2(k)\}$. Now the undesired peak magnitude-squared coherence estimate is 0.9. A tonal power of –8.8 dB (figure 7) yields a near-unity magnitude-squared coherence estimate at $0.6f_N$, and a very substantial tonal indication in the auto spectrum of $\{x_1(k)\}$, figure 7A.

The situation is markedly improved if more data samples are available. When N is increased to 1000, and data are generated via (5) as before, the sample covariance for the particular data set generated is

$$R_o = \begin{bmatrix} 2.60 & .514 \\ .514 & 2.27 \end{bmatrix} \text{ for } N = 1000, \tag{10}$$

for no tone present. When a tone is added to process $\{x_2(k)\}$, with strength –24.6 dB relative to the sample power, 2.27, of the second process, the resultant spectral estimates are as displayed in figure 8. There is a slight hump at $0.6f_N$ in figure 8B, and a near-zero coherence estimate at this frequency. Recall that the ideal characteristics would be identical to figure 1 except for an impulse in figure 8B at $0.6f_N$ and a very sharp null in the magnitude-squared coherence at $0.6f_N$.

The results in figure 8 were achieved by taking Pmax = 8, for which the AIC indicated Pbest = 8 for this particular data set. However, the AIC is a very flat function of filter order P in this range, and it is difficult to justify a particular value of P as "best". Some additional information about the autoregressive portion of the observed process, such as a limit on P, could be useful; for example, when we specified Pmax as 1, the results were very similar to figure 1. There was virtually no indication of the tone in any of the spectral estimates, even though it was in the $\{x_2(k)\}$ data at a relative level of –24.6 dB with respect to the sample power, 2.27, of the second autoregressive component. In fact, the estimated first partial correlation coefficient was

$$A_1^{(1)} = \begin{bmatrix} .8543 & -.7394 \\ .6578 & .5415 \end{bmatrix} \text{ for } N = 1000, \tag{11}$$

which is very close to the true value, (6).

Results for the spectral estimates when the tonal power is increased to –18.6 dB, –12.6 dB, –6.6 dB, and –0.6 dB are given in figures 9, 10, 11, and 12, respectively, all corresponding to Pmax = 8 and Pbest = 8. Even for the nearly 0 dB case in figure 12, there is virtually no indication in auto-spectral estimate 12A of the strong tonal in process $\{x_2(k)\}$, figure 12B. The magnitude-squared coherence estimate in figure 12C appears to have developed a couple of zeros and poles near the frequency $f = 0.6f_N$, where the strong tone is located; recall that we have $4P = 32$ poles available in the approximation for Pbest = 8. Typically, it has been observed that a strong tonal present in only one process manifests itself in the

6A. Auto Spectrum of First Process

6B. Auto Spectrum of Second Process

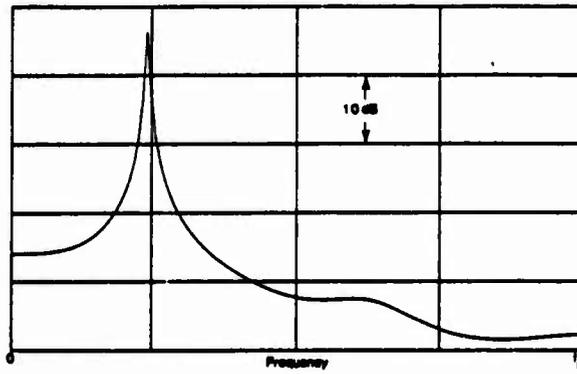6C. Magnitude-Squared Coherence

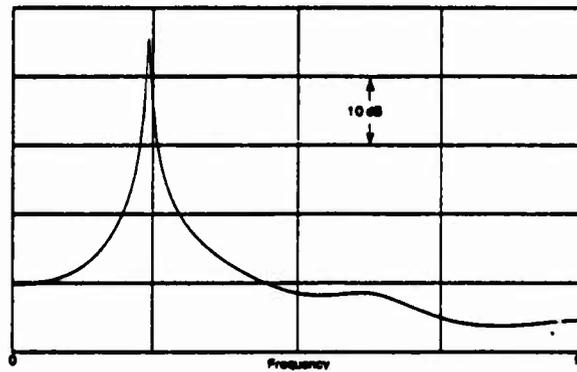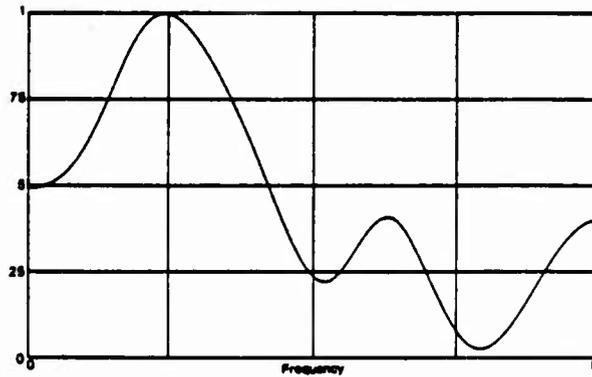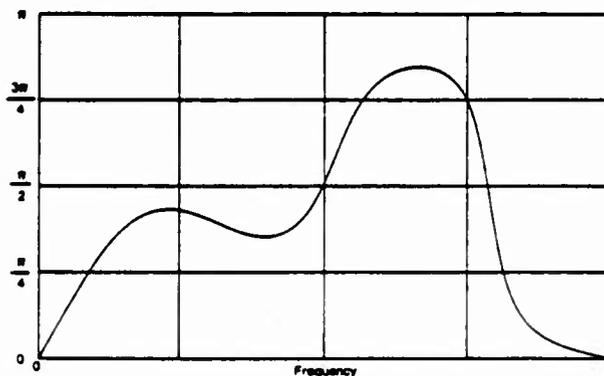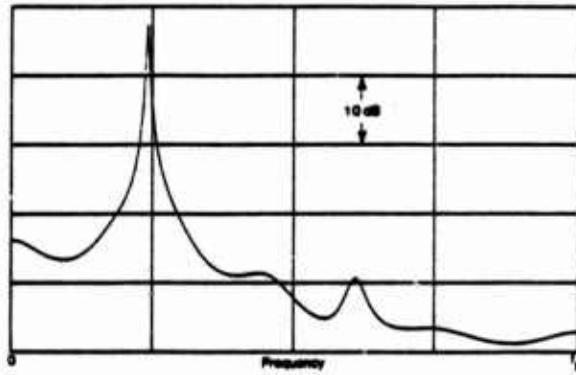6D. Argument of Complex Coherence
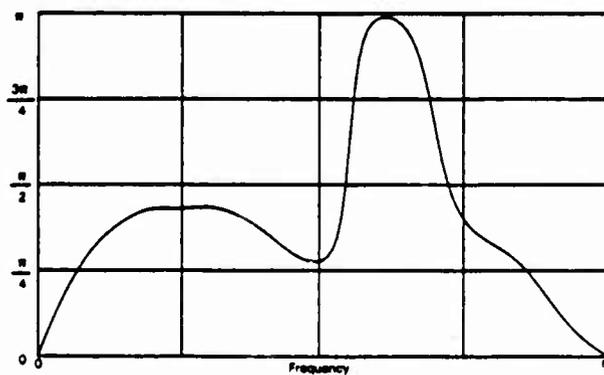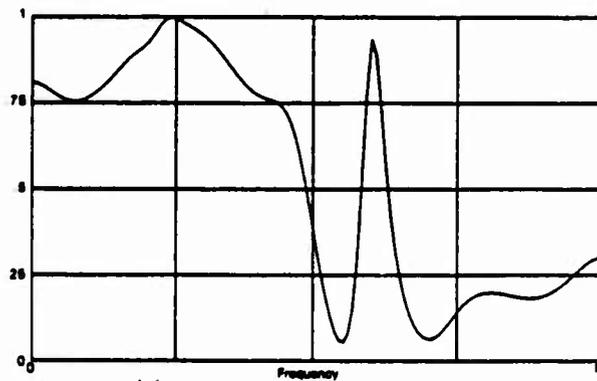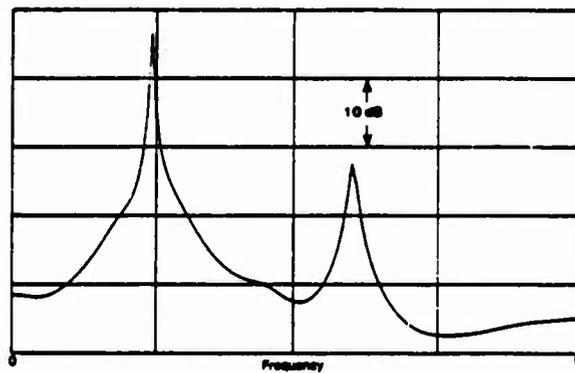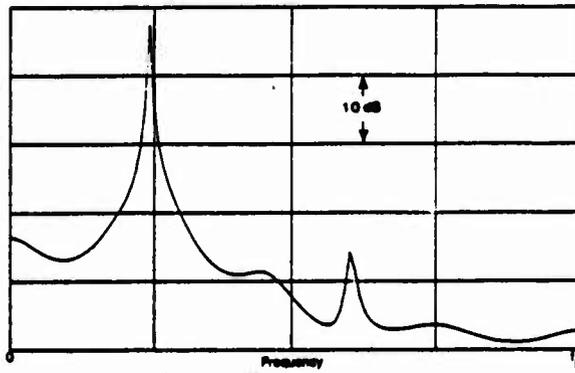
Figure 6. Spectral Estimates for N = 100, Pmax = 15, Pbest = 8, Tone Power = -14.8 dB

7A. Auto Spectrum of First Process

7B. Auto Spectrum of Second Process

7C. Magnitude-Squared Coherence

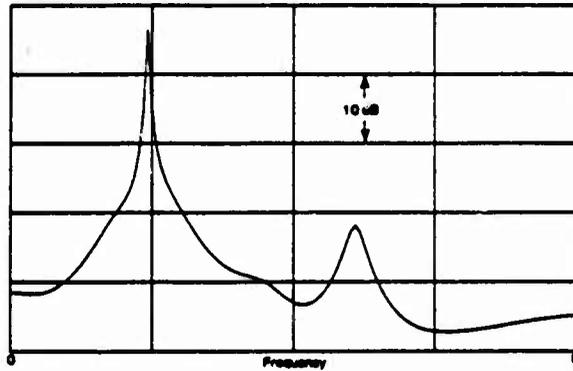7D. Argument of Complex Coherence
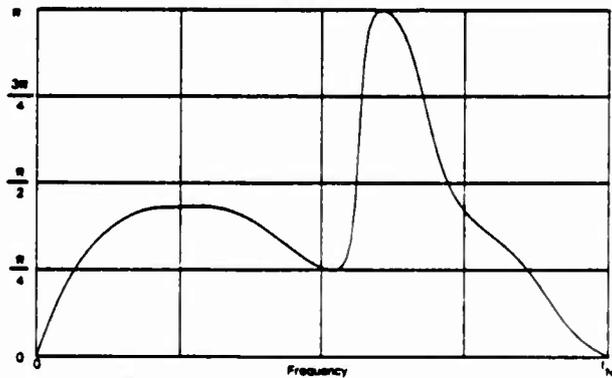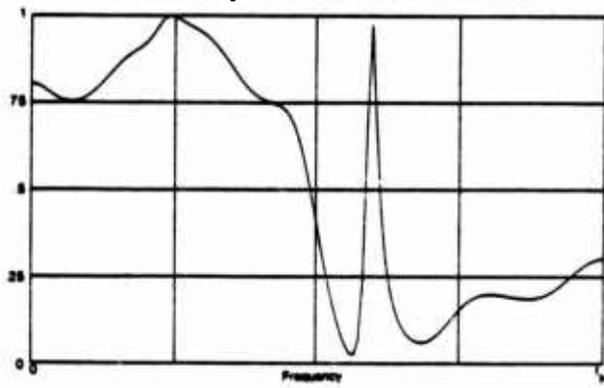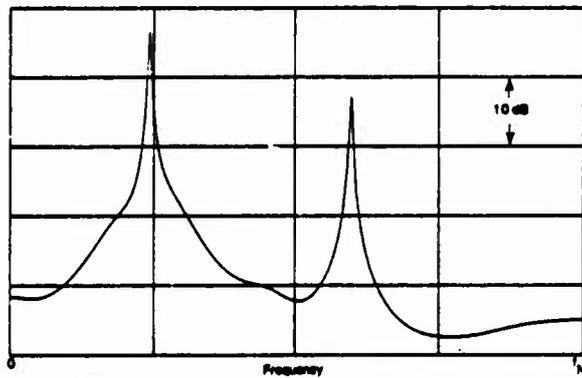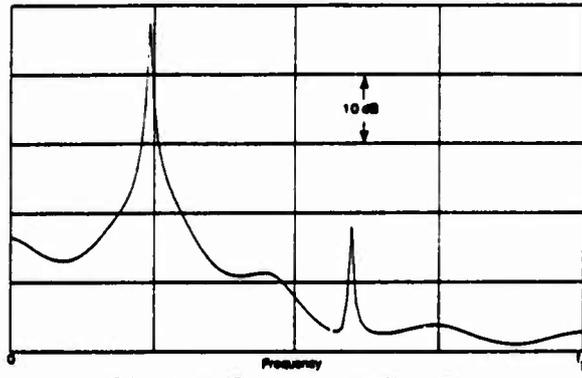
Figure 7. Spectral Estimates for N = 100, Pmax = 15, Pbest = 8, Tone Power = -8.8 dB
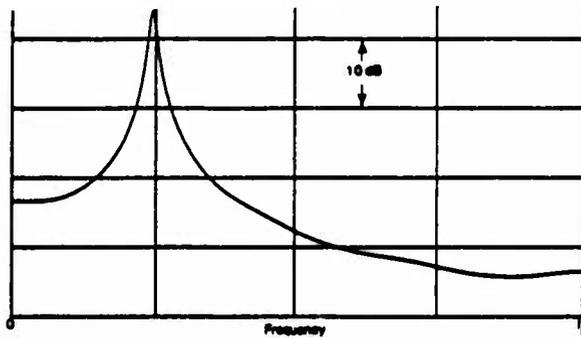
8A. Auto Spectrum of First Process
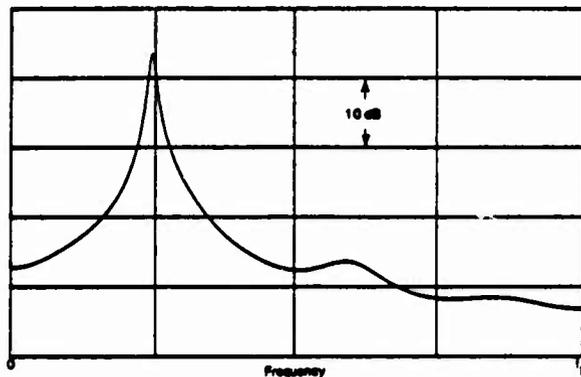
8B. Auto Spectrum of Second Process
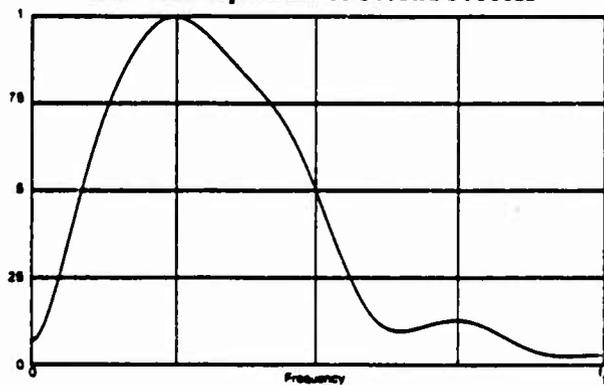
8C. Magnitude-Squared Coherence
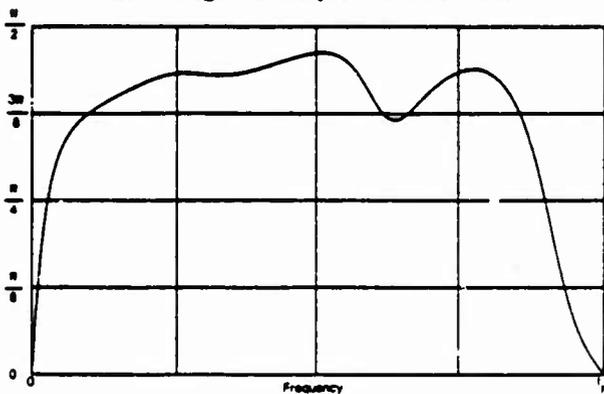
8D. Argument of Complex Coherence
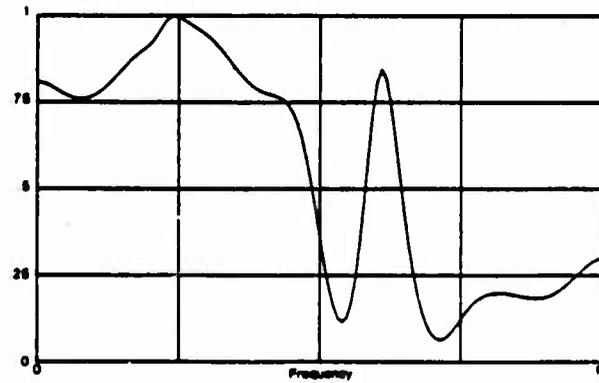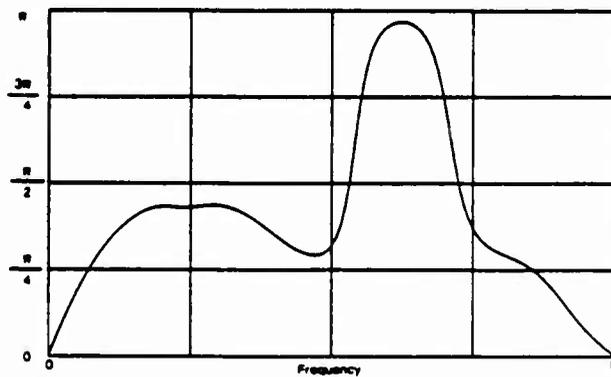
Figure 8. Spectral Estimates for N = 1000, Pmax = 8, Pbest = 8, Tone Power = -24.6 dB
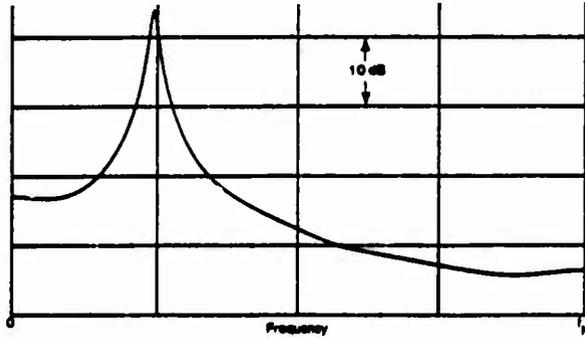
9A. Auto Spectrum of First Process
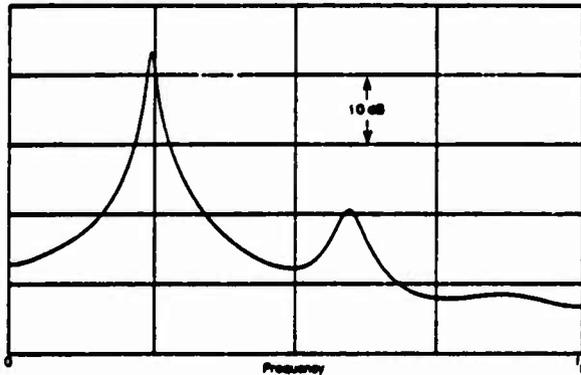
9B. Auto Spectrum of Second Process
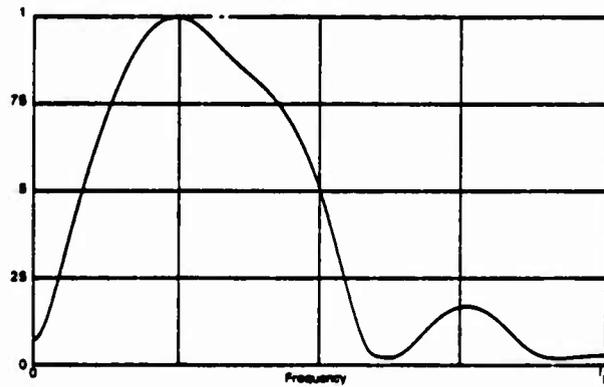
9C. Magnitude-Squared Coherence
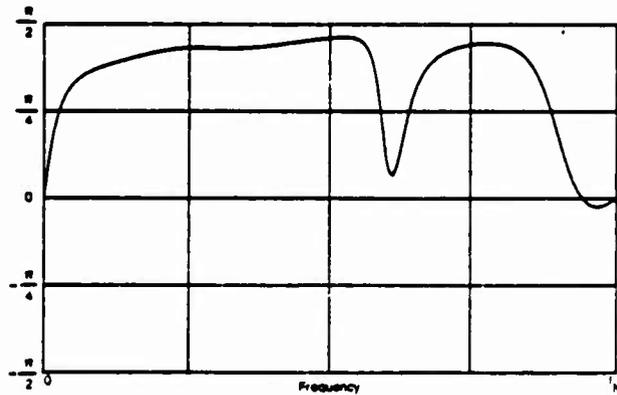
9D. Argument of Complex Coherence

Figure 9. Spectral Estimates for $N = 1000$, Pmax = 8, Pbest = 8, Tone Power = –18.6 dB

16
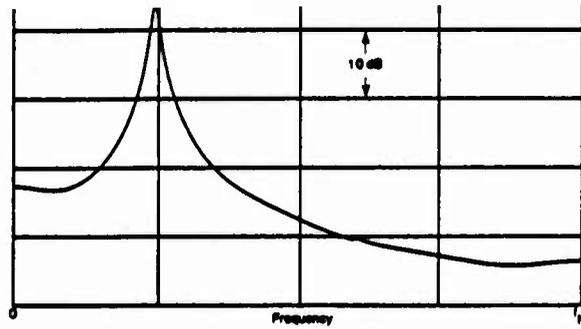
10A. Auto Spectrum of First Process
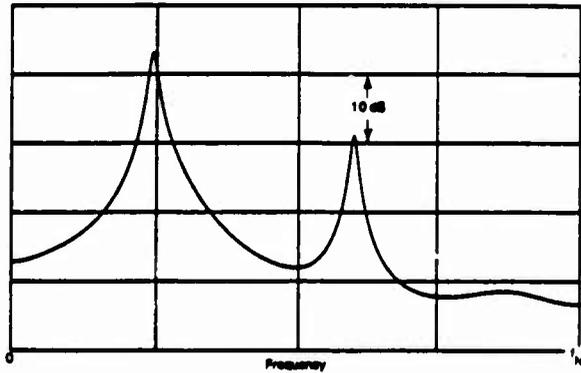
10B. Auto Spectrum of Second Process
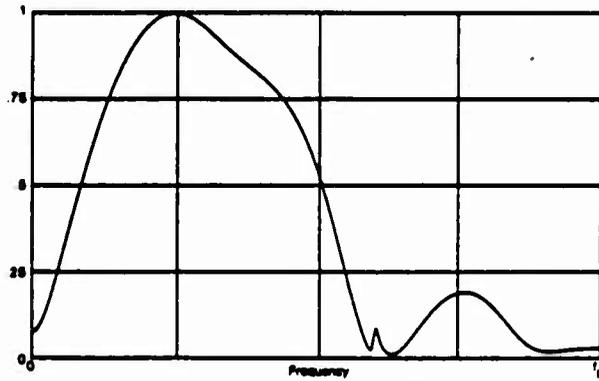
10C. Magnitude-Squared Coherence
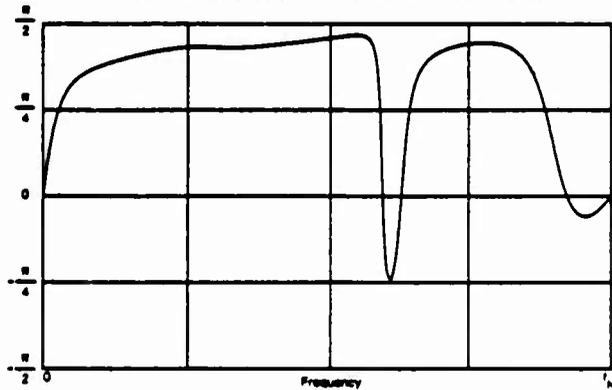
10D. Argument of Complex Coherence

Figure 10. Spectral Estimates for N = 1000, Pmax = 8, Pbest = 8, Tone Power = –12.6 dB

17

11A. Auto Spectrum of First Process

11B. Auto Spectrum of Second Process

11C. Magnitude-Squared Coherence

11D. Argument of Complex Coherence
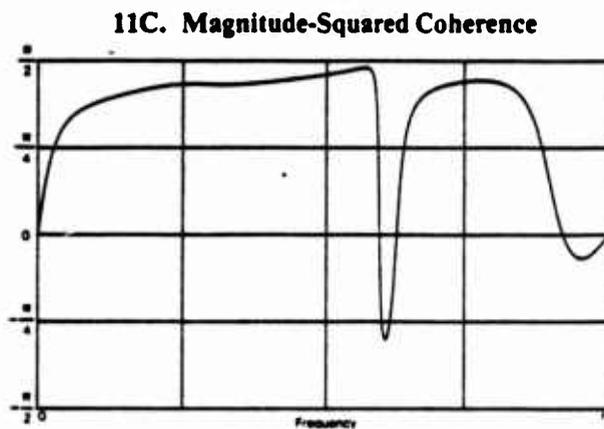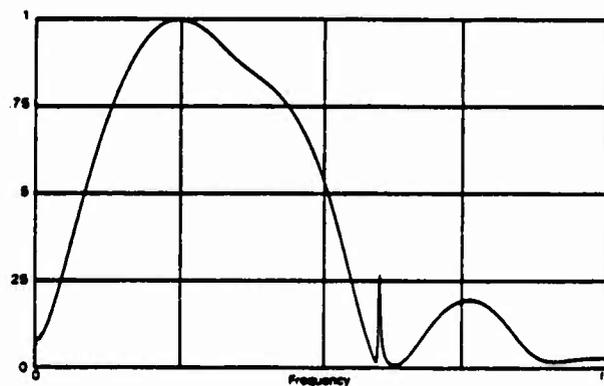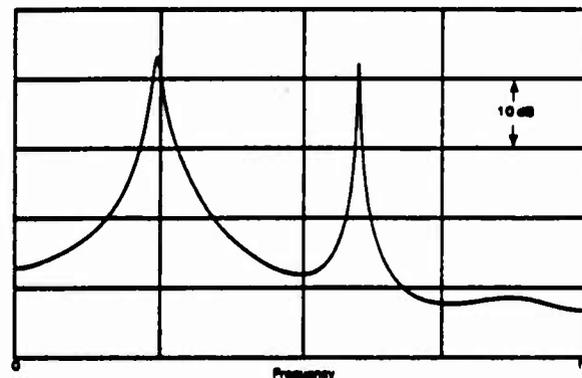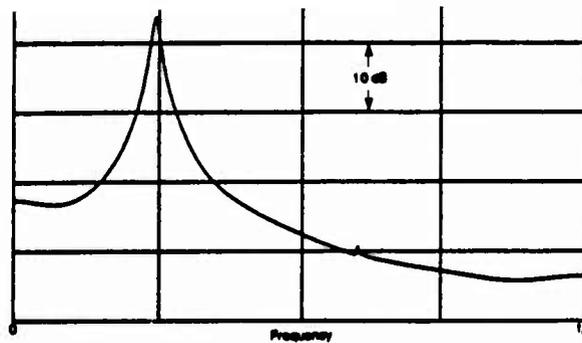
Figure 11. Spectral Estimates for N = 1000, Pmax = 8, Pbest = 8, Tone Power = -6.6 dB
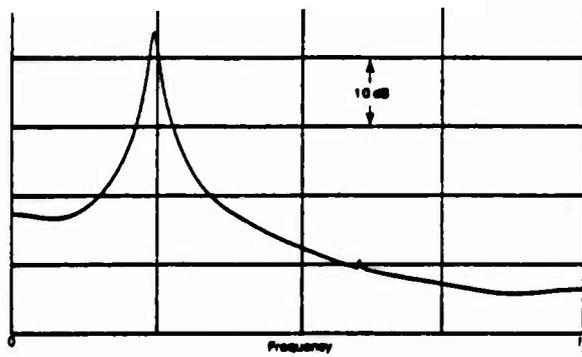
12A. Auto Spectrum of First Process
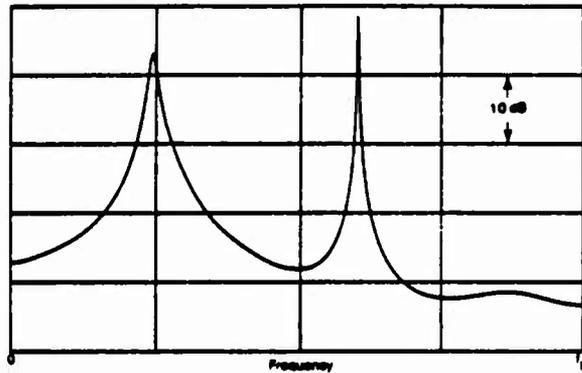
12B. Auto Spectrum of Second Process
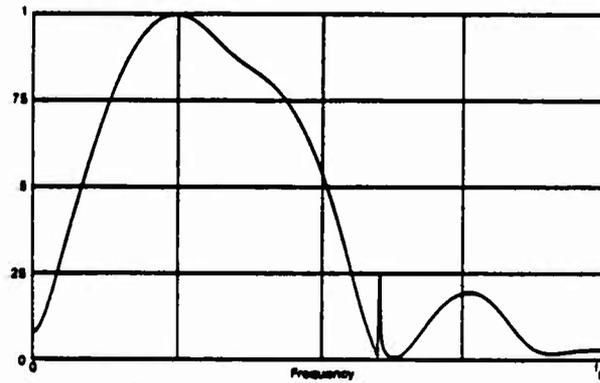
12C. Magnitude-Squared Coherence
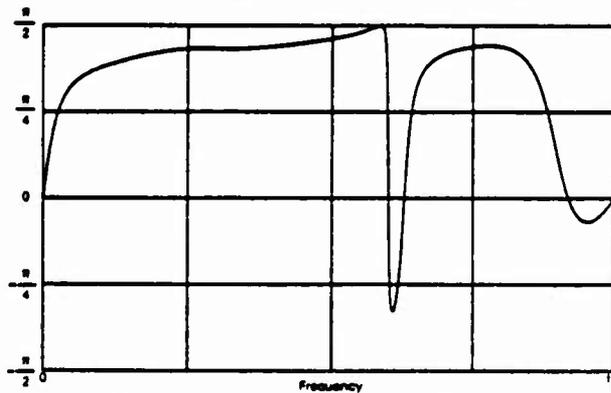
12D. Argument of Complex Coherence

Figure 12. Spectral Estimates for $N = 1000$, Pmax = 8, Pbest = 8, Tone Power = -0.6 dB

19

coherence estimate as a sharp spike at the tone frequency. The change in argument in figure 12D in the neighborhood of this frequency can serve as an indicator of the number of poles and zeros clustered there.

When Pmax was increased to 47, the AIC yielded Pbest = 25 for these last four figures. However, the spectral estimates for Pbest = 25 proved to be too spiky and erratic. Also the selection of Pbest at 25 is rather tenuous, as figure 13 indicates; this is a plot of the AIC versus filter order P in the range (1, 47). Although the absolute minimum occurs at P = 25, there are significant drops in the curve at P = 4, 6, and 8. Selection of P at one of these significant drops appears to be a promising approach, instead of using the absolute minimum of the curve. In addition, the flatness of the curve is brought out by observing that the range of values of AIC is limited to (-4.80, -4.73) for P in the wide range from 10 to 47. Thus, the local minor drops and rises in the AIC curves are not significant; selection of values of P corresponding to significant decreases seems to be a viable approach.
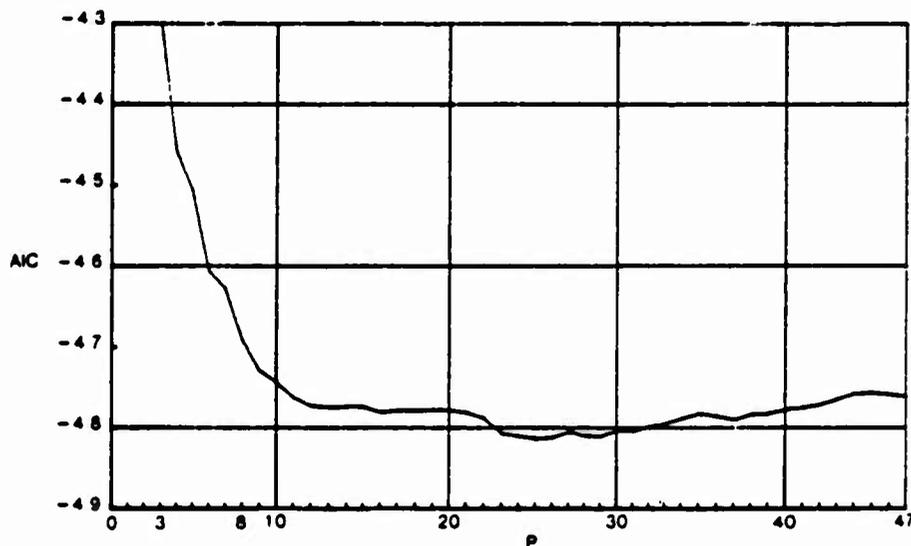
Figure 13. Akaike Information Criterion for N = 1000, Tone Power = -12.6 dB

The last example we consider is a two-channel process of N = 64 data points, composed of several tones, some of which are at common frequencies, and some of which are not; this example was supplied by S. L. Marple (reference 10). In particular, process $\{x_1(k)\}$ has two strong tones at $f = 0.4f_N$ and $0.5f_N$, and a weaker tone (-20 dB) at $f = 0.2f_N$, in addition to some low level, colored background noise. The other process has two strong tones at $f = 0.4f_N$ and $0.8f_N$, and a weaker tone (-20 dB) at $f = 0.2f_N$. Thus the tonal frequencies common to both processes are $0.2f_N$ and $0.4f_N$, whereas the uncommon frequencies are $0.5f_N$ and $0.8f_N$. The two auto-spectral estimates of each process (obtained via the single-channel, forward-backward averaging technique of reference 4) are displayed in figure 14 for prediction length P = 12 (24 poles for each spectral estimate). There is, of course, no cross-feed at frequencies $0.5f_N$ and $0.8f_N$.

The spectral estimates of the same two-channel data (obtained via the program in appendix A which includes coherence estimation) are given in figure 15. The value

of P used was 6, which allows for 24 poles in each spectral estimate. The low number of data points, N = 64, now allows some undesired cross-feed in figures 15A and 15B at f = 0.8f$_N$ and 0.5f$_N$, respectively. This also shows up in the magnitude-squared coherence estimate as two very sharp spikes at these two frequencies, whereas the true coherence is zero at these two frequencies. This limited capability of the multi-channel linear predictive technique can be improved by utilizing larger data sets; N = 64 is too small a data size to accomplish a high quality result for a data set such as this with strong interfering tones.
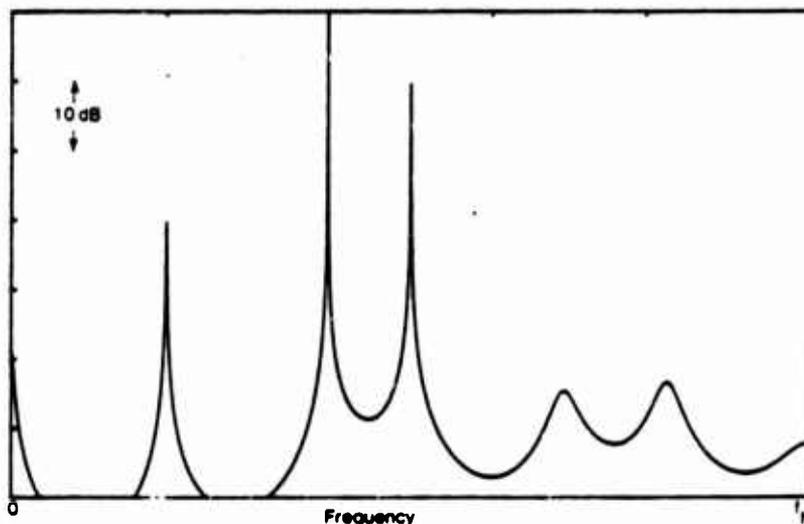
14A. Auto Spectrum of First Process
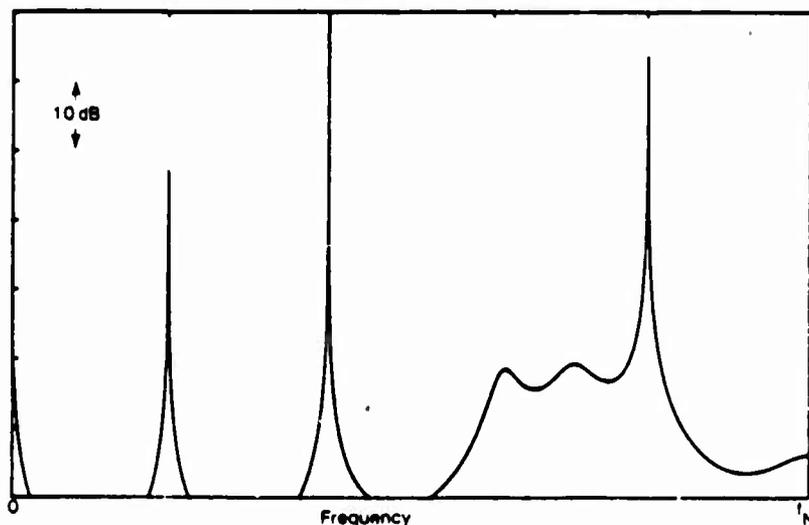
14B. Auto Spectrum of Second Process

Figure 14. Auto-Spectral Estimates for Multitone Example, N = 64, P = 12

21
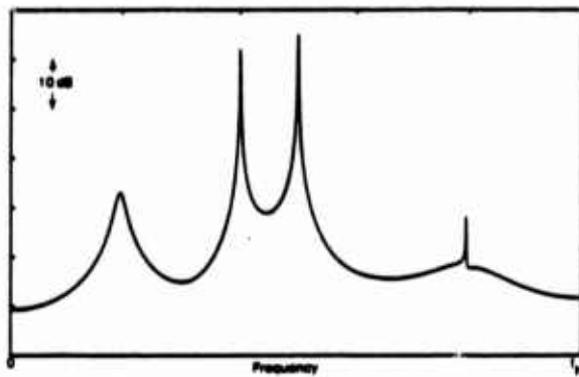
15A. Auto Spectrum of First Process
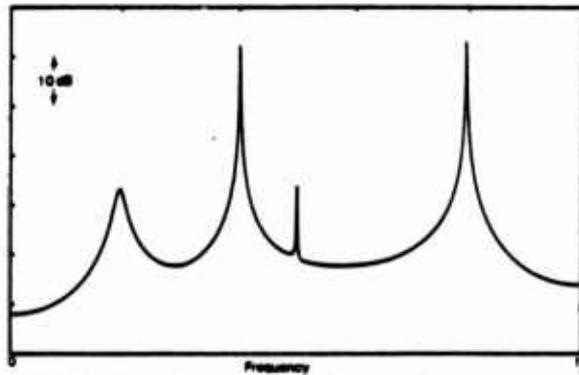
15B. Auto Spectrum of Second Process
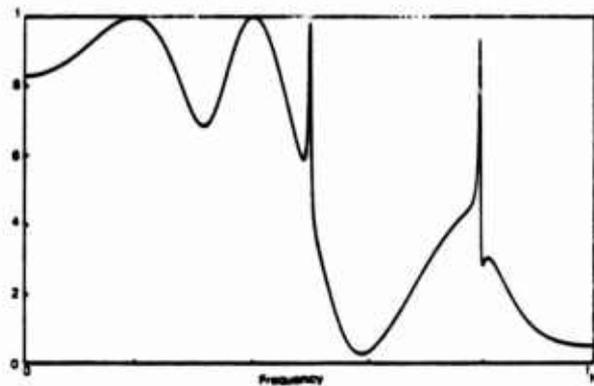
15C. Magnitude-Squared Coherence
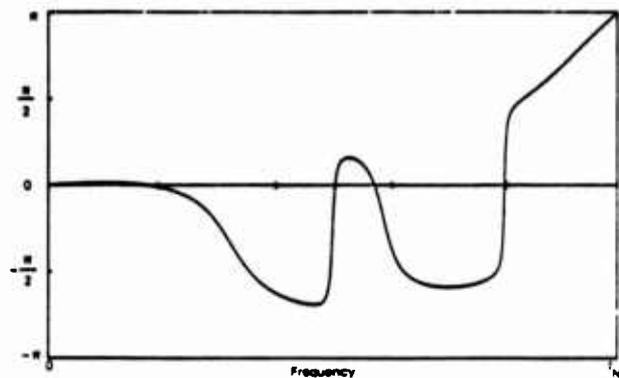
15D. Argument of Complex Coherence

Figure 15. Spectral Estimates for Multitone Example, N = 64, P = 6

22

Discussion and Conclusions

A program for two-channel auto- and cross-spectral estimation has been presented and illustrated for cases including interfering tones. If the number of available points, N, is too small, some misleading estimates may be obtained because of cross-feed between the finite lengths of data from each channel. This cross-feed manifests itself as narrow spurious spikes in the spectral and coherence estimates. Choice of the appropriate value of filter order, P, is possible by observing where the AIC undergoes significant negative jumps, rather than by using the absolute minimum of the curve. It can also be illuminating to overlay plots made with two (or more) different values of P, thereby obtaining different degrees of resolution and stability from the same data set. The recursive nature of the linear-predictive approach makes this practice easy to achieve.

A more fundamental observation about spectral estimation in general is now developed. Suppose we are given finite data records of three stationary processes $x(t)$, $y(t)$, and $z(t)$, and we wish to estimate all the auto spectra and cross spectra involved. The Blackman and Tukey and weighted-FFT approaches evaluate the auto spectrum of each process separately. Thus, the spectrum of $x(t)$ is estimated without interference from $y(t)$ and $z(t)$; the availability of the data records for $y(t)$ and $z(t)$ plays no part in the eventual auto-spectral estimate for $x(t)$. Additionally, the cross-spectral estimate for processes $x(t)$ and $y(t)$ is independent of the available data on the $z(t)$ process. Finally, the coherence estimate between two processes is independent of any additional data records for other (statistically related) processes.

On the other hand, the generalization (in references 7-9) of Burg's single-channel linear-predictive spectral analysis approach to the multichannel case gives auto-spectral estimates of the $x(t)$ process that are dependent on the available values of $y(t)$ and $z(t)$. Also, the cross-spectral estimate between $x(t)$ and $y(t)$ is dependent on the particular $z(t)$ data available. This procedure can be poor for short data lengths if, for example, $y(t)$ contains a strong tone at f_o that is not present in $x(t)$ or $z(t)$. Thus, estimates of spectra $G_{xx}(f)$, $G_{xy}(f)$, and $G_{xz}(f)$ all contain tonal indications at f_o that should not be there. These spurious tonal indications are due to cross-feed between the available finite data segments of the various processes.

This raises the following questions:

- Should the estimate of $G_{xx}(f)$ be determined only from the available $x(t)$ data record ?

- Should the estimate of $G_{xy}(f)$ be determined only from the available $x(t)$ and $y(t)$ data records ?

- If coherence $C_{xy}(f_o) = 0$, why use $y(t)$ to estimate $G_{xx}(f_o)$?

- If coherence $C_{xy}(f_o) = 1$, why use the completely statistically dependent $y(t)$ data to estimate $G_{xx}(f_o)$?

This philosophy of discarding "irrelevant" data would be consistent with the Blackman and Tukey and FFT approaches. Carrying this philosophy on, we are led to the following: estimate $G_{xx}(f)$ solely from the $x(t)$ data by some good single-

channel linear-predictive technique, such as forward-and-backward averaging, coupled with an efficient way of inverting the relevant matrices (e.g., references 4 and 5). Then estimate cross spectrum $G_{xy}(f)$ or coherence $C(f)$ directly, by some (yet unknown) linear predictive technique whose sole goal is linear prediction of $x(t)$ from $y(t)$ and vice versa, with no interest in or diversion from simultaneous estimation of $G_{xx}(f)$ or $G_{yy}(f)$. By this means, we can concentrate on extracting all the relevant cross-spectral information with maximum stability and resolution. Other cross spectra of interest between particular pairs of available processes can be similarly obtained, one at a time. This procedure is currently under investigation.

Appendix A

Program For Two-Channel Linear-Predictive Spectral Analysis

The program listing below in Basic for the HP 9845B Desk Calculator is a translation and update of that given in references 7-9. A complete breakdown and explanation of the components and subroutines of the program are given in reference 8, and in reference 9, appendix D.

Inputs required of the user are the integers listed in lines 20, 30, 40; they are

N	Number of data points in each process;
Pmax	Maximum order of predictive filter to be considered;
Nfft	Size of FFT to be used in spectral analysis.

In addition, the user must modify subroutine SUB Data(N,X(*)) in lines 5430-5490 at the end of the program to read in his own particular two-channel data sets. Pbest can be forced to equal Pmax by setting Fac = 0 in SUB Pcc.

An explanation of the program output is given under equation (1) of the main text of this report. A sample printout for a short (N = 20) data sequence that can be used as a check case on the program is presented following the listing below. Sample plots of the auto-spectral estimates and the coherence estimates conclude the appendix.

```
10  ! TWO-CHANNEL LINEAR PREDICTIVE SPECTRAL ANALYSIS, TR 5501 & 5729
20      N=20                    ! NUMBER OF DATA POINTS IN EACH PROCESS
30      Pmax=6                  ! MAXIMUM ORDER OF PREDICTIVE FILTER
40      Nfft=256                ! SIZE OF FFT
50      OPTION BASE 1
60      REDIM Y(2,N),Z(2,N),Ap(Pmax,2,2),Bp(Pma.,2,2)
70      REDIM Rn(Pmax,2,2),Aic(0:Pmax),X11(Nfft),Y11(Nfft),Y12(Nfft)
80      REDIM Y12(Nfft),X21(Nfft),Y21(Nfft),X22(Nfft),Y22(Nfft)
90      DIM Y(2,1000),Z(2,1000),Ap(25,2,2),Bp(25,2,2)
100     DIM Rn(25,2,2),Aic(0:25),X11(1024),Y11(1024),X12(1024)
110     DIM Y12(1024),X21(1024),Y21(1024),X22(1024),Y22(1024)
120     DIM Ave(2),Ubest(2,2),U(2,2),V(2,2),Ui(2,2),Vi(2,2),A(2,2)
130     DIM B(2,2),R(2,2),Wa(2,2),Wb(2,2),Wc(2,2),Wd(2,2),We(2,2)
140     PRINT "NUMBER OF DATA POINTS IN EACH PROCESS N =";N
150     PRINT "MAXIMUM ORDER OF PREDICTIVE FILTER Pmax =";Pmax
160     PRINT "SIZE OF FFT Nfft =";Nfft
170     PRINT
180     CALL Data(N,Y(*))
190     PRINT "PROCESS NUMBER 1"
200     FOR I=1 TO N
210     PRINT Y(1,I);
220     NEXT I
230     PRINT LIN(1)
240     PRINT "PROCESS NUMBER 2"
250     FOR I=1 TO N
260     PRINT Y(2,I);
270     NEXT I
280     PRINT LIN(2)
290     CALL Pcc(N,Pmax,Y(*),Z(*),Ave(*),Wa(*),Wb(*),Wc(*),Wd(*),We(*),R(*),U(*),V
(*),Aic(*),Pbest,Ubest(*),Ui(*),Vi(*),A(*),B(*),Ap(*),Bp(*))
300     PRINT "MEANS OF INPUT DATA (Ave):"
310     PRINT Ave(1)
320     PRINT Ave(2)
330     PRINT LIN(1)
340     PRINT "COVARIANCE MATRIX OF INPUT DATA (R):",R(*)
350     PRINT "AKAIKE INFORMATION CRITERION:"
360     PRINT "   P          Aic(P)"
370     IMAGE 3D,4(4X,M. 9DE)
380     FOR P=0 TO Pmax
390     PRINT USING 370;P,Aic(P)
400     NEXT P
410     PRINT LIN(1)
420     PRINT "Pbest =";Pbest
430     PRINT LIN(1)
440     PRINT "Ubest:",Ubest(*)
450     PRINT "FORWARD PARTIAL CORRELATION COEFFICIENTS:"
460     PRINT "   P          Ap(1,1)            Ap(2,1)            Ap(1,2)
Ap(2,2)"
470     FOR P=1 TO Pmax
480     PRINT USING 370;P,Ap(P,1,1),Ap(P,2,1),Ap(P,1,2),Ap(P,2,2)
490     NEXT P
500     PRINT LIN(1)
```

```
510    PRINT "BACKWARD PARTIAL CORRELATION COEFFICIENTS:"
520    PRINT "  P         Bp(1,1)            Bp(2,1)            Bp(1,2)
  Bp(2,2)"
530    FOR P=1 TO Pmax
540    PRINT USING 370;P,Bp(P,1,1),Bp(P,2,1),Bp(P,1,2),Bp(P,2,2)
550    NEXT P
560    PRINT LIN(1)
570    IF Pbest=0 THEN 890
580    CALL Pfc(Pmax,Pbest,R(*),Ap(*),Bp(*),Wa(*),Wb(*),Wc(*),Wd(*),Rn(*),R11,R22
,R12)
590    PRINT "FORWARD PREDICTIVE FILTER COEFFICIENTS FOR Pbest:"
600    PRINT "  P         Ap(1,1)            Ap(2,1)            Ap(1,2)
  Ap(2,2)"
610    FOR P=1 TO Pbest
620    PRINT USING 370;P,Ap(P,1,1),Ap(P,2,1),Ap(P,1,2),Ap(P,2,2)
630    NEXT P
640    PRINT LIN(1)
650    PRINT "NORMALIZED CORRELATION MATRICES (Rn):"
660    PRINT "DELAY      AUTO11             CROSS21            CROSS12
  AUTO22"
670    PRINT USING 370;0,R(1,1),R(2,1),R(1,2),R(2,2)
680    FOR P=1 TO Pmax
690    PRINT USING 370;P,Rn(P,1,1),Rn(P,2,1),Rn(P,1,2),Rn(P,2,2)
700    NEXT P
710    PRINT LIN(1)
720    CALL Peftf(Pbest,Nfft,Ap(*),X11(*),Y11(*),X12(*),Y12(*),X21(*),Y21(*),X22(
*),Y22(*))
730    CALL Sdm(Nfft,Ubest(*),Wa(*),Wb(*),Wc(*),Wd(*),Wa(*),X11(*),Y11(*),X12(*),
Y12(*),X21(*),Y21(*),X22(*),Y22(*),S11,S22,S12)
740    PRINT "SPECTRAL DENSITY MATRIX AND COHERENCE, FROM ZERO FREQUENCY (BIN 1):
"
750    PRINT "  BIN    AUTO11       AUTO22     RE(CROSS12) IM(CROSS12)   MAG SQ COH
  ARGUMENT"
760    IMAGE 3D,5(M.6DE,1X),M.6DE
770    FOR I=1 TO 30
780    L=I
790    IF I<16 THEN 840
800    IF I>16 THEN 830
810    PRINT "***"
820    GOTO 850
830    L=I+Nfft/2-29
840    PRINT USING 760;L,X11(L),X22(L),X12(L),Y12(L),Y11(L),Y22(L)
850    NEXT I
860    PRINT LIN(1)
870    PRINT "TRAPEZOIDAL SUMS OF SPECTRA:"
880    PRINT S11,S22,S12
890    PRINT
900    PRINT "COVARIANCES OF INPUT DATA:"
910    PRINT R11,R22,R12
920    PRINT LIN(1)
930    CALL Acm(Nfft,X11(*),X12(*),Y12(*),X21(*),Y21(*),X22(*),X11m1,X22m1,X11m0,
X22m0)
940    N1=Nfft+1
950    N2=Nfft/2
960    N22=N2+2
```

A-3

```
970    PRINT "ALIASED NORMALIZED CORRELATION MATRICES:"
980    PRINT "DELAY         AUTO11              CROSS21              CROSS12
AUTO22"
990    PRINT USING 370;0,X11(N22),X21(1),X21(1),X22(N22)
1000   FOR I=1 TO 27
1010   L=I
1020   IF I<16 THEN 1070
1030   IF I>16 THEN 1060
1040   PRINT "***"
1050   GOTO 1080
1060   L=I+N2-29
1070   PRINT USING 370;L,X11(N22+L),X21(N1-L),X21(1+L),X22(N22+L)
1080   NEXT I
1090   PRINT USING 370;N2-1,X11m1,X21(N22),X21(N2),X22m1
1100   PRINT USING 370;N2,X11m0,X21(N2+1),X21(N2+1),X22m0
1110   PRINT LIN(2)
1120   PRINTER IS 0
1130   PRINT "AUTO SPECTRAL DENSITIES IN DB:"
1140   PLOTTER IS "GRAPHICS"
1150   GRAPHICS
1160   SCALE 0,N2,-5,0
1170   GRID N2/4,1
1180   PENUP
1190   FOR I=0 TO N2
1200   PLOT I,LGT(X11(I+1))
1210   NEXT I
1220   PENUP
1230   FOR I=0 TO N2
1240   PLOT I,LGT(X22(I+1))
1250   NEXT I
1260   PENUP
1270   DUMP GRAPHICS
1280   PRINT LIN(3)
1290   PRINT "MAGNITUDE SQUARED COHERENCE AND ARGUMENT"
1300   PLOTTER IS "GRAPHICS"
1310   SCALE 0,N2,0,1
1320   GRID N2/4,.25
1330   PENUP
1340   FOR I=0 TO N2
1350   PLOT I,Y11(I+1)
1360   NEXT I
1370   PENUP
1380   SCALE 0,N2,-PI,PI
1390   FOR I=0 TO N2
1400   PLOT I,Y22(I+1)
1410   NEXT I
1420   PENUP
1430   DUMP GRAPHICS
1440   PRINT LIN(4)
1450   PRINTER IS 16
1460   END
1470   !
```

```
1480  SUB Pcc(N,Pmax,Y(*),Z(*),Ave(*),Wa(*),Wb(*),Wc(*),Wd(*),We(*),R(*),U(*),V(
*),Aic(*),Pbest,Ubest(*),Ui(*),Vi(*),A(*),B(*),Ap(*),Bp(*))
1490  Ia=INT(1.5*SQR(N))
1500  IF Pmax<=Ia THEN 1520
1510  PRINT "Pmax =";Pmax;"IS TOO LARGE FOR N =";N;" SEARCH LIMITED TO P =";Ia
1520  Ia=MIN(Ia,Pmax)
1530  Fac=3/N    ! Fac=0  WOULD FORCE Pbest EQUAL TO Pmax
1540  MAT Ave=RSUM(Y)
1550  MAT Ave=Ave/(N)
1560  A1=Ave(1)
1570  A2=Ave(2)
1580  FOR I=1 TO N
1590  Y(1,I)=Y(1,I)-A1
1600  Y(2,I)=Y(2,I)-A2
1610  NEXT I
1620  MAT Z=Y
1630  CALL Auto(2,N-1,Y(*),Wc(*))
1640  CALL Auto(1,1,Y(*),Wd(*))
1650  CALL Auto(N,N,Y(*),We(*))
1660  MAT Wa=Wc+We
1670  MAT Wb=Wc+Wd
1680  MAT R=Wb+We
1690  MAT R=R/(N)
1700  MAT U=R
1710  MAT V=R
1720  CALL Cross(2,N,Y(*),Z(*),Wc(*))
1730  Aic(0)=LOG(DET(U))
1740  Aicmin=Aic(0)
1750  Pbest=0
1760  MAT Ubest=U
1770  FOR P=1 TO Ia
1780  MAT Vi=INV(V)
1790  MAT Wd=Vi*Wb
1800  MAT Wb=Wd
1810  MAT Ui=INV(U)
1820  MAT Wd=Wa
1830  MAT Wa=Wd*Ui
1840  MAT Wc=Wc*(2)
1850  CALL Solve(Wa(*),Wb(*),Wc(*),Wd(*),We(*))
1860  MAT A=Wc*Vi
1870  MAT Wd=TRN(Wc)
1880  MAT B=Wd*Ui
1890  Ap(P,1,1)=A(1,1)
1900  Ap(P,1,2)=A(1,2)
1910  Ap(P,2,1)=A(2,1)
1920  Ap(P,2,2)=A(2,2)
1930  Bp(P,1,1)=B(1,1)
1940  Bp(P,1,2)=B(1,2)
1950  Bp(P,2,1)=B(2,1)
1960  Bp(P,2,2)=B(2,2)
1970  MAT We=A*Wd
1980  MAT U=U-We
1990  MAT We=B*Wc
2000  MAT V=V-We
```

```
2010   Aic(P)=LOG(DET(U))+Fac*P
2020   IF Aic(P)>=Aicmin THEN 2060
2030   Aicmin=Aic(P)
2040   Pbest=P
2050   MAT Ubest=U
2060   IF P=Ia THEN 2220
2070   L=P+1
2080   FOR K=N TO L STEP -1
2090   A1=Y(1,K)
2100   A2=Y(2,K)
2110   B1=Z(1,K-1)
2120   B2=Z(2,K-1)
2130   Z(1,K)=B1-B(1,1)*A1-B(1,2)*A2
2140   Z(2,K)=B2-B(2,1)*A1-B(2,2)*A2
2150   Y(1,K)=A1-A(1,1)*B1-A(1,2)*B2
2160   Y(2,K)=A2-A(2,1)*B1-A(2,2)*B2
2170   NEXT K
2180   CALL Auto(P+2,N,Y(*),Wa(*))
2190   CALL Auto(P+1,N-1,Z(*),Wb(*))
2200   CALL Cross(P+2,N,Y(*),Z(*),Wc(*))
2210   NEXT P
2220   A1=.5*(Ubest(1,2)+Ubest(2,1))
2230   Ubest(1,2)=Ubest(2,1)=A1
2240   SUBEND
2250   !

2260   SUB Pfc(Pmax,Pbest,R(*),Ap(*),Bp(*),Wa(*),Wb(*),Wc(*),Wd(*),Rn(*),R11,R22,
R12)
2270   Rn(1,1,1)=Ap(1,1,1)*R(1,1)+Ap(1,1,2)*R(2,1)
2280   Rn(1,1,2)=Ap(1,1,1)*R(1,2)+Ap(1,1,2)*R(2,2)
2290   Rn(1,2,1)=Ap(1,2,1)*R(1,1)+Ap(1,2,2)*R(2,1)
2300   Rn(1,2,2)=Ap(1,2,1)*R(1,2)+Ap(1,2,2)*R(2,2)
2310   FOR P=2 TO Pbest
2320   Wc(1,1)=Ap(P,1,1)*R(1,1)+Ap(P,1,2)*R(2,1)
2330   Wc(1,2)=Ap(P,1,1)*R(1,2)+Ap(P,1,2)*R(2,2)
2340   Wc(2,1)=Ap(P,2,1)*R(1,1)+Ap(P,2,2)*R(2,1)
2350   Wc(2,2)=Ap(P,2,1)*R(1,2)+Ap(P,2,2)*R(2,2)
2360   FOR L=1 TO P-1
2370   Ib=P-L
2380   Wa(1,1)=Ap(P,1,1)*Bp(Ib,1,1)+Ap(P,1,2)*Bp(Ib,2,1)
2390   Wa(1,2)=Ap(P,1,1)*Bp(Ib,1,2)+Ap(P,1,2)*Bp(Ib,2,2)
2400   Wa(2,1)=Ap(P,2,1)*Bp(Ib,1,1)+Ap(P,2,2)*Bp(Ib,2,1)
2410   Wa(2,2)=Ap(P,2,1)*Bp(Ib,1,2)+Ap(P,2,2)*Bp(Ib,2,2)
2420   Wa(1,1)=Ap(L,1,1)-Wa(1,1)
2430   Wa(1,2)=Ap(L,1,2)-Wa(1,2)
2440   Wa(2,1)=Ap(L,2,1)-Wa(2,1)
2450   Wa(2,2)=Ap(L,2,2)-Wa(2,2)
```

A-6

```
2460  Wb(1,1)=Bp(P,1,1)*Ap(L,1,1)+Bp(P,1,2)*Ap(L,2,1)
2470  Wb(1,2)=Bp(P,1,1)*Ap(L,1,2)+Bp(P,1,2)*Ap(L,2,2)
2480  Wb(2,1)=Bp(P,2,1)*Ap(L,1,1)+Bp(P,2,2)*Ap(L,2,1)
2490  Wb(2,2)=Bp(P,2,1)*Ap(L,1,2)+Bp(P,2,2)*Ap(L,2,2)
2500  Bp(Ib,1,1)=Bp(Ib,1,1)-Wb(1,1)
2510  Bp(Ib,1,2)=Bp(Ib,1,2)-Wb(1,2)
2520  Bp(Ib,2,1)=Bp(Ib,2,1)-Wb(2,1)
2530  Bp(Ib,2,2)=Bp(Ib,2,2)-Wb(2,2)
2540  Ap(L,1,1)=Wa(1,1)
2550  Ap(L,1,2)=Wa(1,2)
2560  Ap(L,2,1)=Wa(2,1)
2570  Ap(L,2,2)=Wa(2,2)
2580  Wd(1,1)=Wa(1,1)*Rn(Ib,1,1)+Wa(1,2)*Rn(Ib,2,1)
2590  Wd(1,2)=Wa(1,1)*Rn(Ib,1,2)+Wa(1,2)*Rn(Ib,2,2)
2600  Wd(2,1)=Wa(2,1)*Rn(Ib,1,1)+Wa(2,2)*Rn(Ib,2,1)
2610  Wd(2,2)=Wa(2,1)*Rn(Ib,1,2)+Wa(2,2)*Rn(Ib,2,2)
2620  MAT Wc=Wc+Wd
2630  NEXT L
2640  Rn(P,1,1)=Wc(1,1)
2650  Rn(P,1,2)=Wc(1,2)
2660  Rn(P,2,1)=Wc(2,1)
2670  Rn(P,2,2)=Wc(2,2)
2680  NEXT P
2690  FOR P=Pbest+1 TO Pmax
2700  MAT Wa=ZER
2710  FOR L=1 TO Pbest
2720  Ib=P-L
2730  Wb(1,1)=Ap(L,1,1)*Rn(Ib,1,1)+Ap(L,1,2)+Rn(Ib,2,1)
2740  Wb(1,2)=Ap(L,1,1)*Rn(Ib,1,2)+Ap(L,1,2)+Rn(Ib,2,2)
2750  Wb(2,1)=Ap(L,2,1)*Rn(Ib,1,1)+Ap(L,2,2)*Rn(Ib,2,1)
2760  Wb(2,2)=Ap(L,2,1)*Rn(Ib,1,2)+Ap(L,2,2)*Rn(Ib,2,2)
2770  MAT Wa=Wa+Wb
2780  NEXT L
2790  Rn(P,1,1)=Wa(1,1)
2800  Rn(P,1,2)=Wa(1,2)
2810  Rn(P,2,1)=Wa(2,1)
2820  Rn(P,2,2)=Wa(2,2)
2830  NEXT P
2840  R11=R(1,1)
2850  R22=R(2,2)
2860  R12=R(1,2)
2870  T=SQR(R11*R22)
2880  R(1,1)=R(2,2)=1
2890  R(1,2)=R(2,1)=R12/T
2900  FOR P=1 TO Pmax
2910  Rn(P,1,1)=Rn(P,1,1)/R11
2920  Rn(P,1,2)=Rn(P,1,2)/T
2930  Rn(P,2,1)=Rn(P,2,1)/T
2940  Rn(P,2,2)=Rn(P,2,2)/R22
2950  NEXT P
2960  SUBEND
2970  !
```

A-7

```
2980   SUB Peftf(Pbest,Nfft,Ap(*),X11(*),Y11(*),X12(*),Y12(*),X21(*),Y21(*),X22(*
),Y22(*))
2990   X11(1)=X22(1)=1
3000   FOR L=1 TO Pbest
3010   X11(L+1)=-Ap(L,1,1)
3020   X12(L+1)=-Ap(L,1,2)
3030   X21(L+1)=-Ap(L,2,1)
3040   X22(L+1)=-Ap(L,2,2)
3050   NEXT L
3060   CALL Fft10(Nfft,X11(*),Y11(*))
3070   CALL Fft10(Nfft,X12(*),Y12(*))
3080   CALL Fft10(Nfft,X21(*),Y21(*))
3090   CALL Fft10(Nfft,X22(*),Y22(*))
3100   SUBEND
3110   !

3120   SUB Sdm(Nfft,Ubest(*),Wa(*),Wb(*),Wc(*),Wd(*),We(*),X11(*),Y11(*),X12(*),Y
12(*),X21(*),Y21(*),X22(*),Y22(*),S11,S22,S12)
3130   T=2/Nfft
3140   S11=S22=S12=0
3150   J=Nfft/2+1
3160   FOR L=1 TO J
3170   Wa(1,1)=X22(L)
3180   Wa(1,2)=-X12(L)
3190   Wa(2,1)=-X21(L)
3200   Wa(2,2)=X11(L)
3210   Wb(1,1)=Y22(L)
3220   Wb(1,2)=-Y12(L)
3230   Wb(2,1)=-Y21(L)
3240   Wb(2,2)=Y11(L)
3250   Ta=DET(Wa)-DET(Wb)
3260   Tb=Wa(1,1)*Wb(2,2)+Wa(2,2)*Wb(1,1)-Wa(1,2)*Wb(2,1)-Wa(2,1)*Wb(1,2)
3270   Ta=T/(Ta*Ta+Tb*Tb)
3280   MAT Wc=TRN(Wa)
3290   MAT Wd=Ubest*Wc
3300   MAT Wc=Wb+Wd
3310   Tb=Wc(1,2)-Wc(2,1)
3320   MAT Wc=Wa+Wd
3330   MAT Wd=TRN(Wb)
3340   MAT We=Ubest*Wd
3350   MAT Wd=Wb+We
3360   MAT Wc=Wc+Wd
3370   Y11(L)=(Wc(1,2)^2+Tb*Tb)/(Wc(1,1)*Wc(2,2))
3380   Y22(L)=FNArg(Wc(1,2),Tb)
3390   X11(L)=Ta*Wc(1,1)
3400   X22(L)=Ta*Wc(2,2)
3410   X12(L)=Ta*Wc(1,2)
3420   Y12(L)=Ta*Tb
3430   S11=S11+X11(L)
3440   S22=S22+X22(L)
3450   S12=S12+X12(L)
3460   NEXT L
3470   S11=S11-.5*(X11(1)+X11(J))
3480   S22=S22-.5*(X22(1)+X22(J))
3490   S12=S12-.5*(X12(1)+X12(J))
3500   SUBEND
3510   !
```

```
3520    SUB Acm(Nfft,X11(*),X12(*),Y12(*),X21(*),Y21(*),X22(*),X11m1,X22m1,X11m0,X
22m0)
3530    N2=Nfft/2
3540    N21=N2+1
3550    N22=N2+2
3560    FOR L=1 TO N2
3570    X21(L)=.5*X11(L)
3580    Y21(L)=.5*X22(L)
3590    X21(N2+L)=.5*X11(N22-L)
3600    Y21(N2+L)=.5*X22(N22-L)
3610    NEXT L
3620    CALL Fft10(Nfft,X21(*),Y21(*))
3630    Ta=1/X21(1)
3640    Tb=1/Y21(1)
3650    T=SQR(Ta*Tb)
3660    X11(N22)=X22(N22)=1
3670    FOR L=2 TO N2-1
3680    X11(N21+L)=X21(L)*Ta
3690    X22(N21+L)=Y21(L)*Tb
3700    NEXT L
3710    X11m1=X21(N2)*Ta
3720    X22m1=Y21(N2)*Tb
3730    X11m0=X21(N21)*Ta
3740    X22m0=Y21(N21)*Tb
3750    X21(1)=.5*X12(1)*T
3760    Y21(1)=-.5*Y12(1)*T
3770    FOR L=2 TO N2
3780    X21(L)=X12(L)*T
3790    Y21(L)=-Y12(L)*T
3800    X21(N2+L)=Y21(N2+L)=0
3810    NEXT L
3820    X21(N21)=.5*X12(N21)*T
3830    Y21(N21)=-.5*Y12(N21)*T
3840    CALL Fft10(Nfft,X21(*),Y21(*))
3850    SUBEND
3860    !
```

```
3870   SUB Cross(N1,N2,A(*),B(*),C(*))
3880   S11=S12=S21=S22=0
3890   FOR K=N1 TO N2
3900   A1=A(1,K)
3910   A2=A(2,K)
3920   B1=B(1,K-1)
3930   B2=B(2,K-1)
3940   S11=S11+A1*B1
3950   S12=S12+A1*B2
3960   S21=S21+A2*B1
3970   S22=S22+A2*B2
3980   NEXT K
3990   C(1,1)=S11
4000   C(1,2)=S12
4010   C(2,1)=S21
4020   C(2,2)=S22
4030   SUBEND
4040   !

4050   SUB Auto(N1,N2,A(*),B(*))
4060   S11=S12=S22=0
4070   FOR K=N1 TO N2
4080   A1=A(1,K)
4090   A2=A(2,K)
4100   S11=S11+A1*A1
4110   S12=S12+A1*A2
4120   S22=S22+A2*A2
4130   NEXT K
4140   B(1,1)=S11
4150   B(1,2)=B(2,1)=S12
4160   B(2,2)=S22
4170   SUBEND
4180   !

4190   SUB Solve(Wa(*),Wb(*),Wc(*),Wd(*),We(*))
4200   Ta=Wa(1,1)+Wa(2,2)+Wb(1,1)+Wb(2,2)
4210   Tb=DET(Wa)-DET(Wb)
4220   MAT Wd=Wc*Wb
4230   We(1,1)=Wa(2,2)
4240   We(1,2)=-Wa(1,2)
4250   We(2,1)=-Wa(2,1)
4260   We(2,2)=Wa(1,1)
4270   MAT Wa=We*Wc
4280   MAT Wd=Wd+Wa
4290   MAT Wb=Wb*(Ta)
4300   Wb(1,1)=Wb(1,1)+Tb
4310   Wb(2,2)=Wb(2,2)+Tb
4320   MAT We=INV(Wb)
4330   MAT Wc=Wd*We
4340   SUBEND
4350   !
```

A-10

```
4360  SUB Fft10(N,X(*),Y(*))      ! N <= 2^10 = 1024, N=2 INTEGER
4370  DIM C(1:257)
4380  DATA 1,.999981175283,.999924701839,.999830581796,.999698818696,.9995294175
01,.999322384588,.999077727753,.998795456205,.998475580573,.998118112900
4390  DATA .997723066644,.997290456679,.996820299291,.996312612183,.995767414468
,.995184726672,.994564570734,.993906970002,.993211949235,.992479534599
4400  DATA .991709753669,.990902635428,.990058210262,.989176509965,.988257567731
,.987301418158,.986308097245,.985277642389,.984210092387,.983105487431
4410  DATA .981963869110,.980785280403,.979569765685,.978317370720,.977028142658
,.975702130039,.974339382786,.972939952206,.971503890986,.970031253195
4420  DATA .968522094274,.966976471045,.965394441698,.963776065795,.962121404269
,.960430519416,.958703474896,.956940335732,.955141168306,.953306040354
4430  DATA .951435020969,.949528180593,.947585591018,.945607325381,.943593458162
,.941544065183,.939459223602,.937339011913,.935183509939,.932992798835
4440  DATA .930766961079,.928506080473,.926210242138,.923879532511,.921514039342
,.919113851690,.916679059921,.914209755704,.911706032005,.909167983091
4450  DATA .906595704515,.903989293123,.901348847046,.898674465694,.895966249756
,.893224301196,.890448723245,.887639620403,.884797098431,.881921264348
4460  DATA .879012226429,.876070094195,.873094978418,.870086991109,.867046245516
,.863972856122,.860866938638,.857728610000,.854557988365,.851355193105
4470  DATA .848120344803,.844853565250,.841554977437,.838224705555,.834862874986
,.831469612303,.828045045258,.824589302785,.821102514991,.817584813152
4480  DATA .814036329706,.810457198253,.806847553544,.803207531481,.799537269108
,.795836904609,.792106577300,.788346427627,.784556597156,.780737228572
4490  DATA .776888465673,.773010453363,.769103337646,.765167265622,.761202385484
,.757208846506,.753186799044,.749136394523,.745057785441,.740951125355
4500  DATA .736816563877,.732654271672,.728464390448,.724247082951,.720002507961
,.715730825284,.711432195745,.707106781187,.702754744457,.698376249409
4510  DATA .693971460890,.689540544737,.685083667773,.680600997795,.676092703575
,.671558954847,.666999922304,.662415777590,.657806693297,.653172842954
4520  DATA .648514401022,.643831542890,.639124444864,.634393284164,.629638238915
,.624859488142,.620057211763,.615231590581,.610382806276,.605511041404
4530  DATA .600616479384,.595699304492,.590759701859,.585797857456,.580813958096
,.575808191418,.570780745887,.565731810784,.560661576197,.555570233020
4540  DATA .550457972937,.545324988422,.540171472730,.534997619887,.529803624686
,.524589682678,.519355990166,.514102744193,.508830142543,.503538383726
4550  DATA .498227666973,.492898192230,.487550160148,.482183772079,.476799230063
,.471396736826,.465976495768,.460538710958,.455083587126,.449611329655
4560  DATA .444122144570,.438616238539,.433093818853,.427555093430,.422000270800
,.416429560098,.410843171058,.405241314005,.399624199846,.393992040061
4570  DATA .388345046699,.382683432365,.377007410216,.371317193952,.365612997805
,.359895036535,.354163525420,.348418680249,.342660717312,.336889853392
4580  DATA .331106305760,.325310292162,.319502030816,.313681740399,.307849640042
,.302005949319,.296150888244,.290284677254,.284407537211,.278519689385
4590  DATA .272621355450,.266712757475,.260794117915,.254865659605,.248927605746
,.242980179903,.237023605994,.231058108281,.225083911360,.219101240157
4600  DATA .213110319916,.207111376192,.201104634842,.195090322016,.189068664150
,.183039887955,.177004220412,.170961388760,.164913120490,.158858143334
4610  DATA .152797185258,.146730474455,.140658239333,.134580708507,.128498110794
,.122410675199,.116318630912,.110222207294,.104121633872,.980171403296E-1
4620  DATA .919089564971E-1,.857973123444E-1,.796824379714E-1,.735645635997E-1,.
674439195637E-1,.613207363022E-1,.551952443497E-1,.490676743274E-1
4630  DATA .429382569349E-1,.368072229414E-1,.306748031766E-1,.245412285229E-1,.
184067299058E-1,.122715382857E-1,.613588464915E-2,0
```

```
4640   READ C(*)
4650   M=1024/N
4660   FOR I=0 TO N/4
4670   C(I+1)=C(M*I+1)
4680   NEXT I
4690   N1=N/4
4700   N2=N1+1
4710   N3=N2+1
4720   N4=N1+N3
4730   Log2n=INT(1.4427*LOG(N)+.5)
4740   FOR I1=1 TO Log2n
4750   I2=2^(Log2n-I1)
4760   I3=2*I2
4770   I4=N/I3
4780   FOR I5=1 TO I2
4790   I6=(I5-1)*I4+1
4800   IF I6<=N2 THEN 4840
4810   N6=-C(N4-I6)
4820   N7=-C(I6-N1)
4830   GOTO 4860
4840   N6=C(I6)
4850   N7=-C(N3-I6)
4860   FOR I7=0 TO N-I3 STEP I3
4870   I8=I7+I5
4880   I9=I8+I2
4890   N8=X(I8)-X(I9)
4900   N9=Y(I8)-Y(I9)
4910   X(I8)=X(I8)+X(I9)
4920   Y(I8)=Y(I8)+Y(I9)
4930   X(I9)=N6*N8-N7*N9
4940   Y(I9)=N6*N9+N7*N8
4950   NEXT I7
4960   NEXT I5
4970   NEXT I1
4980   I1=Log2n+1
4990   FOR I2=1 TO 10                    ! 2^10=1024
5000   C(I2)=1
5010   IF I2>Log2n THEN 5030
5020   C(I2)=2^(I1-I2)
5030   NEXT I2
```

```
5040    K=1
5050    FOR I1=1 TO C(10)
5060    FOR I2=I1 TO C(9) STEP C(10)
5070    FOR I3=I2 TO C(8) STEP C(9)
5080    FOR I4=I3 TO C(7) STEP C(8)
5090    FOR I5=I4 TO C(6) STEP C(7)
5100    FOR I6=I5 TO C(5) STEP C(6)
5110    FOR I7=I6 TO C(4) STEP C(5)
5120    FOR I8=I7 TO C(3) STEP C(4)
5130    FOR I9=I8 TO C(2) STEP C(3)
5140    FOR I10=I9 TO C(1) STEP C(2)
5150    J=I10
5160    IF K>J THEN 5230
5170    A=X(K)
5180    X(K)=X(J)
5190    X(J)=A
5200    A=Y(K)
5210    Y(K)=Y(J)
5220    Y(J)=A
5230    K=K+1
5240    NEXT I10
5250    NEXT I9
5260    NEXT I8
5270    .NEXT I7
5280    NEXT I6
5290    NEXT I5
5300    NEXT I4
5310    NEXT I3
5320    NEXT I2
5330    NEXT I1
5340    SUBEND
5350    !

5360    DEF FNArg(X,Y)              ! PRINCIPAL ARG(Z)
5370    IF X=0 THEN A=.5*PI*SGN(Y)
5380    IF X<>0 THEN A=ATN(Y/X)
5390    IF X<0 THEN A=A+PI*(1-2*(Y<0))
5400    RETURN A
5410    FNEND
5420    !

5430    SUB Data(N,X(*))
5440    OPTION BASE 1
5450    REDIM X(2,N)
5460    DATA 1,2,6,3,1,1,2,1,4,5,3,2,1,5,6,1,2,4,8,9
5470    DATA 2,1,0,1,5,3,0,1,2,6,2,2,4,2,3,5.6,9,8,2
5480    READ X(*)
5490    SUBEND
```

```
NUMBER OF DATA POINTS IN EACH PROCESS N = 20
MAXIMUM ORDER OF PREDICTIVE FILTER Pmax = 6
SIZE OF FFT Nfft = 256
```

PROCESS NUMBER 1
```
 1  2  6  3  1  1  2  1  4  5  3  2  1  5  6  1  2  4  8  9
```

PROCESS NUMBER 2
```
 2  1  0  1  5  3  0  1  2  6  2  2  4  2  3  5  6  9  8  2
```

MEANS OF INPUT DATA (Ave):
3.35
3.2

COVARIANCE MATRIX OF INPUT DATA (R):
```
 5.7275              .83

 .83                 6.16
```

AKAIKE INFORMATION CRITERION:

P	Aic(P)
0	.354363690E+01
1	.324147848E+01
2	.313209545E+01
3	.342425965E+01
4	.235777553E+01
5	.248589498E+01
6	.266117185E+01

Pbest = 4

Ubest:
```
 .54593344383         .226987553923

 .226987553923        4.00240808163
```

FORWARD PARTIAL CORRELATION COEFFICIENTS:

P	Ap(1,1)	Ap(2,1)	Ap(1,2)	Ap(2,2)
1	.326847861E+00	-.296055023E-01	.433508958E+00	.492296101E+00
2	-.497079412E+00	.851055338E-01	.177974150E+00	-.275017512E+00
3	-.172083146E+00	.237335358E+00	.586121047E-01	.154803723E+00
4	-.143631688E+00	-.110686655E+00	.734743629E+00	-.124715363E+00
5	-.515644665E-01	-.440540782E+00	.193329089E+00	.883560194E+00
6	.901306062E-01	-.736616304E-01	-.256580155E+00	.734752737E+00

BACKWARD PARTIAL CORRELATION COEFFICIENTS:

P	Bp(1,1)	Bp(2,1)	Bp(1,2)	Bp(2,2)
1	.391694481E+00	.451665794E+00	-.139717314E-01	.427449481E+00
2	-.657848877E+00	.178422853E+00	.271402210E-01	-.192143512E+00
3	-.218767381E+00	.685350177E-01	.177465355E+00	.129759954E+00
4	-.128639908E+00	.100617477E+01	-.365594759E-01	-.944141856E-01
5	-.257137623E-01	.194124206E+00	-.219204308E+00	.117920510E+00
6	.182731463E+00	-.284463163E+00	.626220731E-01	.161491679E+00

FORWARD PREDICTIVE FILTER COEFFICIENTS FOR Ptest:

P	Ap(1,1)	Ap(2,1)	Ap(1,2)	Ap(2,2)
1	.235726249E+00	.174118149E+00	.296570575E+00	.670170572E+00
2	-.600483360E+00	-.164936772E+00	.412699673E+00	-.442413930E+00
3	-.404360932E+00	.342089409E+00	-.235743642E+00	.251445573E+00
4	-.143631688E+00	-.110686655E+00	.734743629E+00	-.124715363E+00

NORMALIZED CORRELATION MATRICES (Rm):

DELAY	AUTO11	CROSS21	CROSS12	AUTO22
0	.100000000E+01	.139734996E+00	.139734996E+00	.100000000E+01
1	.369669762E+00	.402437203E-01	.495250946E+00	.488307047E+00
2	-.284751305E+00	.912480910E-01	.511090660E+00	.640872368E-01
3	-.372945922E+00	.195208145E+00	.117107113E+00	.663036688E-01
4	-.223941701E-01	.876238663E-01	.190792716E+00	.100976458E+00
5	.396697360E+00	-.896791098E-01	.112701803E+00	.122691975E+00
6	.330388400E+00	-.836262522E-01	-.951869127E-01	.188886781E-01

SPECTRAL DENSITY MATRIX AND COHERENCE, FROM ZERO FREQUENCY (BIN 1):

BIN	AUTO11	AUTO22	RE(CROSS12)	IM(CROSS12)	MAG SQ COH	ARGUMENT
1	.563234E-01	.130480E+00	.848704E-01	.000000E+01	.980117E+00	.000000E+01
2	.562462E-01	.130295E+00	.847113E-01	-.254607E-02	.980064E+00	-.300468E-01
3	.560152E-01	.129740E+00	.842359E-01	-.506959E-02	.979904E+00	-.601108E-01
4	.556327E-01	.128822E+00	.834493E-01	-.754840E-02	.979634E+00	-.902094E-01
5	.551025E-01	.127549E+00	.823603E-01	-.996105E-02	.979250E+00	-.120360E+00
6	.544301E-01	.125934E+00	.809811E-01	-.122873E-01	.978746E+00	-.150581E+00
7	.536226E-01	.123994E+00	.793277E-01	-.145083E-01	.978113E+00	-.180892E+00
8	.526890E-01	.121751E+00	.774195E-01	-.166075E-01	.977342E+00	-.211311E+00
9	.516400E-01	.119230E+00	.752793E-01	-.185704E-01	.976419E+00	-.241858E+00
10	.504882E-01	.116460E+00	.729330E-01	-.203857E-01	.975331E+00	-.272557E+00
11	.492476E-01	.113474E+00	.704068E-01	-.220448E-01	.974059E+00	-.303429E+00
12	.479335E-01	.110309E+00	.677369E-01	-.235426E-01	.972584E+00	-.334499E+00
13	.465623E-01	.107002E+00	.649485E-01	-.248774E-01	.970833E+00	-.365795E+00
14	.451507E-01	.103593E+00	.620751E-01	-.260508E-01	.968928E+00	-.397345E+00
15	.437158E-01	.100121E+00	.591473E-01	-.270676E-01	.966688E+00	-.429181E+00

***

BIN	AUTO11	AUTO22	RE(CROSS12)	IM(CROSS12)	MAG SQ COH	ARGUMENT
116	.495654E-02	.527756E-02	.180088E-02	.366875E-02	.638529E+00	.111448E+01
117	.464126E-02	.492382E-02	.190282E-02	.323458E-02	.616259E+00	.103904E+01
118	.436922E-02	.461992E-02	.198172E-02	.284179E-02	.594634E+00	.961851E+00
119	.413522E-02	.435955E-02	.204244E-02	.248491E-02	.573915E+00	.882822E+00
120	.393482E-02	.413733E-02	.208889E-02	.215898E-02	.554351E+00	.801897E+00
121	.376422E-02	.394877E-02	.212416E-02	.185948E-02	.536176E+00	.719053E+00
122	.362025E-02	.379007E-02	.215073E-02	.158236E-02	.519605E+00	.634308E+00
123	.350027E-02	.365814E-02	.217053E-02	.132398E-02	.504835E+00	.547727E+00
124	.340211E-02	.355042E-02	.218509E-02	.108104E-02	.492036E+00	.459427E+00
125	.332403E-02	.346487E-02	.219558E-02	.850527E-03	.481358E+00	.369582E+00
126	.326465E-02	.339991E-02	.220289E-02	.629579E-03	.472923E+00	.278418E+00
127	.322295E-02	.335433E-02	.220767E-02	.415919E-03	.466828E+00	.196214E+00
128	.319821E-02	.332731E-02	.221037E-02	.296809E-03	.463144E+00	.932912E-01
129	.319001E-02	.331836E-02	.221125E-02	.000000E+01	.461911E+00	.000000E+01

TRAPEZOIDAL SUMS OF SPECTRA:
5.72678260464   6.15993376582   .830232954125

COVARIANCES OF INPUT DATA:
5.7275   6.16   .83

ALIASED NORMALIZED CORRELATION MATRICES:

DELAY	AUTO11	CROSS21	CROSS12	AUTO22
0	.100000000E+01	.139783721E+00	.139783721E+00	.100000000E+01
1	.389662823E+00	.402483863E-01	.495316301E+00	.488307810E+00
2	-.284711075E+00	.912157098E-01	.511114181E+00	.640949626E-01
3	-.372868877E+00	.195185513E+00	.117074400E+00	.663147971E-01
4	-.223628911E-01	.876380084E-01	.190781841E+00	.100979179E+00
5	.396652516E+00	-.896408926E-01	.112729049E+00	.122684113E+00
6	.330310631E+00	-.836012969E-01	-.951530121E-01	.188794710E-01
7	-.345026269E-01	.609690818E-02	-.999179326E-01	-.796893320E-02
8	-.308860295E+00	.824851230E-01	.131798148E-01	.197442248E-01
9	-.262499156E+00	.629998356E-01	.176921946E+00	-.191113963E-01
10	.795873936E-01	-.238734308E-01	.105767456E+00	-.211019705E-01
11	.310325251E+00	-.669643776E-01	-.976171228E-01	.506845393E-02
12	.192490120E+00	-.429532422E-01	-.148308163E+00	.290497257E-01
13	-.123481195E+00	.241857936E-01	-.432258845E-01	.212691423E-01
14	-.294362011E+00	.638641478E-01	.104657982E+00	-.219543907E-01
15	-.128164922E+00	.302603975E-01	.123725501E+00	-.315105589E-01
***				
116	-.244194112E-02	.913297510E-03	.313869734E-02	-.782794585E-03
117	.270370651E-02	.703065771E-03	.135635926E-02	-.620489604E-03
118	.411244195E-02	-.122243230E-03	-.167993100E-02	.177358396E-03
119	.865212010E-03	-.844466456E-03	-.265889725E-02	.750518242E-03
120	-.280736944E-02	-.822305920E-03	-.764920706E-03	.519056852E-03
121	-.264788177E-02	-.179269851E-04	.172870237E-02	-.239248236E-03
122	.440106016E-03	.889991252E-03	.216969761E-02	-.719997787E-03
123	.238535013E-02	.101284639E-02	.329164628E-03	-.439526451E-03
124	.112686107E-02	.992673477E-04	-.164236260E-02	.286993480E-03
125	-.121802675E-02	-.105565279E-02	-.171669788E-02	.698618350E-03
126	-.151757804E-02	-.124026530E-02	-.491074185E-04	.378592478E-03
127	.306556500E-03	-.866410300E-04	.146460666E-02	-.330374132E-03
128	.148247837E-02	.133650385E-02	.133650385E-02	-.690972582E-03
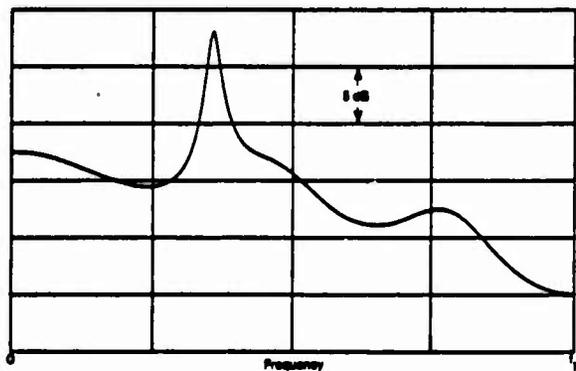
A-1A. Auto Spectrum of First Process
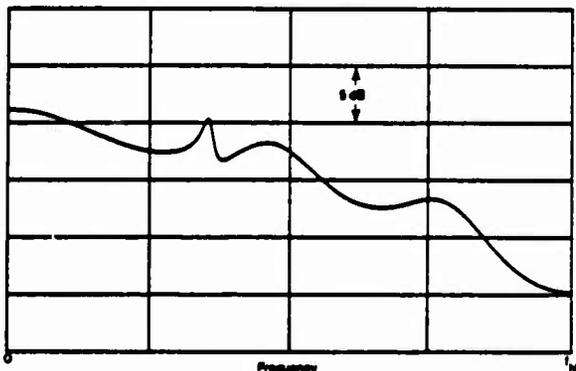
A-1B. Auto Spectrum of Second Process
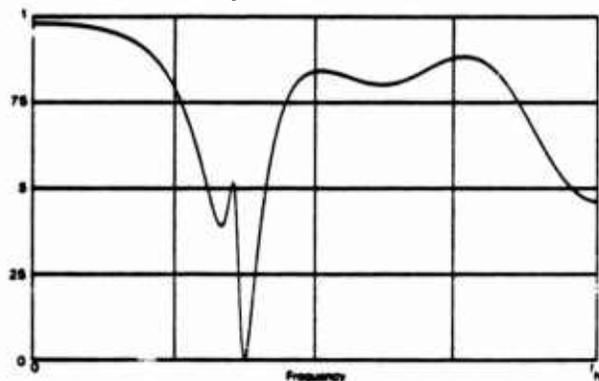
A-1C. Magnitude-Squared Coherence
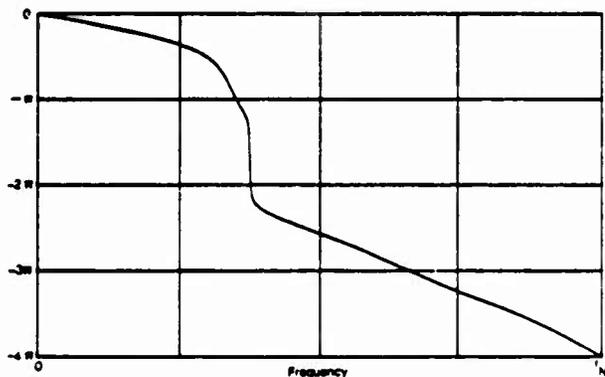
A-1D. Argument of Complex Coherence

Figure A-1. Spectral Estimates for N = 20, Pmax = 6, Pbest = 4

Appendix B

General Filter and Spectral Relations

For each integer k, let H_k be a rectangular matrix of (complex) constants, to be called the filter impulse response at delay $k\Delta$, where Δ is the time sampling increment. The number of filter outputs need not equal the number of filter inputs. For multichannel filter excitation W_n at time $n\Delta$, the filter output at time $n\Delta$ is given by the discrete convolution

$$X_n = \sum_k H_k W_{n-k} \quad , \tag{B-1}$$

where the summation extends over all nonzero summands.

For a stationary excitation, let the correlation matrix of complex input process $\{W_n\}$ at delay $\ell\Delta$ be

$$\overline{W_n W_{n-\ell}^H} = P_\ell \quad \text{(non-diagonal matrix)}, \tag{B-2}$$

where the overbar denotes an ensemble average, and the superscript H denotes a conjugate transpose. The z-transform of input correlation sequence $\{P_\ell\}$ is

$$Q(z) \equiv \Delta \sum_\ell z^{-\ell} P_\ell \quad , \tag{B-3}$$

and the spectrum of input process $\{W_n\}$ is, for real frequency f,

$$Q(f) = Q(\exp(i2\pi f\Delta)) = \Delta \sum_\ell \exp(-i2\pi f\Delta\ell) P_\ell$$
$$= \Delta \sum_\ell \exp(-iu\ell) P_\ell \quad , \tag{B-4}$$

where we let

$$u \equiv 2\pi f\Delta \quad . \tag{B-5}$$

The correlation matrix of the filter output process $\{X_n\}$, at delay $\ell\Delta$, is given by using (B-1) and (B-2):

$$R_\ell \equiv \overline{X_n X_{n-\ell}^H} = \sum_{k,m} H_k \overline{W_{n-k} W_{n-\ell-m}^H} H_m^H = \sum_{k,m} H_k P_{\ell+m-k} H_m^H \quad . \tag{B-6}$$

The z-transform of output correlation sequence $\{R_\ell\}$ is, by use of (B-6),

$$G(z) \equiv \Delta \sum_{\ell} z^{-\ell} R_{\ell} = \Delta \sum_{\ell} z^{-\ell} \sum_{k,m} H_k P_{\ell+m-k} H_m^H$$

$$= \sum_{k} z^{-k} H_k \Delta \sum_{\ell} z^{-(\ell+m-k)} P_{\ell+m-k} \sum_{m} z^m H_m^H \quad . \quad \text{(B-7)}$$

Now define the z-transform of filter sequence $\{H_k\}$ by

$$H(z) = \sum_{k} z^{-k} H_k \tag{B-8}$$

and define the quantity

$$H^H(z) \equiv [H(z)]^H \quad , \quad \text{including conjugation of } z. \tag{B-9}$$

Then

$$H^H\left(\frac{1}{z^*}\right) = \left[H\left(\frac{1}{z^*}\right)\right]^H = \left[\sum_{k} z^{*k} H_k\right]^H = \sum_{k} z^k H_k^H \quad . \tag{B-10}$$

We now employ (B-8), (B-3), and (B-10) in (B-7) to obtain

$$G(z) = H(z) \ Q(z) \ H^H\left(\frac{1}{z^*}\right) \quad . \tag{B-11}$$

The spectrum of output process $\{X_n\}$ is then, at real frequency f,

$$G(f) = G(\exp(i2\pi f\Delta)) = \Delta \sum_{\ell} \exp(-i2\pi f\Delta\ell) \ R_{\ell}$$

$$= H(f) \ Q(f) \ H^H(f) \quad , \tag{B-12}$$

where we employed (B-11), (B-4), and set

$$H(f) = H(\exp(i2\pi f\Delta)) = \sum_{k} \exp(-i2\pi f\Delta k) \ H_k \quad . \tag{B-13}$$

We also employed the property that

$$H^H(f) = [H(f)]^H = \sum_{k} \exp(i2\pi f\Delta k) \ H_k^H \quad , \tag{B-14}$$

or

$$H^H\left(\frac{1}{\exp(-i2\pi f\Delta)}\right) = H^H(\exp(i2\pi f\Delta)) = H^H(f) \quad . \qquad \text{(B-15)}$$

Finally, from convolution (B-1), we obtain the z-transform of output data sequence $\{X_n\}$ as

$$\chi(z) \equiv \sum_n z^{-n} X_n = H(z) \, W(z) \quad , \qquad \text{(B-16)}$$

where we used (B-8) and defined

$$W(z) = \sum_n z^{-n} W_n \quad . \qquad \text{(B-17)}$$

The major results thus far are given by (B-1), (B-16), (B-11), and (B-12) for a general filter and excitation.

ARMA Process

For a multichannel, autoregressive, moving-average process, the recursion is given by

$$X_n = \sum_k E_k \, X_{n-k} + \sum_k F_k \, W_{n-k} \quad . \qquad \text{(B-18)}$$

The z-transform of this equation is

$$X(z) = E(z) \, X(z) + F(z) \, W(z) \quad , \qquad \text{(B-19)}$$

where

$$E(z) \equiv \sum_k z^{-k} E_k \quad , \quad F(z) \equiv \sum_k z^{-k} F_k \quad . \qquad \text{(B-20)}$$

Then we can solve (B-19) for $X(z)$ as

$$\chi(z) = [I - E(z)]^{-1} F(z) \, W(z) = H(z) \, W(z) \quad , \qquad \text{(B-21)}$$

where the transfer function from input to output is

$$H(z) \equiv [I - E(z)]^{-1} F(z) \qquad \text{(B-22)}$$

in terms of the parameters of recursion (B-18). But now (B-22) is in the framework of the presentation above; namely, the spectrum of output process $\{X_n\}$ is, from (B-12),

$$G(f) = H(f) \, Q(f) \, H^H(f) \quad , \qquad \text{(B-23)}$$

where

$$H(f) = [I - E(f)]^{-1} F(f) \quad ,$$

$$E(f) = E(\exp(i2\pi f\Delta)) = \sum_k \exp(-i2\pi f\Delta k) \, E_k \quad ,$$

$$F(f) = F(\exp(i2\pi f\Delta)) = \sum_k \exp(-i2\pi f\Delta k) \, F_k \quad . \tag{B-24}$$

Example

As an example of (B-18), consider the multichannel first-order recursion

$$X_n = E_1 \, X_{n-1} + W_n \tag{B-25}$$

with the input spectrum for $\{W_n\}$,

$$Q(f) = \Delta \, I \quad \text{for all} \quad f \quad . \tag{B-26}$$

This is a white process, uncorrelated from channel to channel. Then

$$E(z) = z^{-1} \, E_1 \quad , \quad F(z) = I \quad ,$$

$$H(z) = (I - z^{-1} \, E_1)^{-1} \quad , \tag{B-27}$$

$$G(f) = \Delta \, H(f) \, H^H(f) \quad .$$

Specialization to Two-Channel Process

We further specialize example (B-25) to the two-input, two-output channel process characterized by coefficient matrix

$$E_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{(complex coefficients)}. \tag{B-28}$$

Then (B-27) and (B-22) yield transfer function

$$H(z) = \begin{bmatrix} 1 - z^{-1} a & -z^{-1} b \\ -z^{-1} c & 1 - z^{-1} d \end{bmatrix}^{-1}$$

$$= \frac{1}{1 - z^{-1}(a + d) + z^{-2}(ad - bc)} \begin{bmatrix} 1 - z^{-1} d & z^{-1} b \\ z^{-1} c & 1 - z^{-1} a \end{bmatrix} \quad \text{(B-29)}$$

There follows

$$H^H(1/z^*) = \left[H(1/z^*) \right]^H$$

$$= \frac{1}{1 - z(a + d)^* + z^2(ad - bc)^*} \begin{bmatrix} 1 - zd^* & zc^* \\ zb^* & 1 - za^* \end{bmatrix} \quad \text{(B-30)}$$

and

$$G(z) = \Delta \, H(z) \, H^H(1/z^*) \equiv \begin{bmatrix} g_{11}(z) & g_{12}(z) \\ g_{21}(z) & g_{22}(z) \end{bmatrix}$$

$$= \frac{1}{D} \begin{bmatrix} 1 + |b|^2 + |d|^2 - z^{-1} d - zd^* & -a^* b - c^* d + z^{-1} b + zc^* \\ -ab^* - cd^* + z^{-1} c + zb^* & 1 + |a|^2 + |c|^2 - z^{-1} a - za^* \end{bmatrix} ,$$

$$\text{(B-31)}$$

where

$$D = \left[1 - z^{-1}(a + d) + z^{-2}(ad - bc) \right] \left[1 - z(a + d)^* + z^2(ad - bc)^* \right] .$$

$$\text{(B-32)}$$

Inspection of denominator D in (B-32) reveals that $G(z)$ has poles (i.e., all the elements of matrix $G(z)$ have poles) at

$$z_\infty = \frac{a + d \pm \sqrt{(a + d)^2 - 4(ad - bc)}}{2}$$

(B-33)

and at

$$\frac{1}{z_\infty^*} = \frac{1}{|z_\infty|} \exp(i \arg(z_\infty)) \ .$$

(B-34)

That is, even though recursion (B-25) is limited to first-order regressive and white independent excitations (see (B-26)), $G(z)$ has *four* poles in the finite z-plane. Generally, if $H(z)$ has a pole at z_∞, then $H(1/z^*)$ has a pole at $1/z_\infty^*$; so if

$$z_\infty = r \exp(i\theta), \qquad \text{then} \qquad 1/z_\infty^* = \exp(i\theta)/r$$

(B-35)

has the same argument.

In addition, element $g_{11}(z)$ in (B-31) has a zero at ∞ and three zeros in the finite z-plane, at

$$z_0 = 0 \quad \text{and} \quad z_0 = \frac{1 + |b|^2 + |d|^2 \pm \sqrt{\left(1 + |b|^2 + |d|^2\right)^2 - 4|d|^2}}{2d^*}$$

(B-36)

The product of the latter two zero locations is $d/d^* = \exp(i2\arg(d))$. Thus, the auto spectrum of process 1 has three zeros and four poles, even though the multichannel recursion, (B-25), is first-order regressive. Similar behavior is true of the auto spectrum of the second process, as well as the cross spectrum between the two processes.

The magnitude-squared coherence for this example is, from (B-31) and (B-5),

$$\frac{g_{12}(\exp(iu)) \ g_{21}(\exp(iu))}{g_{11}(\exp(iu)) \ g_{22}(\exp(iu))}$$

$$= \frac{\left| - a^*b - c^*d + b \exp(-iu) + c^* \exp(iu)\right|^2}{\left[\left|1 - d \exp(-iu)\right|^2 + |b|^2\right]\left[\left|1 - a \exp(-iu)\right|^2 + |c|^2\right]} \ .$$

(B-37)

This has four zeros and four poles in the finite z-plane.

Numerical Example

We now specialize (B-28) to example (6) in the main text:

$$E_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} .85 & -.75 \\ .65 & .55 \end{bmatrix} \quad . \tag{B-38}$$

The four poles of the spectrum $G(z)$ are located at

$$z_\infty = .7 \pm i \; .6819 \; , \qquad 1/z_\infty^* = .7330 \pm i \; .7140 \; , \tag{B-39}$$

and the zeros of $g_{11}(z)$ are at

$$z_0 = 0 \; , \quad 3.0646, \quad .3263, \quad \infty \quad . \tag{B-40}$$

The zeros of $g_{12}(z)$ are located instead at

$$z_0 = 0 \; , \quad .8802, \quad -1.3109, \quad \infty \quad . \tag{B-41}$$

The magnitude-squared coherence simplifies to

$$\frac{1.0634 - .056 \cos(u) - .975 \cos(2u)}{[1.865 - 1.1 \cos(u)] \; [2.145 - 1.7 \cos(u)]} \quad . \tag{B-42}$$

This example was used frequently in the main body of this report. The peak value is .9990128 at $u = .772564$, or $2f\Delta = .245915$.

References

1. J.P. Burg, "A New Analysis Technique for Time Series Data," *NATO Advanced Study Institute on Signal Processing*, Enschede, Netherlands, vol. 1, August 1968.

2. R.T. Lacoss, "Data-Adaptive Spectral Analysis Methods," *Geophysics*, vol. 36, no. 4, August 1971, pp. 661-675.

3. J. Makhoul, "Linear Prediction: A Tutorial Review," *Proceedings of the IEEE*, vol. 63, no. 4, April 1975, pp. 561-580.

4. A.H. Nuttall, "Spectral Analysis of a Univariate Process with Bad Data Points, via Maximum Entropy and Linear Predictive Techniques," NUSC Technical Report 5303, 26 March 1976.

5. S.L. Marple, "A New Autoregressive Spectrum Analysis Algorithm,' *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 4, August 1980, pp. 441-454.

6. A.H. Nuttall, "Application of Linear Predictive Spectral Analysis to Multiple Tones in Noise," NUSC Technical Memorandum 791218, 12 December 1979.

7. A.H. Nuttall, *FORTRAN Program for Multivariate Linear Predictive Spectral Analysis, Employing Forward and Backward Averaging*, NUSC Technical Document 5419, 19 May 1976.

8. A.H. Nuttall, *Multivariate Linear Predictive Spectral Analysis, Employing Weighted Forward and Backward Averaging: A Generalization of Burg's Algorithm*, NUSC Technical Report 5501, 13 October 1976.

9. A.H. Nuttall, *Positive Definite Spectral Estimate and Stable Correlation Recursion for Multivariate Linear Predictive Spectral Analysis*, NUSC Technical Report 5729, 14 November 1977.

10. S.L. Marple, personal communication.

Experimental Comparison Of Three Multichannel Linear Prediction Spectral Estimators

S. L. Marple
A. H. Nuttall

ABSTRACT

Single-channel spectral estimators based on linear prediction techniques, such as the maximum-entropy method, have been shown to often provide better spectral stability and resolution than standard FFT procedures for short data sequences. Based on this improved performance, a multitude of multichannel linear prediction techniques have been promoted for processing multichannel data sequences. Three of these are examined in the paper: a multichannel generalization of the single-channel Burg algorithm by Nuttall, a maximum-entropy type of algorithm by Morf, Vieira, Lee and Kailath, and a multichannel extension of the covariance method of linear prediction implemented by Marple. For purposes of experimental comparison, various two-channel data sets were processed by the three methods to produce the two autospectra, the magnitude-squared coherence and the coherence phase associated with each data set. A possible deleterious effect of signal 'feed-across' between autospectra and in the coherence has been discovered in all three methods. The phenomenon, due to inexact pole-zero cancellation, is especially prominent for short data sequences. Based on the multichannel results given here, the Nuttall method generally produced the best spectral estimates.

Experimental comparison of three multichannel linear prediction spectral estimators

S. Lawrence Marple Jr. and Albert H. Nuttall

Indexing term: Signal processing

Abstract: Single-channel spectral estimators based on linear prediction techniques, such as the maximum-entropy method, have been shown to often provide better spectral stability and resolution than standard FFT procedures for short data sequences. Based on this improved performance, a multitude of multichannel linear prediction techniques have been promoted for processing multichannel data sequences. Three of these are examined in the paper: a multichannel generalisation of the single-channel Burg algorithm by Nuttall, a maximum-entropy type of algorithm by Morf, Vieira, Lee and Kailath, and a multichannel extension of the covariance method of linear prediction implemented by Marple. For purposes of experimental comparison, various two-channel data sets were processed by the three methods to produce the two autospectra, the magnitude-squared coherence and the coherence phase associated with each data set. A possible deleterious effect of signal 'feed-across' between autospectra and in the coherence has been discovered in all three methods. The phenomenon, due to inexact pole-zero cancellation, is especially prominent for short data sequences. Based on the multichannel results given here, the Nuttall method generally produced the best spectral estimates.

1   Introduction

Multichannel digital signal processing is being used in an increasing number of application areas, particularly in the sonar and geoseismic communities. Until recently, most multichannel digital signal processing was based on fast Fourier transform (FFT) methods. The success of unichannel high-resolution spectral estimation techniques, like the autoregressive or so-called maximum-entropy methods, has encouraged researchers to develop multichannel extensions in hope of obtaining performance improvements for multichannel applications comparable with that seen in unichannel applications. The multichannel extensions are all based on linear prediction concepts, since a linear prediction filter whitens an autoregressive process.

Three multichannel linear prediction spectrum analysis algorithms are examined in this paper. They are the multichannel generalisation of the Burg algorithm developed by Nuttall [1–4], a multichannel maximum-entropy spectral estimate by Morf, Vieira, Lee and Kailath [5, 6], and a multichannel generalisation of the covariance method of linear prediction as implemented by Marple [7]. All three algorithms make estimates of the multichannel linear prediction coefficients, from which the multichannel autoregressive autospectra and cross-spectra may be computed. From the cross-spectra, the magnitude-squared coherence and coherence phase may be computed.

There has been no previous attempt in the literature to compare and characterise the various multichannel linear-prediction/autoregressive spectral estimators. An experimental approach is used in this paper to empirically characterise performance with respect to two signal classes. One class is an actual two-channel autoregressive process. The other class is a set of tones (sinusoids) imbedded in a coloured noise process. An analytic approach for calculating multichannel algorithm performance was felt to be mathematically intractable, given the very complex analysis that was required to characterise the single-channel autoregressive spectral estimate [8, 9].

Relative to the particular multichannel results reported here, the Nuttall algorithm generally produced the best results. Best in this case means with less frequency estimation bias and variance than the other methods when tested against autoregressive and tonal processes. During the testing, a 'feed-across' effect was discovered that was common to all techniques. Narrowband components that should be present in only one channel's autospectra were found coupled into another channel's autospectral estimate. This is shown to be due to inexact pole-zero cancellation in the autospectral estimate. Conditions that would cause spectral line splitting in the single-channel case were found to also cause splitting in two of the three linear prediction techniques considered here.

2   Summary of techniques

This Section provides an overview of the three multichannel linear prediction methods considered. If we define X_n as the vector of M channel samples at time index n

$$X_n = \begin{bmatrix} x_n(1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_n(M) \end{bmatrix}$$

from a stationary zero-mean multichannel process, then the covariance function at time lag k is given by

$$R_k = E[X_{n+k} X_n^H] = R_{-k}^H$$

The symbol H denotes Hermitian transpose, and E denotes statistical expectation. Define the forward linear prediction error for prediction order p as

$$Y_k = \sum_{n=0}^{p} A_n^{(p)} X_{k-n}$$

and the backward linear prediction error as

$$Z_k = \sum_{n=0}^{p} B_n^{(p)} X_{k-p+n}$$

where $\{A_n^{(p)}\}$ and $\{B_n^{(p)}\}$ are, respectively, the forward and backward linear prediction coefficient matrices of dimension

$M \times M$. Note that $A_0^{(p)} = B_0^{(p)} = I$ (the identity matrix) by convention.

Minimisation of the forward and backward linear prediction mean square errors $\mathrm{tr}\{E[Y_k Y_k^H]\} = \mathrm{tr}\{U_p\}$ and $\mathrm{tr}\{E[Z_k Z_k^H]\} = \mathrm{tr}\{V_p\}$, respectively, (tr denotes the matrix trace) will yield coefficient matrices that satisfy the block-Toeplitz normal equation

$$\begin{bmatrix} I & A_1^{(p)} & \cdots & A_{p-1}^{(p)} & A_p^{(p)} \\ B_p^{(p)} & B_{p-1}^{(p)} & \cdots & B_1^{(p)} & I \end{bmatrix} \begin{bmatrix} R_0 & R_1^H & \cdots & R_p^H \\ R_1 & R_0 & \cdots & R_{p-1}^H \\ \vdots & & \ddots & \vdots \\ \vdots & & & \vdots \\ R_p & R_{p-1} & \cdots & R_0 \end{bmatrix} = \begin{bmatrix} U_p & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & V_p \end{bmatrix}$$

The solution to this normal equation, when the covariance matrices $\{R_k\}_0^p$ are known, is provided by the multichannel (vector) Levinson-Wiggins-Robinson (LWR) algorithm. Briefly, this algorithm relates the order p solution to the order $p-1$ solution according to the recursions

$$\Delta_p = \sum_{n=0}^{p-1} A_n^{(p-1)} R_{p-n}$$

$$A_p^{(p)} = -\Delta_p (V_{p-1})^{-1}$$

$$B_p^{(p)} = -\Delta_p^H (U_{p-1})^{-1}$$

$$U_p = (I - A_p^{(p)} B_p^{(p)}) U_{p-1}$$

$$V_p = (I - B_p^{(p)} A_p^{(p)}) V_{p-1}$$

$$A_k^{(p)} = A_k^{(p-1)} + A_p^{(p)} B_{p-k}^{(p-1)} \qquad \text{for} \quad 1 < k < p-1$$

$$B_k^{(p)} = B_k^{(p-1)} + B_p^{(p)} A_{p-k}^{(p-1)} \qquad \text{for} \quad 1 < k < p-1$$

with initial conditions

$$U_0 = V_0 = R_0$$

$$A_0^{(k)} = B_0^{(k)} = I \qquad \text{for} \quad 0 < k < p$$

The $M \times M$ matrix Δ_p is often called the reflection coefficient matrix.

Based on the linear prediction coefficients, the multichannel autoregressive power spectral density matrix estimate may be shown to be [1]

$$G_A(z) = \Delta t \, [A(z)]^{-1} U_p \, [A(1/z^*)]^{-H}$$
$$= \Delta t \, [B(z)]^{-1} V_p \, [B(1/z^*)]^{-H}$$
$$= G_B(z)$$

where Δt is the sample interval, $-H$ denotes the Hermitian transpose of the inverse, the asterisk denotes complex conjugate, and

$$A(z) = \sum_{n=0}^{p} A_n^{(p)} z^{-n}$$

$$B(z) = \sum_{n=0}^{p} B_n^{(p)} z^{-n}$$

The substitution $z = \exp(j2\pi f \Delta t)$ is made, and G_A is evaluated as a function of the frequency f. With this substitution having been made, then

$$G(f) = \begin{bmatrix} G_{11}(f) & G_{12}(f) \\ G_{21}(f) & G_{22}(f) \end{bmatrix}$$

is the power spectral density matrix in the two-channel case. Here $G_{11}(f)$ and $G_{22}(f)$ are the autospectra of channel 1 and channel 2, respectively, and the magnitude-squared coherence (MSC) is given by

$$\mathrm{MSC} = |G_{12}(f)|^2 / \{G_{11}(f) G_{22}(f)\}$$

The coherence phase spectrum (CPS) is simply

$$\mathrm{CPS} = \arg[G_{12}(f)]$$

3 Unknown covariance: Measured data

The three linear prediction algorithms examined in this paper are concerned with the situation in which there are data samples available, but no covariance values are known at any lags. Assuming N vectors X_n of channel samples have been collected, then the following squared-error and cross-error covariance estimates may be formed:

$$E_p = \frac{1}{N-p} \sum_{k=p+1}^{N} Y_k Y_k^H$$

$$F_p = \frac{1}{N-p} \sum_{k=p+1}^{N} Z_{k-1} Z_{k-1}^H$$

$$G_p = \frac{1}{N-p} \sum_{k=p+1}^{N} Y_k Z_{k-1}^H$$

The Nuttall extension of the Burg algorithm to the multichannel case makes use of the LWR algorithm, with the exception of how the reflection coefficient Δ_p is computed, since covariance values are now unavailable. Using the error covariance estimates, Δ_p is obtained as the solution of the bilinear equation

$$(V_{p-1})^{-1} F_p \Delta_p + \Delta_p (U_{p-1})^{-1} E_p = -2G_p$$

in order to satisfy a weighted arithmetic mean criterion between the forward and backward squared errors E_p and F_p.

The so-called maximum entropy algorithm of Morf et al. also uses the LWR algorithm, but the reflection coefficient Δ_p is computed as the weighted geometric mean of the forward and backward squared errors:

$$\Delta_p = [E_p]^{-1/2} G_p [F_p^H]^{-1/2}$$

The superscript notation $-1/2$ means the square-root matrix of the inverse, since a normalised form of the LWR algorithm is used by the authors of this linear prediction method.

Instead of forcing the LWR recursions to hold in the given data case, as in the two previous algorithms, one could separately minimise the forward and backward squared errors

$$\mathrm{tr}\left[\sum_{n=p+1}^{N} Y_n Y_n^H\right]$$

$$tr\left[\sum_{n=p+1}^{N} Z_{n-1} Z_{n-1}^{H}\right]$$

in the least-squares sense. This yields the normal equation

$$\begin{bmatrix} I & A_1^{(p)} \ldots \ldots A_{p-1}^{(p)} & A_p^{(p)} \\ B_p^{(p)} & B_{p-1}^{(p)} \ldots \ldots B_1^{(p)} & I \end{bmatrix} \begin{bmatrix} R_{00}^{(p)} & \ldots \ldots R_{0p}^{(p)} \\ \vdots & \ddots \quad \vdots \\ R_{p0}^{(p)} & \ldots \ldots R_{pp}^{(p)} \end{bmatrix}$$

$$= \begin{bmatrix} U_p & 0 \ldots \ldots 0 & 0 \\ 0 & 0 \ldots \ldots 0 & V_p \end{bmatrix}$$

where

$$R_{ij}^{(p)} = \sum_{n=p+1}^{N} X_{n-i} X_{n-j}^{H} \qquad for\ 0 \leqslant i, j \leqslant p$$

and the matrix of $R^{(p)}$ terms is no longer block-Toeplitz. However, a fast recursive algorithm similar in structure to the LWR algorithm can solve the above least-squares normal equation; see Reference 7 for details. The fast algorithm is a multichannel extension of the single-channel covariance linear prediction algorithm used in speech processing [10]. Since the block-Toeplitz property is lost in the least-squares minimisation of this method, the estimated power spectral densities $G_A(z)$ and $G_B(z)$ do not, in general, yield mathematically identical values, although the plotted spectra are often of comparable shapes.

4 Experimental comparisons

Plots of autospectra, MSC and CPS from various two-channel test cases are used here to examine relative experimental performance differences among the three linear prediction spectrum analysis techniques. The discussion is restricted to real-valued data and filters.

In the first test, a first-order ($p = 1$) two-channel autoregressive process was used to generate data ensembles. Data were generated according to the recursion

$$\left. \begin{aligned} X_n &= A_1 X_{n-1} + W_n \\ A_1 &= \begin{bmatrix} 0.85 & -0.75 \\ 0.65 & 0.55 \end{bmatrix} \end{aligned} \right\} \qquad (1)$$

where W_n is a white process, uncorrelated from channel to channel and of unit variance in both channels. The exact autospectra for each of the two channels are shown in Figs. 1a and b, while the MSC and phase of the complex coherence (cross-spectrum) are shown in Figs. 1c and d. There is seen to be a strong narrowband component at approximately 1/8 of the sampling frequency; the peak magnitude of the squared coherence is 0.999013 at this point. 20 sequences, each of 100 sample points, were processed by each of the three algorithms, with the order p restricted to be 1. The 20 spectral plots were overlapped, as shown in Figs. 2–5. All plots are made to the same reference values established in Fig. 1. Of the four sets of plots, clearly the Nuttall linear prediction algorithm has less variance and bias (location of the spectral
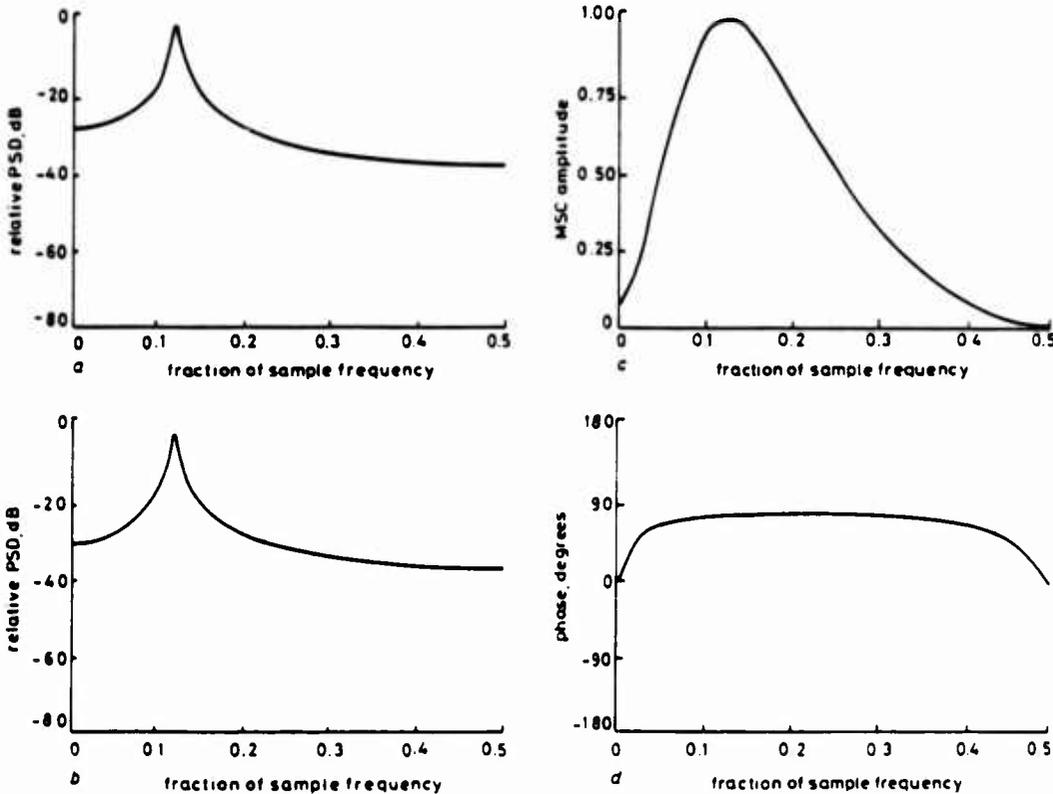
Fig. 1 *True autospectra and coherence of the AR (1) process*

a G_{11} autospectrum
b G_{22} autospectrum
c Magnitude-squared coherence
d Coherence phase
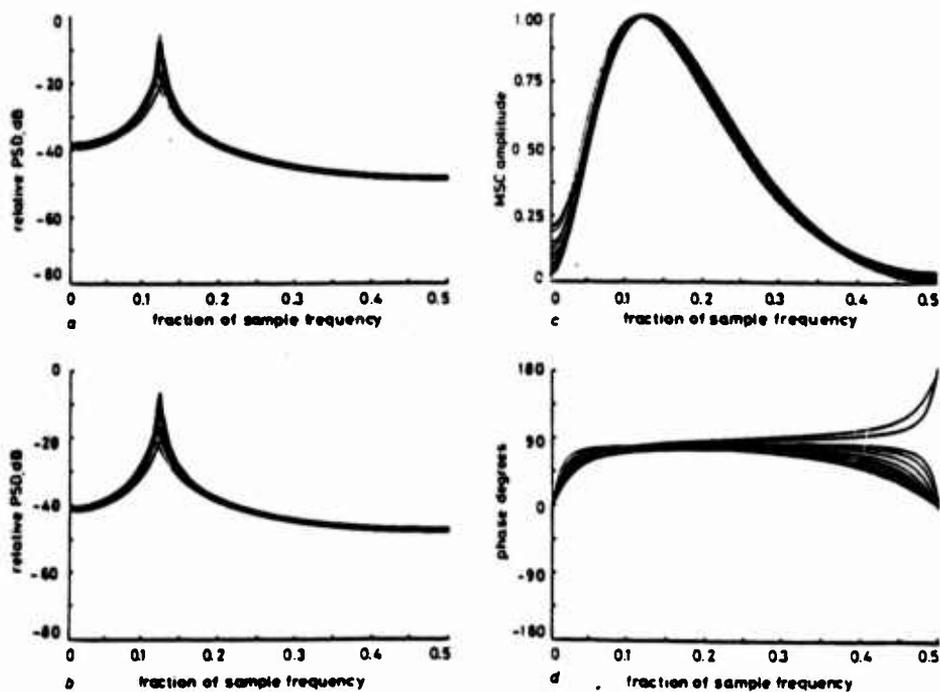
Fig. 2 *Autospectra and coherence of 20 overlapped data sets: Nuttall-Burg algorithm generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
c Magnitude-squared coherence
d Coherence phase

Fig. 3 *Autospectra and coherence of 20 overlapped data sets: Morf et al. maximum-entropy generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
c Magnitude-squared coherence
d Coherence phase

Fig. 4 *Autospectra and coherence of 20 overlapped data sets: Marple covariance generalisation (forward)*

a G_{11} autospectrum
b G_{22} autospectrum
c   Magnitude-squared coherence
d   Coherence phase

Fig. 5 *Autospectra and coherence of 20 overlapped data sets: Marple covariance generalisation (backward)*

a G_{11} autospectrum
b G_{22} autospectrum
c   Magnitude-squared coherence
d   Coherence phase

peak) than the other methods. In fact, 18 out of 20 of the phase estimates were close to the truth for the Nuttall algorithm, whereas none of those given by the other techniques was correct at all. Note that the four plots of Fig. 4, based on the forward linear prediction estimate, differ little from those in Fig. 5, based on the backward linear prediction estimate, even though the covariance method does not guarantee that the two cases will yield the same result.

This first test has very important implications. All three techniques should work well against a simple first-order autoregressive process, since this is the data model appropriate for linear prediction spectral estimators. Clearly, however, the Nuttall technique works best, at least for short data records. The other two techniques exhibit more variance than is typically found in single-channel versions of these methods. For a multichannel pth-order regression and independent white excitations, it may easily be shown that the auto- and cross-spectra of the processes each possess $4p$ poles and $3p$ zeros in the finite z-plane (of which p zeros occur at the origin). This is in contrast to the single-channel case, where only $2p$ poles (and a pth-order zero only at the origin) can occur. Thus, while a single-channel linear prediction requires estimation of only p parameters, an M-channel approximation requires estimation of $M^2 p$ parameters per channel. For a fixed number N of data points from each process, the estimation of an increased number of parameters can only be done with increased variance since the total number of data points is only MN. This is a partial explanation of the results seen in Figs. 2–5.

The next test consisted of data sequences with three sinusoids in a coloured noise process. In channel 1, sinusoids with fractions of sampling frequency of 0.1, 0.2 and 0.24 were used. The respective amplitudes were 0.1, 1 and 1, while the respective initial phases were 0, 90 and 235°. The coloured noise process, which had most of its power above the frequencies of the sinusoids, was generated by passing white noise of 0.05 variance through a digital filter with a raised-cosine spectral response between 0.2 and 0.5 of the sampling rate. In channel 2, sinusoids of fractional frequencies 0.1, 0.2 and 0.4, amplitudes of 0.1, 1 and 1, and initial phases of 0, 210 and 25° were generated. A coloured noise process similar to channel 1, but independently generated, was added to the sinusoids. Note that the sinusoid components at fractional frequencies 0.1 and 0.2 are common to both channels. A data sequence of 64 points was generated for each channel. Figs. 6–8 depict the autospectra and coherence magnitude and phase for order $p = 12$ estimates by the three linear prediction methods. Note that in all plots, some energy of the 0.24 fractional frequency sinusoid, present only in channel 1, has coupled into the G_{22} estimate for channel 2, and, conversely, the 0.4 fractional frequency sinusoid, present only in channel 2, has coupled into the G_{11} estimate for channel 1. The MSC plots also show spikes at the tone frequencies of 0.24 and 0.4, where ideally the coherence should be zero. This 'feed-across' artifact is an undesirable effect; however, it appears unavoidable whenever processing short data

Fig. 6 *Autospectra and coherence of sinusoids in noise (64 samples): Nuttall-Burg algorithm generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
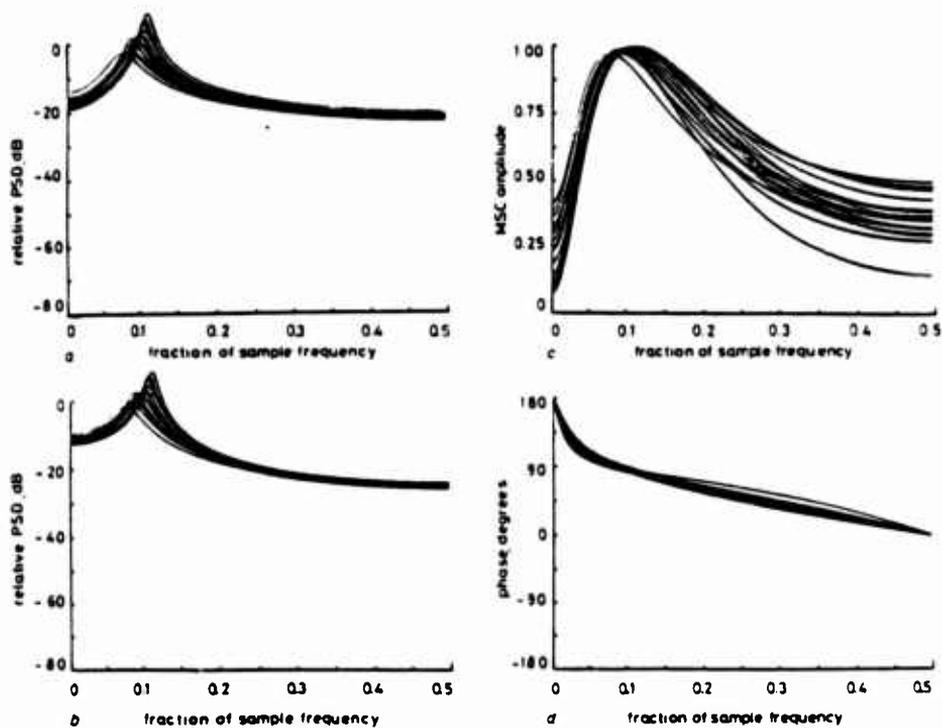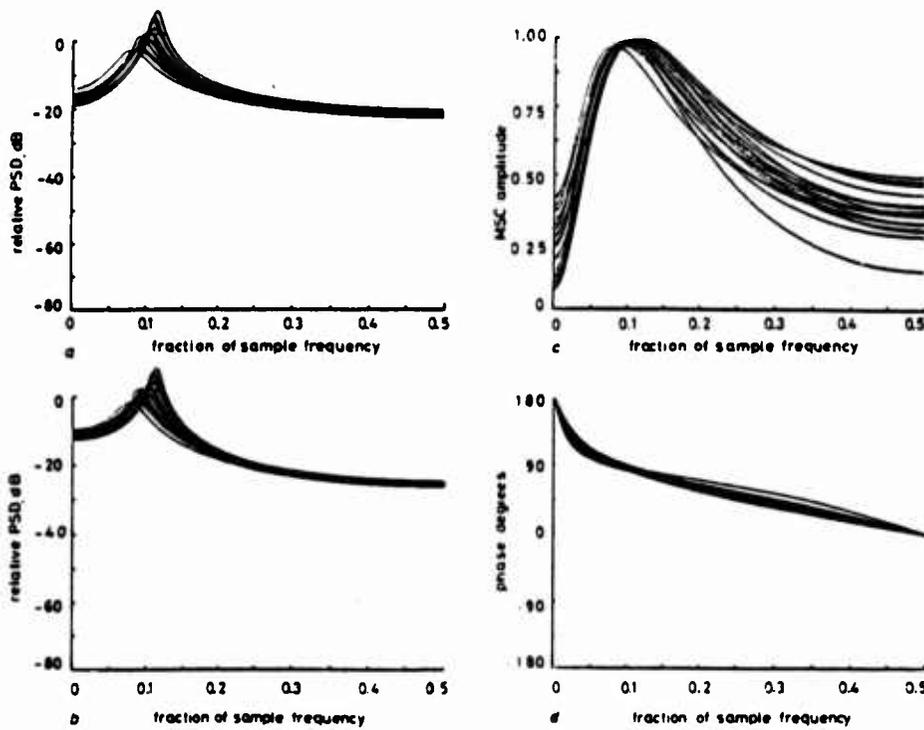c   Magnitude-squared coherence
d   Coherence phase

sequences with multichannel algorithms. If the data sequences from each channel are processed separately by the single-channel Burg linear prediction algorithm, the autospectra shown in Fig. 9 are the result. Note that these spectra do not give a false tonal indication. By increasing the data sequence length to 500 samples, there is an improvement in the feed-across effect in the multichannel autospectra in Fig. 10, although the coherence still has strong sharp spikes. Thus it seems that the attractive properties of single-channel linear prediction spectrum analysis relative to periodogram spectrum analysis when processing short data records do not extend to the multichannel case without problems. The feed-across effect leads to false indications of narrowband components, especially when processing short data sequences. A detailed analysis of the cause of the feed-across effect is presented in the following Section.

The final test case examined the spectral line splitting phenomenon that has been reported in the literature for single-channel linear prediction spectrum analysis. A data sequence of 101 samples of a single tone at 0.0725 fraction of sampling frequency, unit amplitude and 45° initial phase, added to the same coloured noise process (variance 10^{-4}) as in the last test, was generated for channel 1. A comparable 101-sample sequence was generated for channel 2, but the tone was placed at a fractional frequency of 0.2175. Both cases are known to cause line splitting in the single-channel Burg algorithm. Figs. 11—13 show the autospectra generated by the three multichannel linear prediction methods for order $p = 15$. Line splitting is present in the Nuttall and Morf methods, but not in the covariance method. Thus line splitting behaviour carries over into some of the multichannel techniques.

5 Properties of linear predictive spectral analysis techniques

The behaviour of the linear predictive techniques in the presence of tonals is of importance to anyone using these spectral analysis procedures. We now present some of the important properties of multichannel linear prediction techniques and illustrate these properties by an example from the Nuttall algorithm for the two-channel autoregressive process in eqn. 1, with a tone added solely to the channel-2 process. The tone strength is $- 6.6$ dB relative to the sample power in the second component of X_n, and the tone is located at fractional frequency $f_c = 0.3$; this is basically the example considered in Fig. 11 of Reference 4.

The autospectra and coherence estimates for a data run with $N = 1000$ data points and $p = 8$ are given in Fig. 14. There is the desired strong tonal indication at f_c in G_{22}, but, in addition, there is a weak undesired contribution at f_c in G_{11}. Since this tone was never added to process 1, this indication in G_{11} has 'leaked across' in the mathematical data manipulations of the linear predictive algorithm. This leakage is unavoidable and will be present in all multichannel linear predictive procedures; it is due to the fact that we must work with finite data sets, thereby resulting in slightly inaccurate filter coefficients. The effects on the poles and zeros of the autospectral and coherence estimates are discussed below. The MSC estimate in Fig. 14c has developed two notches and a sharp spike near frequency f_c, while the phase in Fig. 14d has gone through an abrupt 2π change in that neighbourhood.

In order to explain the various behaviours of the spectral
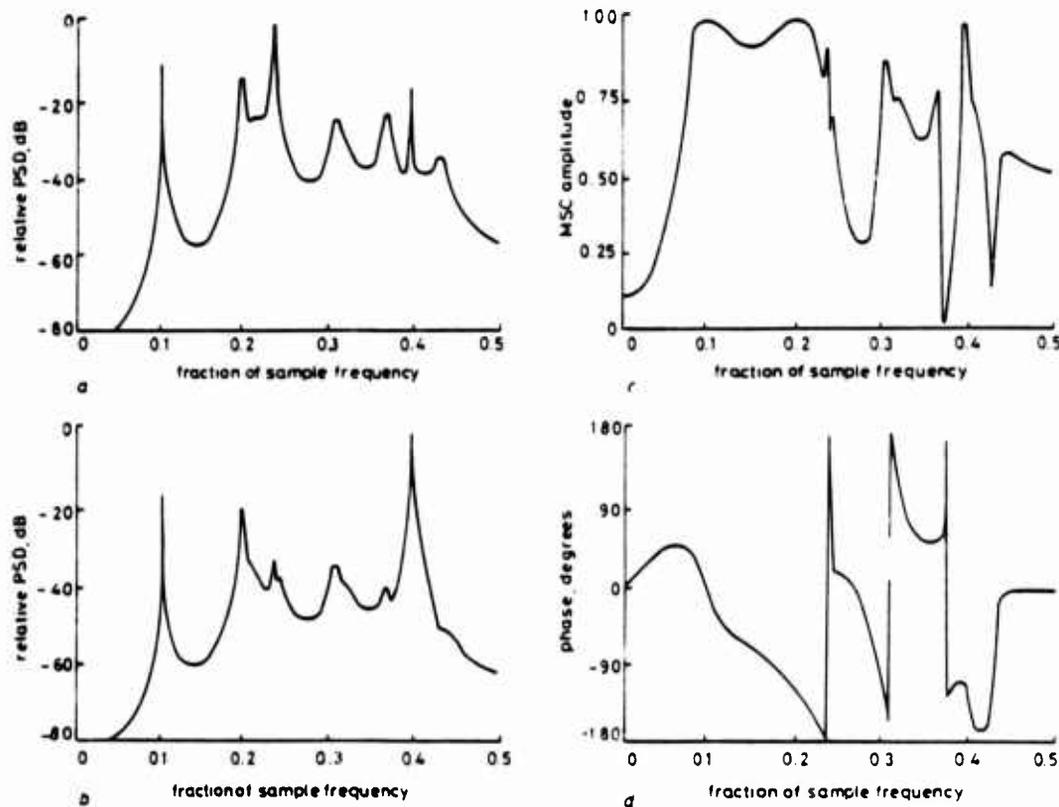
Fig. 7   Autospectra and coherence of sinusoids in noise (64 samples). Morf et al. maximum-entropy generalisation

a G_{11} autospectrum
b G_{22} autospectrum
c   Magnitude-squared coherence
d   Coherence phase
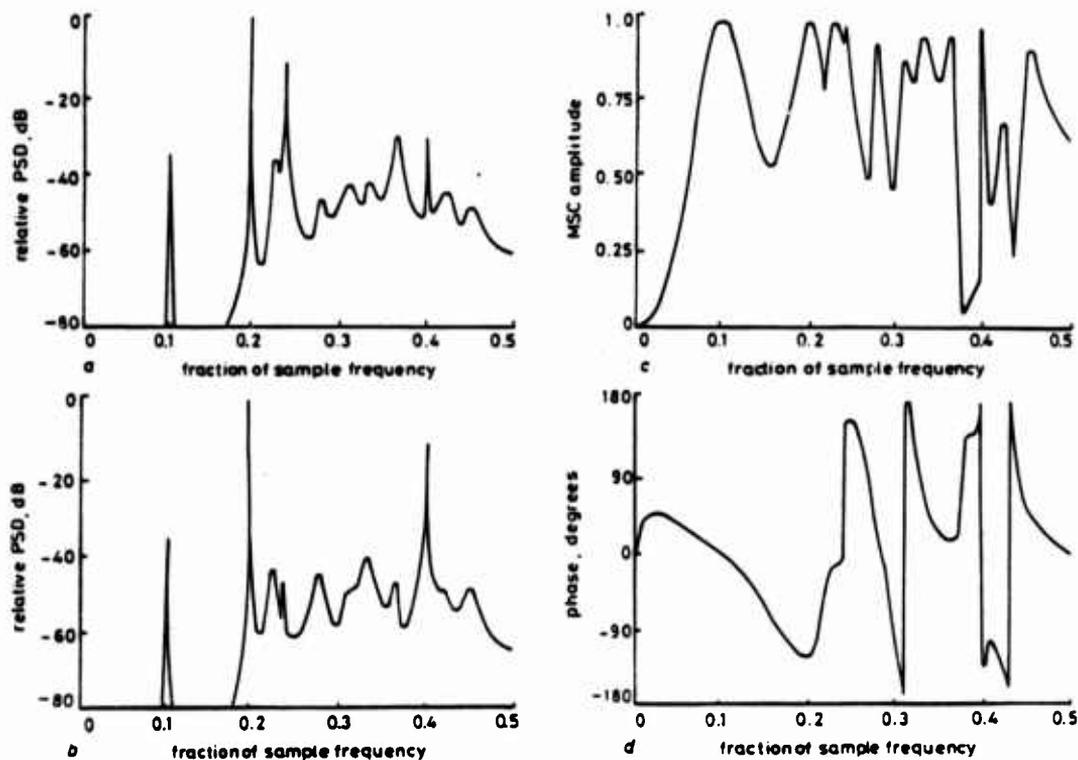
Fig. 8 *Autospectra and coherence of sinusoids in noise (64 samples): Marple covariance generalisation (forward)*

a G_{11} autospectrum
b G_{22} autospectrum
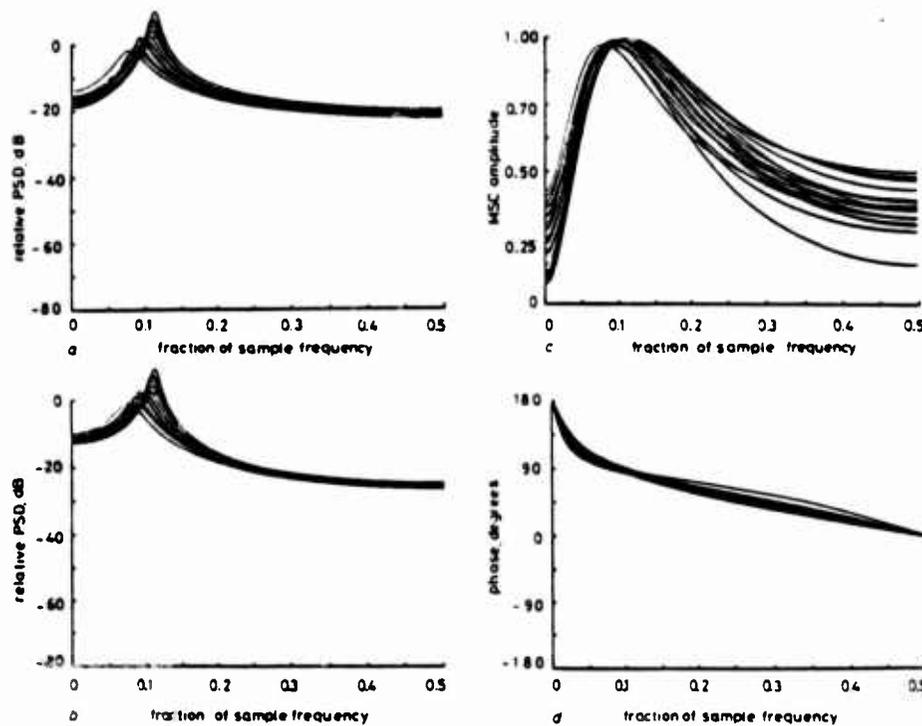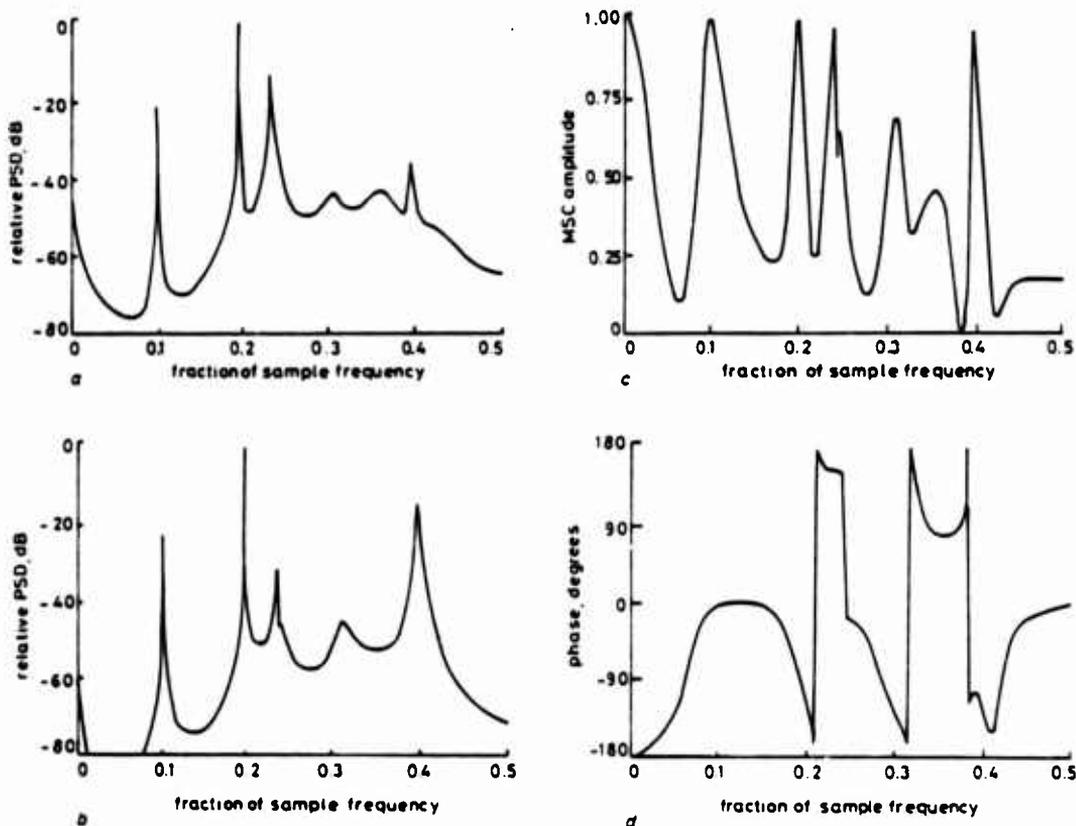c   Magnitude-squared coherence
d   Coherence phase
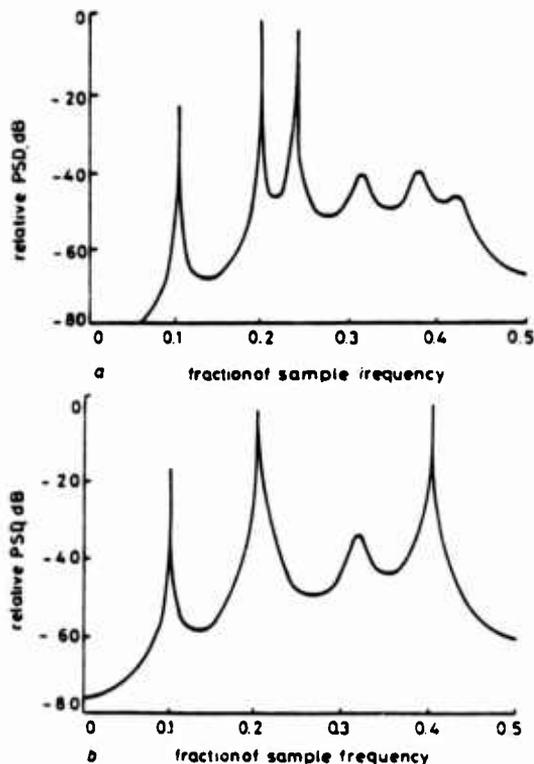
Fig. 9 *Autospectra of sinusoids in noise (64 samples) Single-channel Burg algorithm*

a G_{11} autospectrum
b G_{22} autospectrum

estimates in Fig. 14, we need to develop, in more detail, the matrix spectral estimate (see eqn. F-16 of Reference 1)

$$G(z) = [A(z)]^{-1} U_p [A(1/z^*)]^{-H} \tag{2}$$

where

$$A(z) = I + \sum_{n=1}^{p} z^{-n} A_n^{(p)} \tag{3}$$

For real data and filters, U_p and $\{A_n^{(p)}\}$ are real. For a two-channel application, eqn. 3 can be expressed as

$$A(z) = \begin{bmatrix} h_{11}(z) & h_{12}(z) \\ h_{21}(z) & h_{22}(z) \end{bmatrix} \tag{4}$$

where $\{h_{kl}(z)\}$ are scalars of the form

$$h_{kl}(z) = \sum_{n=0}^{p} z^{-n} a_{kl}(n) \tag{5}$$

There follows for eqn. 2

$$G(z) = \frac{1}{D(z)D(1/z)} \begin{bmatrix} h_{22}(z) & -h_{12}(z) \\ -h_{21}(z) & h_{11}(z) \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

$$\times \begin{bmatrix} h_{22}(1/z) & -h_{21}(1/z) \\ -h_{12}(1/z) & h_{11}(1/z) \end{bmatrix} \tag{6}$$

where

$$D(z) = h_{11}(z)h_{22}(z) - h_{12}(z)h_{21}(z) \tag{7}$$

We denote

$$h_{kl}(z) = z^{-p} q_{kl}(z)$$
$$h_{kl}(1/z) = r_{kl}(z)$$

$\qquad\qquad(8)$

where $q_{kl}(z)$ and $r_{kl}(z)$ are polynomials of degree p in z with real coefficients. Then

$$D(z) = z^{-2p} Q(z)$$
$$D(1/z) = R(z)$$

$\qquad\qquad(9)$

where $Q(z)$ and $R(z)$ are polynomials of degree $2p$ in z. It follows that

$$Q(0) = a_{11}(p) a_{22}(p) - a_{12}(p) a_{21}(p)$$

$$R(0) = 1$$

neither of which is zero; thus $Q(z)$ and $R(z)$ have no zeros at the origin of the z-plane. The spectral estimate can now be expressed as

$$G(z) = \frac{z^p}{Q(z) R(z)} \begin{bmatrix} N_{11}(z) & N_{12}(z) \\ N_{21}(z) & N_{22}(z) \end{bmatrix}$$

$\qquad\qquad(10)$

where the 2×2 matrix in eqn. 10 is given by

$$\begin{bmatrix} q_{22}(z) & -q_{12}(z) \\ -q_{21}(z) & q_{11}(z) \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} r_{22}(z) & -r_{21}(z) \\ -r_{12}(z) & r_{11}(z) \end{bmatrix}$$

$\qquad\qquad(11)$

All the functions in eqn. 10 and expr. 11 are polynomials in z.

For the real data and filters employed here, the $2p$ zeros of $Q(z)$ occur in conjugate pairs (or in real pairs). Furthermore, they are all inside the unit circle C_1 in the complex z-plane; this property was proved in eqn. 33 of Reference 3. Thus we only need to search the interior of the unit hemisphere in the upper-half z-plane for p zeros of $Q(z)$.

Furthermore, if $Q(z)$ has a zero at z_0, then $R(z)$ has a zero at $1/z_0$; thus $G(z)$ in eqn. 10 has $4p$ poles in the z-plane, $2p$ in the upper halfplane, of which p are inside the unit hemisphere. A typical 4-tuple of poles in $G(z)$ is given by z_0, z_0^*, $1/z_0$, $1/z_0^*$; there are p such 4-tuples. These poles are common to all the auto- and cross-spectral estimates in eqn. 10.

Similarly, a typical term, $N_{kl}(z)$, in the numerator of $G(z)$ in eqn. 10 is a polynomial of degree $2p$ in z, of which p zeros will be located in the upper-half z-plane. Thus eqn. 10 shows that (every term of) $G(z)$ has p zeros at 0, p zeros at ∞ and $2p$ zeros in the finite plane. We need only search for the p zeros of $N_{kl}(z)$ in the upper-half plane, for $k, l = 1, 2$.

In order to explain the behaviour of the coherence estimates given by the linear predictive techniques, we start with the complex coherence at frequency f as given (for $\Delta t = 1$) by

$$C(f) = \frac{N_{12}(\exp(j2\pi f))}{[N_{11}(\exp(j2\pi f)) N_{22}(\exp(j2\pi f))]^{1/2}}$$

$\qquad\qquad(12)$

The squared coherence, generalised to the entire z-plane, is

$$\mathscr{C}^2(z) = \frac{N_{12}^2(z)}{N_{11}(z) N_{22}(z)} \quad \left(\text{not} \quad \frac{N_{12}(z) N_{21}(z)}{N_{11}(z) N_{22}(z)} \right)$$

$\qquad\qquad(13)$

Fig. 10   *Autospectra and coherence of sinusoids in noise (500 samples). Nuttall-Burg algorithm generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
c   Magnitude-squared coherence
d   Coherence phase

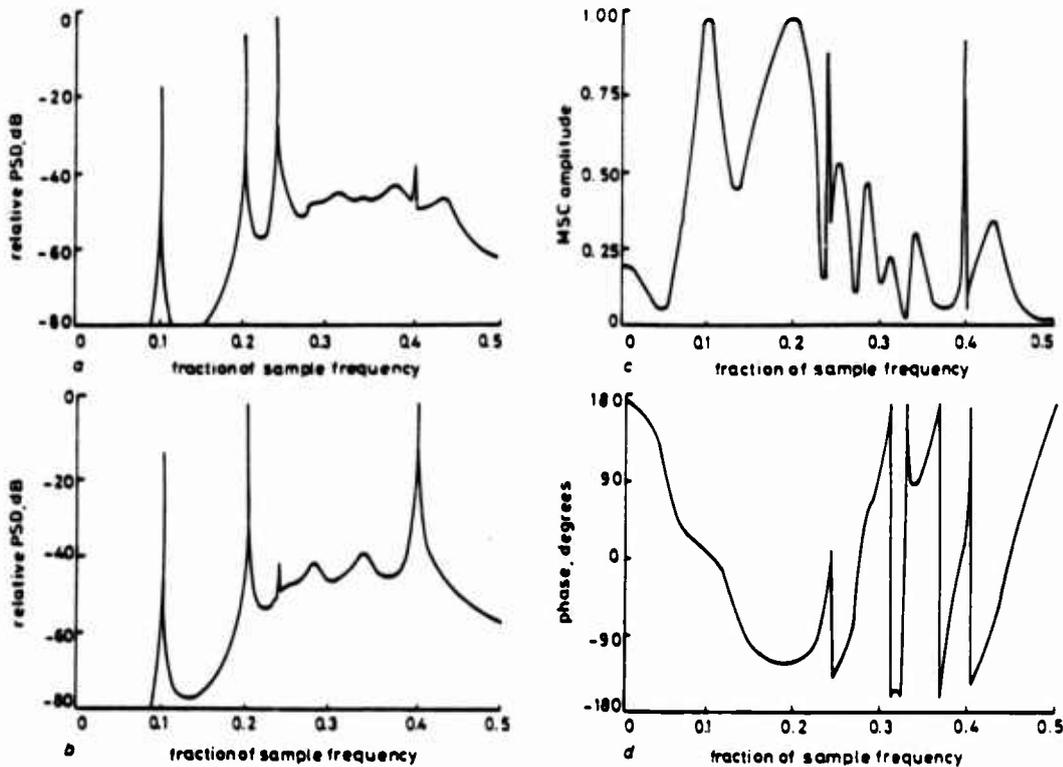The following are general properties of the $N_{kl}(z)$ polynomials:

(a) If $N_{11}(z)$ has a zero at $r_1 \exp(j\theta_1)$, it also has a zero at $(1/r_1) \exp(j\theta_1)$.

(b) If $N_{22}(z)$ has a zero at $r_2 \exp(j\theta_2)$, it also has a zero at $(1/r_2) \exp(j\theta_2)$.

(c) If $N_{12}(z)$ has a zero at $r_0 \exp(j\theta_0)$, $N_{21}(z)$ has a zero at $(1/r_0) \exp(j\theta_0)$.

In fact, property (c) is a special case of the general rule that

$$N_{21}(z) = z^{2p} N_{12}(1/z) \qquad \text{for all } z \qquad (14)$$

Now we return to the example of Fig. 14, where a tone was present only in the channel-2 data, at frequency f_c. Since process 1 has no tone at f_c, $N_{11}(z)$ develops *two* zeros near $z_c \equiv \exp(j2\pi f_c) = \exp(j\pi 0.6)$, one inside the unit circle C_1 and one outside C_1, tending to cancel the one zero of $D(z)$ inside C_1 and the one zero of $D(1/z)$ outside C_1, which are near z_c, so that the autospectral estimate $G_{11}(f)$ is well behaved near f_c. However, the cancellation is not perfect, and the small spike at f_c in Fig. 14a remains. For this example, the pole location inside C_1 is $0.993591 \exp(j\pi\ 0.600104)$, whereas the zero location inside C_1 is $0.992697\ (j\pi\ 0.600015)$. Since the pole is closer to C_1, a positive-going perturbation occurs in Fig. 14a at f_c.

Similarly, the cross-spectrum of $G(z)$ ideally should have no indication at f_c since there is no tone in process 1. In order to counter the zeros of $D(z)$ and $D(1/z)$ near z_c, $N_{12}(z)$ develops *two* zeros near z_c. For the same example, they are at $0.996958 \exp(j\pi 0.592593)$ and $0.983192 \exp(j\pi 0.615533)$, both of which happen to be inside C_1. Other

examples have shown that these two zeros can both be o side C_1, or one can be inside and one outside C_1.

Finally, since process 2 does have a tone at f_c, $N_{22}(f)$ does *not* develop any zeros near z_c. Thus the zeros of $D(z)$ and $D(1/z)$ dominate near z_c, and the estimate $G_{22}(f)$ in Fig. 14b develops a large value near f_c, as desired.

The squared coherence estimate in eqn. 13 is independent of $D(z)$ or $Q(z)$. However, according to the discussion above, it has two double-zeros near z_c, owing to the $N_{12}^2(z)$ term. These four zeros can lie either all inside C_1, all outside C_1, or two inside and two outside C_1. $\mathscr{C}^2(z)$ also has two poles near z_c, owing to the two zeros of $N_{11}(z)$ near z_c: one pole lies inside C_1; the other lies outside C_1. Since there are no poles and zeros near z_c that cancel exactly, some very fine detail can develop in the coherence estimate in the neighbourhood of z_c. Sharp notches and spikes are the rule, not the exception, in the MSC evaluated on C_1 in the neighbourhood of a tonal frequency owing to this imperfect cancellation of poles and zeros. The phase variations can be so rapid that large FFT sizes are required to track it accurately. The two double-zeros of $N_{12}^2(z)$ cause this rapid variation, especially if they are all located on the same side of C_1; the two zeros of $N_{11}(z)$ are always located on opposite sides of C_1 and so do not themselves lead to a very rapid phase change near z_c, although they greatly influence the MSC in that region. In general, the squared coherence $\mathscr{C}^2(z)$ has $2p$ double-zeros and $4p$ poles in the complex z-plane, none restricted to be inside or outside of C_1; however, they all occur in conjugate pairs.

Since order $p = 8$ used in this example is larger than
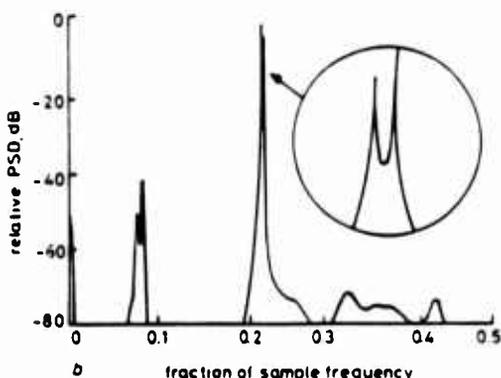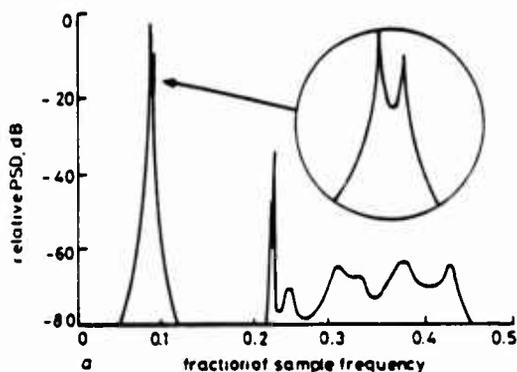
Fig. 11 *Autospectra of data set producing line splitting. Nuttall-Burg algorithm generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
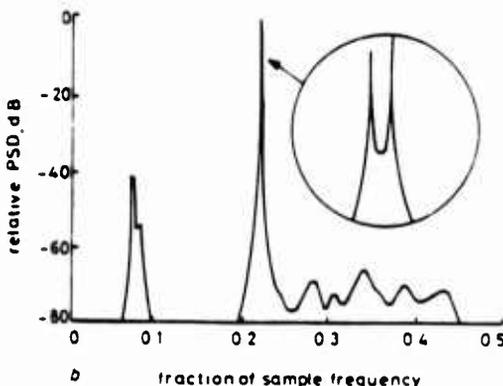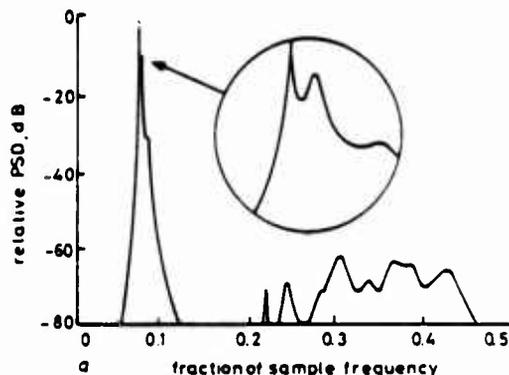
Fig. 12 *Autospectra of data set producing line splitting. Morf et al. maximum-entropy generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
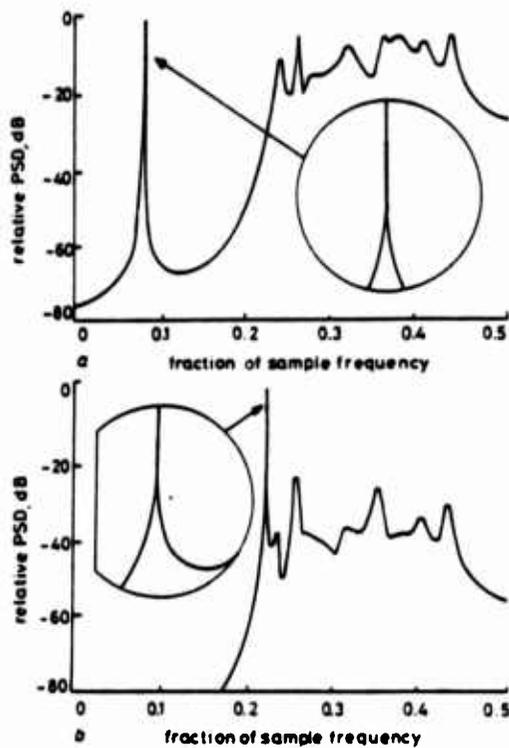
Fig. 13 *Autospectra of data set that produced line splitting in Figs. 11 and 12: Marple covariance algorithm generalisation (forward)*
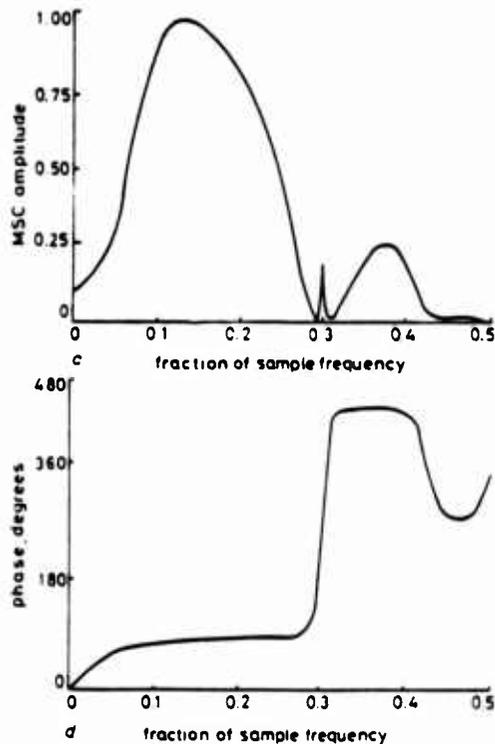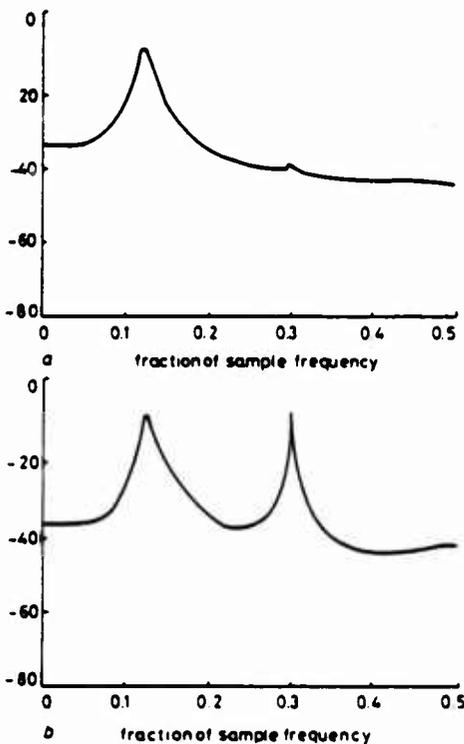
a G_{11} autospectrum
b G_{22} autospectrum

necessary to account for a single tone and a narrowband spectral component, the extra poles [zeros of $Q(z)$] are distributed fairly uniformly inside C_1, with radii in the range 0.5 to 0.7. The extra zeros of polynomials $N_{k1}(z)$ can lie anywhere, either inside or outside of C_1.

6 Discussion and conclusions

Experimental results for two-channel linear prediction auto- and cross-spectral estimation have been presented for a first-order autoregressive process and for cases with interfering tones. It has been shown that some misleading estimates may be obtained because of feed-across in the mathematical manipulations of the finite lengths of data from each channel. This feed-across manifests itself as narrow spurious spikes in the spectral and coherence estimates. In order to circumvent this problem, while maintaining the high-resolution properties of linear prediction techniques, the following philosophy for multichannel spectrum analysis is suggested.

Suppose we are given finite data records of three stationary processes $x(t)$, $y(t)$ and $z(t)$, and we wish to estimate all the autospectra and cross-spectra involved. The Blackman and Tukey and weighted-FFT approaches evaluate the autospectrum of each process separately. Thus the spectrum of $x(t)$ is estimated without interference from $y(t)$ and $z(t)$; the availability of the data records for $y(t)$ and $z(t)$ plays no part in the eventual autospectral estimate for $x(t)$. Additionally, the cross-spectral estimate for processes $x(t)$ and $y(t)$ is independent of the available data on the $z(t)$ process. Finally, the coherence estimate between two processes is independent of any additional data records for other (statistically related) processes.

Fig. 14 *Autospectra and coherence of first-order autoregressive process plus single sinusoid (1000 samples, $p = 8$) Nuttall-Burg algorithm generalisation*

a G_{11} autospectrum
b G_{22} autospectrum
c Magnitude-squared coherence
d Coherence phase

On the other hand, the three multichannel linear predictive spectral analysis approaches give autospectral estimates of the $x(t)$ process that are dependent on the available values of $y(t)$ and $z(t)$. Also, the cross-spectral estimate between $x(t)$ and $y(t)$ is dependent on the particular $z(t)$ data available. This procedure can be poor for short data lengths if, for example, $y(t)$ contains a strong tone at f_0 that is not present in $x(t)$ or $z(t)$. Thus estimates of spectra $G_{xx}(f)$, $G_{xy}(f)$ and $G_{xz}(f)$ all contain tonal indications at f_0 that should not be there. These spurious tonal indications are due to feed-across between the available finite data segments of the various processes.

This raises the following questions:

(i) Should the estimate of $G_{xx}(f)$ be determined only from the available $x(t)$ data record?

(ii) Should the estimate of $G_{xy}(f)$ be determined only from the available $x(t)$ and $y(t)$ data records?

(iii) If coherence $C_{xy}(f_0) = 0$, why use $y(t)$ to estimate $G_{xx}(f_0)$?

(iv) If coherence $C_{xy}(f_0) = 1$, why use the completely statistically dependent $y(t)$ data to estimate $G_{xx}(f_0)$?

This philosophy of discarding 'irrelevant' data would be consistent with the Blackman and Tukey and FFT approaches. Carrying this philosophy on, we are led to the following: estimate $G_{xx}(f)$ solely from the $x(t)$ data by some single-channel linear predictive technique. Then estimate cross-spectrum $G_{xy}(f)$ or coherence $C(f)$ directly, by some linear predictive technique whose sole goal is linear prediction of $x(t)$ from $y(t)$ and vice versa, with no interest in or diversion from simultaneous estimation of $G_{xx}(f)$ or $G_{yy}(f)$. By this means, we can concentrate on extracting all the relevant cross-spectral information with maximum stability and resolution. Other cross-spectra of interest between particular pairs of available processes can be similarly obtained, one at a time.

7 References

1 NUTTALL, A.H.: 'Multivariate linear predictive spectral analysis employing weighted forward and backward averaging: A generalization of Burg's Algorithm'. NUSC technical report 5501, Oct. 1976

2 NUTTALL, A.H.: 'FORTRAN program for multivariate linear predictive spectral analysis, employing forward and backward averaging'. NUSC technical report 5419, May 1976

3 NUTTALL, A.H.: 'Positive definite spectral estimate and stable correlation recursion for multivariate linear predictive spectral analysis'. NUSC technical report 5729, Nov. 1977

4 NUTTALL, A.H.: 'Two-channel linear-predictive spectral analysis: Program for the HP 9845 desk calculator'. NUSC technical report 6533, Oct. 1981

5 MORF, M., VIEIRA, A., LEE, D.T.L., and KAILATH, T.: 'Recursive multichannel maximum entropy spectral estimation', IEEE Trans., 1978, GE-16, pp. 85–94

6 MORF, M., VIEIRA, A., and KAILATH, T.: 'Covariance characterization by partial autocorrelation matrices', Ann. Stat., 1978, 6, pp. 643–648

7 MARPLE, S.L. Jr.: 'Multichannel spectral analysis: A generalization of the fast covariance linear prediction algorithm'. Final report to NUSC on contract N00140-81-C-BT82, Nov. 1981

8 BERK, K.N.: 'Consistent autoregressive spectral estimates', Ann. Stat., 1974, 2, pp. 489–502

9 KROMER, R.E.: 'Asymptotic properties of the autoregressive spectral estimator'. Ph.D. dissertation, Department of Statistics, Stanford University, California, USA, Dec. 1969

10 MORF, M., DICKINSON, B., KAILATH, T., and VIEIRA, A.: 'Efficient solution of covariance equations for linear prediction', IEEE Trans., 1977, ASSP-25, pp. 429–433

Direct Coherence Estimation Via A Constrained Least-Squares Linear-Predictive Fast Algorithm

A. H. Nuttall

ABSTRACT

Estimation of multi-channel auto-spectra, cross-spectra, and coherences via linear-predictive techniques suffers from cross-feed between the limited number of data points available of each channel process. Here, the coherence between two channels is estimated directly, without diversion from auto-spectra estimation, so as to minimize only the valid computable linear-predictive errors that can be formed from given finite data records. Furthermore, the least-squares minimization procedure is constrained so as to yield physically-consistent magnitude-squared coherence estimates, and can be accomplished by a fast algorithm in a recursive fashion. The resulting nonlinear programming problem that is encountered is solvable via a recursion employing the Hessian of an unconstrained error measure. Examples of application of the technique to short data records yield good coherence estimates.

DIRECT COHERENCE ESTIMATION VIA A CONSTRAINED LEAST-SQUARES LINEAR-PREDICTIVE FAST ALGORITHM

Albert H. Nuttall

Naval Underwater Systems Center
New London, Connecticut
06320 USA

ABSTRACT

Estimation of multi-channel auto-spectra, cross-spectra, and coherences via linear-predictive techniques suffers from cross-feed between the limited number of data points available of each channel process. Here, the coherence between two channels is estimated directly, without diversion from auto-spectra estimation, so as to minimize only the valid computable linear-predictive errors that can be formed from given finite data records. Furthermore, the least-squares minimization procedure is constrained so as to yield physically-consistent magnitude-squared coherence estimates, and can be accomplished by a fast algorithm in a recursive fashion. The resulting non-linear programming problem that is encountered is solvable via a recursion employing the Hessian of an unconstrained error measure. Examples of application of the technique to short data records yields good coherence estimates.

INTRODUCTION

Estimation of the magnitude-squared coherence (MSC) between two stationary processes x and y has usually been accomplished by estimating the auto-spectra and cross-spectrum of the two processes and forming the ratio

$$\hat{C}(f) \equiv \frac{|\hat{G}_{xy}(f)|^2}{\hat{G}_{xx}(f)\,\hat{G}_{yy}(f)}.$$

Eq. 1

Here, f is frequency, $\hat{G}_{xx}(f)$ and $\hat{G}_{yy}(f)$ are the auto spectral estimates, and $\hat{G}_{xy}(f)$ is the cross spectrum estimate.

Recently, a more direct method of estimating the MSC was presented (1,2), whereby two filters are used to linearly estimate y from x and also x from y. The product of the two filter transfer functions is then used as an estimate of the MSC. In this way, no denominator estimates (as in Eq. 1) are ever required. Some related work is given in (3).

However, there are some limitations in this past work that need further consideration. In (1, Eq. 4), the squared errors were not confined to the valid errors that can be computed from given finite data records of x and y. This results in the claim (1, page 647) that the matrix that must be inverted

1

is Toeplitz. Although this is true, the use of invalid error terms and Toeplitz matrices has been found in (4) to yield poorer auto-spectral estimates, and probably does so for MSC estimation as well. The adaptive approach in (2, Eq. 6) uses only available data, but does not make optimum use of given limited data segments, in terms of variance reduction.

In addition, in both (1) and (2), since no constraints are imposed on the two filters that do the linear predictions of y from x, and x from y, respectively, the product of the transfer functions need not be real, although the true (unknown) MSC is always real. Thus, in (1, Eq. 14), the magnitude of the product is taken, whereas in (2, Eq. 7a), no mention is made of the complex nature of the right-hand side. The reason for this behavior is the independent calculation of the two linear-predictive filters from finite data records.

Finally, in both (1, Eqs. 4 and 13) and (2, Eq. 2), the number of forward weights is set equal to the number of backward weights. This unnecessary limitation should be eliminated, as may be readily seen by considering the case where one process is a (noisy) time-delayed version of the other. Insistence on an equal number of forward and backward weights can result in the unnecessary computation of many near-zero weights.

In this paper, we limit the least-squares calculation solely to the valid errors that can be formed from finite data records; we constrain the two filters that do the linear predictions so as to yield a real MSC estimate; and we allow the number of forward and backward weights to be arbitrary. We also use a fast recursive algorithm to solve for the filter coefficients.

BASIS OF TECHNIQUE

For continuous-time stationary processes x and y, the linear non-causal filter with input x, that minimizes the mean-square error in predicting the value of y at the same time instant, has a transfer function given by $G_{yx}(f)/G_{xx}(f)$. Conversely, linear prediction of x, based upon data record y, is accomplished by a linear filter with transfer function $G_{xy}(f)/G_{yy}(f)$. The product of these two transfer functions is then precisely the MSC between x and y.

For discrete-time stationary processes x and y, exactly the same relations ensue, except that the spectra are discrete Fourier transforms of discrete correlation functions. Again, the two filters are allowed to be non-causal and of infinite extent in the impulse response domain.

In the practical case, the various required auto- and cross-correlations and spectra above are not known, only finite data records are available, and the filters must have limited impulse response durations. Also, the mean-square (ensemble average) error must be replaced by a measurable physically-meaningful sum of squared errors and then minimized by choice of the free parameters. The details of such a procedure to accomplish MSC estimation are presented below.

2

FORMATION OF VALID ERRORS

Suppose that data records of $x(m)$ for $M_1 \leq m \leq M_2$ and of $y(n)$ for $N_1 \leq n \leq N_2$ are available. We linearly estimate $y(n)$ from the available data record $\{x(m)\}$ according to

$$\hat{y}(n) \equiv \sum_{k=-q}^{p} a(k)\, x(n-k), \qquad \text{Eq. 2}$$

where p and q are arbitrary integers. We select $-q \leq p$ without loss of generality; i.e., $p+q \geq 0$. These integers should be chosen so that the region of $\{x(m)\}$, on which $y(n)$ is statistically dependent, is covered by (non-causal) impulse response $\{a(k)\}$. Similarly, estimation of $x(m)$ is accomplished according to the linear filtering operation

$$\hat{x}(m) \equiv \sum_{k=-p}^{q} b(k)\, y(m-k). \qquad \text{Eq. 3}$$

Eqs. 2 and 3 are consistent in their selection of the extents of the non-zero coefficients in impulse responses $\{a(k)\}$ and $\{b(k)\}$, in that the statistically dependent portions of x and y are covered equally well, regardless of which process is predicted from the other.

In order to compare $\hat{y}(n)$ with $y(n)$, and $\hat{x}(m)$ with $x(m)$, and thereby form valid errors, we must restrict n and m according to

$$N_3 \leq n \leq N_4, \text{ where } N_3 = \max(N_1, M_1+p), \quad N_4 = \min(N_2, M_2-q), \quad \text{Eq. 4}$$

$$M_3 \leq m \leq M_4, \text{ where } M_3 = \max(M_1, N_1+q), \quad M_4 = \min(M_2, N_2-p), \quad \text{Eq. 5}$$

where it is presumed that $N_3 \leq N_4$ and $M_3 \leq M_4$. The sums of squared errors are then

$$E_y \equiv \sum_{n=N_3}^{N_4} \left| y(n) - \hat{y}(n) \right|^2 = \sum_{n=N_3}^{N_4} \left| y(n) - \sum_{k=-q}^{p} a(k)\, x(n-k) \right|^2, \quad \text{Eq. 6}$$

$$E_x \equiv \sum_{m=M_3}^{M_4} \left| x(m) - \hat{x}(m) \right|^2 = \sum_{m=M_3}^{M_4} \left| x(m) - \sum_{k=-p}^{q} b(k)\, y(m-k) \right|^2. \quad \text{Eq. 7}$$

The data points required on the right-hand sides of Eqs. 6 and 7 are all available in the given records; i.e., these equations involve only valid errors.

The set of optimum filter weights $\{a_0(k)\}$ that minimize error E_y is given by the solution of

$$\sum_{k=-q}^{p} R_{xx}(\ell,k)\, a_0(k) = R_{yx}(\ell) \quad \text{for} \; -q \leq \ell \leq p, \qquad \text{Eq. 8}$$

where

$$R_{xx}(\ell,k) \equiv \sum_{n=N_3}^{N_4} x(n-k)\, x^*(n-\ell) \quad \text{for} \; -q \leq \ell, k \leq p, \qquad \text{Eq. 9}$$

$$R_{yx}(\ell) \equiv \sum_{n=N_3}^{N_4} y(n)\, x^*(n-\ell) \quad \text{for} \; -q \leq \ell \leq p. \qquad \text{Eq. 10}$$

Eq. 8 constitutes p+q+1 linear equations in the p+q+1 unknowns $\{a_0(k)\}_{-q}^{p}$. Furthermore, with an appropriate shift of time origin, a recursive fast algorithm (5) exists for the minimization of E_y in Eq. 6, i.e., the solution of Eq. 8. Similarly, the independent minimization of E_x in Eq. 7, by choice of optimum filter weights $\{b_0(k)\}$, can be realized by a second call to that same fast algorithm. Thus, we have an efficient procedure for determining the optimum unconstrained filter weights that separately minimize E_x and E_y; these are called the unconstrained least-squares filters.

We can now evaluate transfer functions, and form the product $A_0(f)B_0(f)$, where

$$A_0(f) = \sum_{k=-q}^{p} a_0(k) \exp(-i2\pi f k), \quad B_0(f) = \sum_{k=-p}^{q} b_0(k) \exp(-i2\pi f k), \qquad \text{Eq. 11}$$

as an approximation to the MSC for $|f| < 1/2$. An example of the result for 1000 common data points in each sequence and for $p = q = 6$ is given in Fig. 1. The actual data points were generated by a two-channel first-order auto-regressive process according to (6, pages 5-6)

$$x(k) = .85\, x(k-1) - .75\, y(k-1) + u(k),$$

$$y(k) = .65\, x(k-1) + .55\, y(k-1) + v(k), \qquad \text{Eq. 12}$$

where $\{u(k)\}$ and $\{v(k)\}$ are independent white uniformly-distributed random processes.

The real and imaginary parts of $A_0(f)B_0(f)$ are plotted as solid curves in Fig. 1, while the true MSC for Eq. 12 is plotted as a dashed curve. Since errors E_x and E_y were independently minimized, the imaginary part of

4

$A_0(f)B_0(f)$ is not zero, which is not consistent with a MSC which must be real. However, it is conceivable that the size of the imaginary part of $A_0(f)B_0(f)$ could serve as an indicator of the statistical accuracy of the real part as an estimator of the MSC, if p and q are large enough; this possibility has not been pursued.

CONSTRAINED LEAST-SQUARES SOLUTIONS

In order that the transfer functions in Eq. 11 have a real product for all f, it may be shown (for real filter weights) that the convolution of filter weight sequences $\{a(k)\}$ and $\{b(k)\}$ must be even. Explicitly, this means that we must have (for $p+q \geq 1$)

$$\sum_{k=-q}^{p-n} a(k)b(-n-k) = \sum_{k=n-q}^{p} a(k)b(n-k) \quad \text{for} \quad 1 \leq n \leq p+q. \qquad \text{Eq. 13}$$

This equation constitutes p+q quadratic equality constraints on the $2(p+q+1)$ unknown filter coefficients.

We now must form a single physically-meaningful error quantity and minimize it subject to the p+q quadratic constraints in Eq. 13. We select the average of the two-squared errors defined in Eqs. 6 and 7; this seems reasonable when the two available sequences are normalized to have the same energy. Mathematically, this translates into determining the stationary points (not minima) of the quantity

$$Q \equiv \frac{1}{2}(E_x + E_y) + \sum_{n=1}^{p+q} \lambda(n) \sum_k a(k)\left[b(n-k) - b(-n-k)\right], \qquad \text{Eq. 14}$$

where $\{\lambda(n)\}$ are Lagrange multipliers; we set $\{a(k)\}$ and $\{b(k)\}$ equal to zero except for the ranges utilized above in Eqs. 2 and 3.

The solution for the stationary points of Eq. 14, obtained by setting all the partial derivatives of Q with respect to $\{a(k)\}$, $\{b(k)\}$, and $\{\lambda(n)\}$ equal to zero, results in $3(p+q)+2$ non-linear equations. Ordinarily, the numerical solution to these equations would be very difficult; in fact, there are generally many solutions. However, we already have a good approximation (at least for many data points) to these solutions, in the form of the solutions to the unconstrained least-squares approach of the previous section. Furthermore, recall that the unconstrained least-squares filters can be determined by a fast algorithm. Thus, if we use a recursive technique and employ the Hessian (7) of the quantity Q in Eq. 14, along with the starting values from the unconstrained solutions and zero starting values for the Lagrange multipliers, we will have an efficient recursive method of getting the desired result, namely a purely real MSC estimate via minimization of a physically-meaningful error criterion.

The results of this constrained procedure for the same pair of 1000 data-point sequences as above, and for $p = q = 6$, is shown in Fig. 2, after the recursion for the stationary point of Q has reached its final values. The

5

imaginary part is zero, and the real part is substantially the same as Fig. 1, except for a slightly better behavior for the small MSC values near the Nyquist frequency.

DISCUSSION

If p and q are not chosen large enough to encompass the regions of linear dependence of processes x and y, poor estimates of the MSC can result, even though the number of data points gets large. On the other hand, too large values of p and q results in excessive statistical fluctuations in the MSC estimate. Thus, the familiar trade-off between frequency resolution and stability must be faced.

For many cases, the real part of $A_0(f)B_0(f)$ for the unconstrained approach may suffice as the MSC estimate, without bothering to resort to the recursive approach for a zero imaginary part. This would also avoid the need for inverting a large non-Toeplitz matrix of size $3(p+q)+2$ in the case when p+q is large. If so, use of a fast algorithm, as indicated here, affords a very efficient procedure for getting good MSC estimates via minimization of valid errors only.

REFERENCES

(1) Y. T. Chan, J. M. Riley, and J. B. Plant, "A Wiener Filter Approach to Coherence Estimation," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, Colorado, pp. 646-649, 9-11 April 1980.

(2) D. H. Youn, N. Ahmed, G. C. Carter, and S. Vijayendra, "Magnitude-Squared Coherence Function Estimation: An Adaptive Approach," submitted to IEEE Transactions on Acoustics, Speech, and Signal Processing in January 1982.

(3) D. B. Harris, "Recursive Least Squares with Linear Constraints," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, Georgia, pp. 526-529, 30 March - 1 April 1981.

(4) A. H. Nuttall, "Spectral Analysis of a Univariate Process with Bad Data Points, via Maximum Entropy and Linear Predictive Techniques," NUSC Technical Report 5303, New London, CT, 26 March 1976.

(5) S. L. Marple, "Efficient Least-Squares FIR System Identification," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol ASSP-29, No. 1, pp. 62-73, February 1981.

(6) A. H. Nuttall, "Two-Channel Linear-Predictive Spectral Analysis; Program for the HP 9845 Desk Calculator," NUSC Technical Report 6533, New London, CT, 1 October 1981.

(7) The International Dictionary of Applied Mathematics, D. Van Nostrand Co., Inc., Princeton, N.J., 1960. See p. 445.
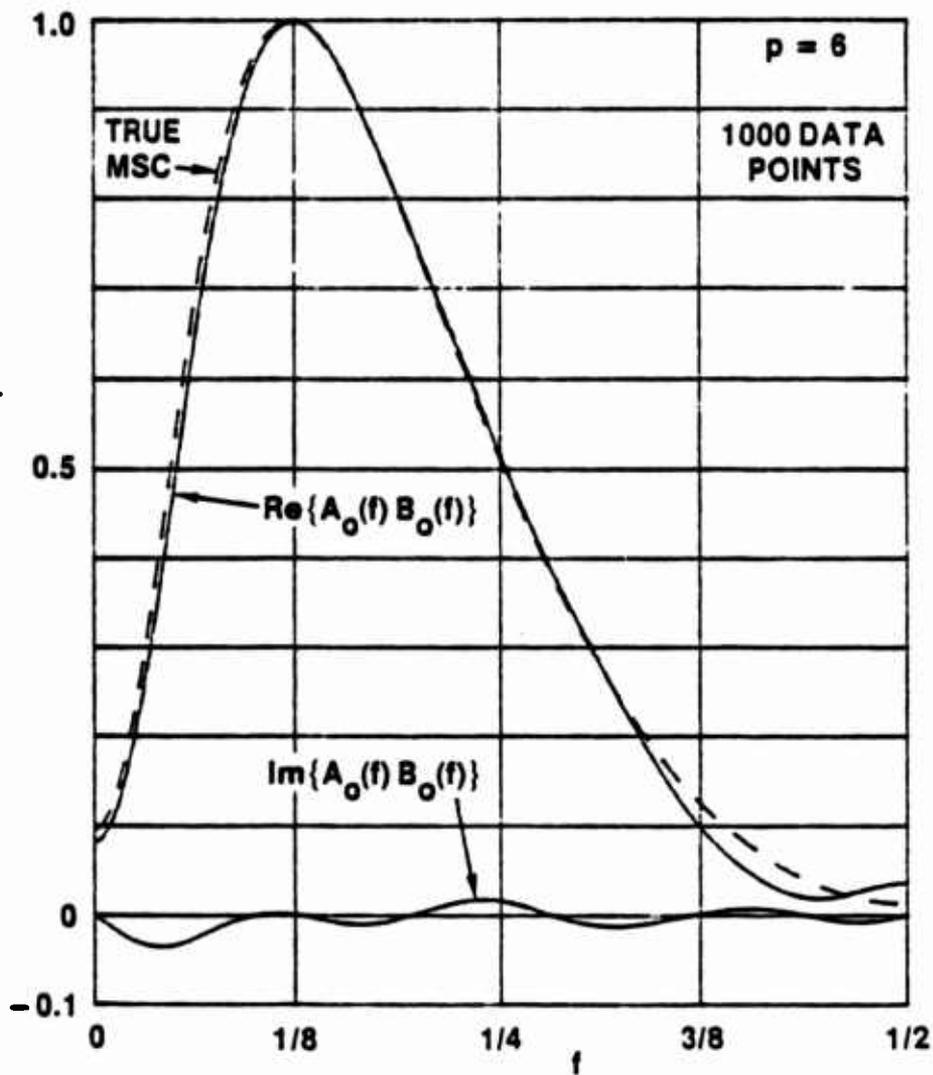
FIG. I.   UNCONSTRAINED ESTIMATE
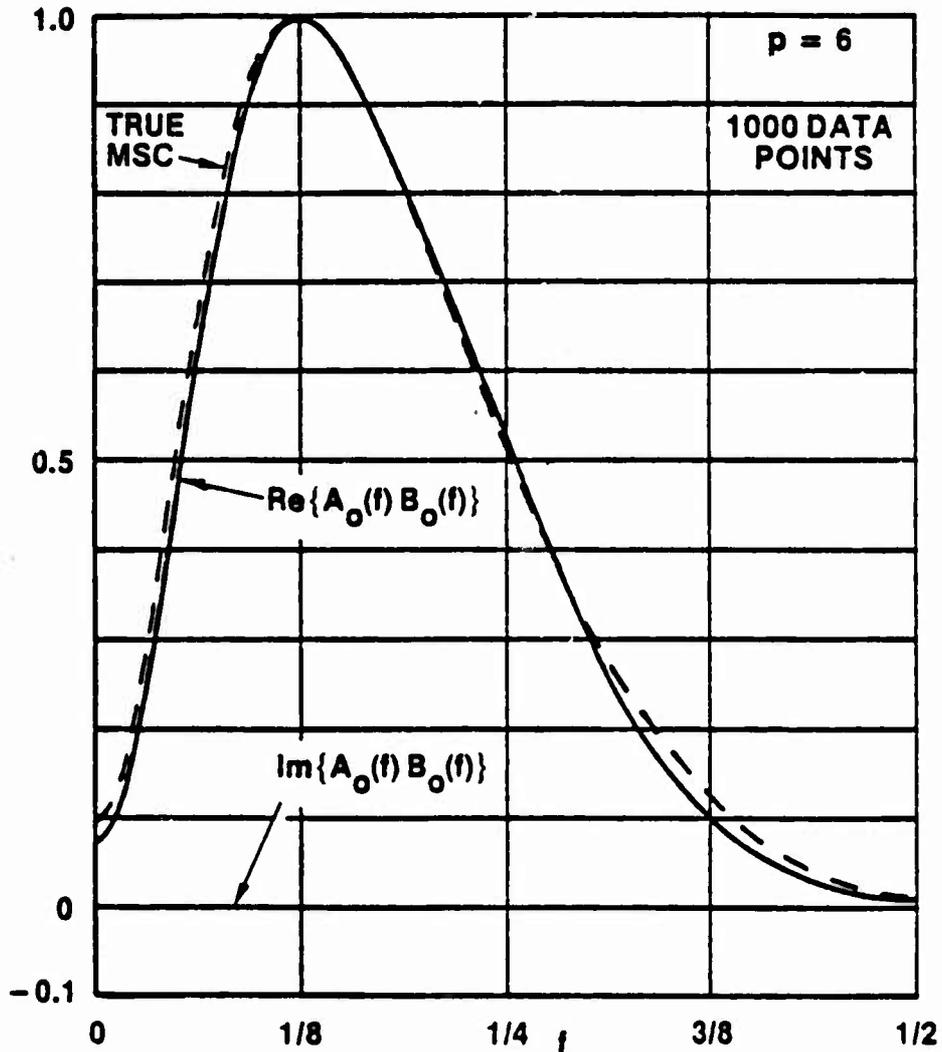
FIG. 2. CONSTRAINED ESTIMATE

A Two-Parameter Class of Bessel Weightings With Controllable Side-Lobe Behavior for Linear, Planar-Circular, Volumetric-Spherical Arrays; The Ideal Weighting-Pattern Pairs

A. H. Nuttall

ABSTRACT

A two-parameter class of Bessel weightings is presented whose patterns have controllable sidelobe decay and mainlobe-to-peak-sidelobe ratio, for one-, two-, and three-dimensional arrays. In the one-dimensional application, the class of weightings subsumes the Kaiser-Bessel weighting as a special case, and extends to the ideal van der Maas weighting as a limiting case. In the two- and three-dimensional applications, all the results are new; the ideal patterns in these latter cases are also derived and found to require weightings with generalized functions that are more singular than the delta functions required for one dimension. Approximations to the generalized functions are presented.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF SYMBOLS

|	
R	radius or half-length of array
λ	wavelength of plane-wave arrival
ϕ_a, θ_a	arrival angles of plane wave
ϕ_χ, θ_χ	look angles of array
s	normalized distance on array
$w(s)$	normalized weighting of array
$g(u)$	voltage response pattern of array
u	dimensionless parameter of array; (4), (6), (8)
$J_\alpha(x)$	Bessel function of order α and argument x
$I_\alpha(x)$	modified Bessel function of order α and argument x
K	kernel of transform (10)
μ	parameter indicating dimensionality of array; table 1
$\mathcal{J}_\alpha(z)$	Bessel function ratio, (17)
$\mathcal{Q}_\alpha(z)$	Bessel function ratio, (18)
ν, B	weighting parameters, (22)
$\Gamma(x)$	gamma function
α	composite order of Bessel function, (29)
$g_i(u)$	ideal pattern, (34)
z_α	smallest nonzero null location of $J_\alpha(z)$, (38)
u_0	first null location of pattern $g(u)$
u_p	first sidelobe peak location of $g(u)$
sub G	generalized function, (50)

A TWO-PARAMETER CLASS OF BESSEL WEIGHTINGS WITH CONTROLLABLE SIDELOBE BEHAVIOR FOR LINEAR, PLANAR-CIRCULAR, AND VOLUMETRIC-SPHERICAL ARRAYS; THE IDEAL WEIGHTING-PATTERN PAIRS

INTRODUCTION

A wide variety of time-domain weightings for spectral analysis, whose frequency-domain windows have very good sidelobe behavior, have been presented in [1,2]. Since the basic mathematics describing the response of a weighted linear array can also be written as a Fourier transform, these weighting-window pairs have immediate application to one-dimensional array processing as well as spectral analysis.

Most of the weighting-window pairs in [1,2] have no parameters in their design equations; that is, the windows are fixed and cannot be altered, as for example, in the Hanning and Hamming windows. A few windows, such as the Dolph-Chebyshev and Kaiser-Bessel [3,4], do have a single parameter in their design equations that allows for a tradeoff between the mainlobe width and the ratio of mainlobe to peak sidelobe. However, neither have any control over the rate of decay of the sidelobes, the Dolph-Chebyshev case having no decay, and the Kaiser-Bessel case a 6 dB/octave decay. It is obvious that in order to control both the sidelobe decay and the mainlobe-to-peak-sidelobe ratio, a two-parameter family of weightings is necessary. And it is desirable (although not necessary) for the window to possess a simple analytical form that can be easily understood and evaluated for a range of parameter values. Such a class of Bessel weightings is presented in this report.

For the array application, the weighting is applied as a multiplicative factor in the spatial domain; the response to plane wave arrivals from various directions is called the pattern, rather than the window. Here we will give a two-parameter family of weighting-pattern pairs for use with arrays in one, two, or three dimensions, and shall indicate the ideal weightings and corresponding patterns in all cases. Special cases of this family will be shown to include some of the weightings that are currently employed in array and signal processing.

RESPONSES OF ONE-, TWO-, AND THREE-DIMENSIONAL ARRAYS

LINEAR ARRAY

We consider a continuous line array located on the x-axis in the range $(-R,R)$ and subject to symmetric weighting $w_1(x)$ for $|x| \leq R$. For a single-frequency plane wave of wavelength λ, arriving at angle ϕ_a relative to the normal to the line array, the array voltage response, by use of time-delay steering to look-angle ϕ_ℓ, is

$$g(u) = \int_{-R}^{R} dx \, \exp\left[-i2\pi\frac{x}{\lambda}(\sin\phi_a - \sin\phi_\ell)\right] w_1(x). \tag{1}$$

By letting $s = x/R$ and by using the symmetry of weighting w_1, we can express response (1) as

$$g(u) = \int_{0}^{1} ds \left(\frac{2}{\pi}\right)^{1/2} \cos(us) \, w(s), \tag{2}$$

where normalized weighting

$$w(s) = (2\pi)^{1/2} R \, w_1(Rs), \tag{3}$$

and dimensionless parameter

$$u = 2\pi\frac{R}{\lambda}(\sin\phi_a - \sin\phi_\ell) \tag{4}$$

incorporates the relevant features of array geometry, look angle, and the arrival wavelength and angle.

Thus the response (2) of a line array is a cosine transform of the normalized weighting. As an example, rectangular weighting yields response $g(u)$ proportional to $\sin u / u$, which has its first few nulls at $u = \pi$, 2π, 3π.

PLANAR ARRAY

The voltage response of a continuous planar-circular array of radius R, to a plane wave of wavelength λ arriving at (polar, azimuthal) angles $(\phi_a,\ \theta_a)$, and subject to weighting which depends only on the distance from the center of the array, is derived in appendix A, culminating in the result (A-11). It is

$$g(u) = \int_0^1 ds\ s\ J_0(us)\ w(s), \tag{5}$$

where $w(s)$ is the normalized weighting and

$$u = 2\pi\frac{R}{\lambda}\left[\sin^2\phi_\ell + \sin^2\phi_a - 2\sin\phi_\ell \sin\phi_a \cos(\theta_\ell - \theta_a)\right]^{1/2}. \tag{6}$$

Here $(\phi_\ell,\ \theta_\ell)$ are the (polar, azimuthal) look angles; that is, the response (5) of a planar-circular array is basically a zero-th order Bessel transform of the normalized weighting. As an example, rectangular weighting w yields response $g(u)$ proportional to $J_1(u)/u$, which has its first few nulls at $u = 3.83$, 7.02, 10.17.

VOLUMETRIC ARRAY

The voltage response of a continuous volumetric-spherical array of radius R, with weighting dependent only on the distance from the center of the array, is derived in appendix B. The result is given by (B-10) in the form

$$g(u) = \int_0^1 ds\ \left(\frac{2}{\pi}\right)^{1/2} s\ \frac{\sin(us)}{u}\ w(s), \tag{7}$$

where now

$$u \equiv 2\pi\frac{R}{\lambda}\left[2 - 2\cos\phi_\ell \cos\phi_a - 2\sin\phi_\ell \sin\phi_a \cos(\theta_\ell - \theta_a)\right]^{1/2}. \tag{8}$$

The other parameters are as explained in the previous subsection.

Equation (7) has the basic form of a sine transform. As an example, rectangular weighting w yields response g(u) proportional to

$$\frac{\sin u - u \cos u}{u^3} , \tag{9}$$

which has its first few nulls at u = 4.49, 7.73, 10.90.

UNIFIED FORM FOR ARRAY RESPONSES IN DIFFERENT DIMENSIONS

The results in (2), (5), and (7) for the array voltage response in one, two, and three dimensions, respectively, appear to be quite different. However, they can all be written in the basic form

$$g(u) = \int_0^1 ds \, K(u, s) \, w(s), \tag{10}$$

where the kernel

$$K(u, s) = s \left(\frac{s}{u}\right)^\mu J_\mu(us) = \begin{cases} \left(\frac{2}{\pi}\right)^{1/2} \cos(us) & \text{for } \mu = -\frac{1}{2} \\ s \, J_0(us) & \text{for } \mu = 0 \\ \left(\frac{2}{\pi}\right)^{1/2} s \, \frac{\sin(us)}{u} & \text{for } \mu = \frac{1}{2} \end{cases} . \tag{11}$$

Here J_μ is a Bessel function of order μ, and we have used [5; 10.1.1, 10.1.11, 10.1.12]. Thus all the responses are basically Bessel transforms of the normalized weighting w, with the correspondences given in the following table.

Table 1. Identification of Values of μ in (11)

Number of Dimensions	Value of μ
1	-1/2
2	0
3	1/2

If we substitute (11) in (10), we have the explicit result for the response pattern:

$$g(u) = \int\limits_{0}^{1} ds \ s\left(\frac{s}{u}\right)^{\mu} J_{\mu}(us) \ w(s). \tag{12}$$

Inspection of the properties of the Bessel function reveals that $g(u)$ as given by (12) is even in u; see [5; 9.1.10]. Thus we only need to investigate $g(u)$ for $u \geq 0$.

HANKEL TRANSFORM PAIRS

The Hankel transform pair of order μ is given by [6; page 136]

$$f(u) = \int_0^\infty ds \, (us)^{1/2} J_\mu(us) \, F(s) \text{ for } u > 0,$$

$$F(s) = \int_0^\infty du \, (su)^{1/2} J_\mu(su) \, f(u) \text{ for } s > 0. \tag{13}$$

Thus knowledge of either function f or F for positive arguments enables determination of the other function by an integral transform. Under the identification

$$F(s) = s^{\mu+\frac{1}{2}} w(s), \qquad f(u) = u^{\mu+\frac{1}{2}} g(u), \tag{14}$$

(13) takes the form

$$g(u) = \int_0^\infty ds \, s\left(\frac{s}{u}\right)^\mu J_\mu(us) \, w(s) \text{ for } u > 0, \tag{15}$$

$$w(s) = \int_0^\infty du \, u\left(\frac{u}{s}\right)^\mu J_\mu(su) \, g(u) \text{ for } s > 0. \tag{16}$$

Equation (15) is more general than (12) in that it allows for weighting $w(s)$ to be nonzero for $s > 1$. Equation (16) is complementary in that, given a desired pattern $g(u)$, it indicates what weighting $w(s)$ is required for $s > 0$. However, if we attempt to specify some desirable pattern $g(u)$, and then solve (16) for the required weighting $w(s)$, it will generally turn out that the resultant $w(s)$ will be nonzero for $s > 1$. Thus not _any_ pattern $g(u)$ can be selected if we insist on a finite-support weighting $w(s)$; rather, desirable candidate patterns can be substituted in (16) and the corresponding weighting $w(s)$ evaluated to see if it is zero for $s > 1$. If not, the candidate pattern is disallowed and must be modified or discarded.[*] We will use precisely this procedure in a later section when we determine the weightings that realize the ideal patterns in various numbers of dimensions.

---

[*] If pattern $g(u)$ yields $w(s) = 0$ for $s > a$, the scaled pattern $g(u/a)$ yields a modified weighting $a^{2\mu+2} w(as)$, which is zero for $s > 1$, as desired.

DEFINITION OF TWO BESSEL FUNCTION RATIOS

It will be very convenient notationally in the following to employ the shorthand notations

$$\oint_\alpha(z) \equiv \frac{J_\alpha(z)}{z^\alpha} = \frac{1}{2^\alpha} \sum_{k=0}^{\infty} \frac{(-z^2/4)^k}{k!\Gamma(\alpha+1+k)} , \qquad (17)$$

$$\mathcal{Q}_\alpha(z) \equiv \frac{I_\alpha(z)}{z^\alpha} = \frac{1}{2^\alpha} \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!\Gamma(\alpha+1+k)} , \qquad (18)$$

for these two Bessel function ratios; these types of functions have already been encountered in (11), (12), (15), (16). They are extensions of the $\psi_n(z)$ functions discussed in [7; page 56]. Both functions, $\oint_\alpha(z)$ and $\mathcal{Q}_\alpha(z)$, are single-valued and are entire in z for any α, as well as being entire in α for any z [5; 6.1.3, 9.1.1]. Special cases of these functions that are useful here are given by [5; 9.6.6, 10.1.1, 10.1.11, 10.1.12, 10.2.13, 10.2.14]

$$\mathcal{Q}_\alpha(0) = \oint_\alpha(0) = \frac{1}{2^\alpha \Gamma(\alpha+1)} ,$$

$$\mathcal{Q}_0(z) = I_0(z), \mathcal{Q}_1(z) = \frac{I_1(z)}{z}, \mathcal{Q}_{-1}(z) = zI_1(z),$$

$$\mathcal{Q}_{1/2}(z) = \left(\frac{2}{\pi}\right)^{1/2} \frac{\sinh z}{z} , \quad \mathcal{Q}_{-1/2}(z) = \left(\frac{2}{\pi}\right)^{1/2} \cosh z,$$

$$\mathcal{Q}_{3/2}(z) = \left(\frac{2}{\pi}\right)^{1/2} \frac{z \cosh z - \sinh z}{z^3} , \quad \mathcal{Q}_{-3/2}(z) = \left(\frac{2}{\pi}\right)^{1/2} (z \sinh z - \cosh z) ,$$

$$\mathcal{Q}_{5/2}(z) = \left(\frac{2}{\pi}\right)^{1/2} \frac{(3+z^2) \sinh z - 3z \cosh z}{z^5} ,$$

$$\mathcal{Q}_{-5/2}(z) = \left(\frac{2}{\pi}\right)^{1/2} [(3+z^2) \cosh z - 3z \sinh z]. \qquad (19)$$

A useful property, which is obvious from (17) and (18) and which will find frequent application here, is

$$\mathcal{J}_\alpha(\pm i z) = \mathcal{U}_\alpha(z), \quad \mathcal{U}_\alpha(\pm i z) = \mathcal{J}_\alpha(z). \tag{20}$$

These relations hold true for all values of z, real or complex. Properties similar to (19) can be obtained for the $\mathcal{J}_\alpha(z)$ functions; for example,

$$\mathcal{J}_{-1/2}(z) = \mathcal{U}_{-1/2}(\pm i z) = \left(\frac{2}{\pi}\right)^{1/2} \cosh(\pm i z) = \left(\frac{2}{\pi}\right)^{1/2} \cos(z).$$

Other useful properties follow from the use of [5; 9.1.30 and 9.6.28],

$$\mathcal{J}'_\alpha(z) = -z \mathcal{J}_{\alpha+1}(z), \quad \mathcal{U}'_\alpha(z) = z \mathcal{U}_{\alpha+1}(z), \tag{21}$$

and from [5; 9.1.27 and 9.6.26],

$$\mathcal{U}_\alpha(z) = \frac{1}{z^2}\left[\mathcal{U}_{\alpha-2}(z) - 2(\alpha-1)\mathcal{U}_{\alpha-1}(z)\right],$$

$$\mathcal{J}_\alpha(z) = -\frac{1}{z^2}\left[\mathcal{J}_{\alpha-2}(z) - 2(\alpha-1)\mathcal{J}_{\alpha-1}(z)\right].$$

A CLASS OF BESSEL WEIGHTINGS

The class of Bessel weightings that we are suggesting, regardless of the number of dimensions of the array, is given by

$$w(s) = \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) \text{ for } 0 < s < 1, \quad \text{where } \nu > -1, B > 0, \quad (22)$$

and $w(s) = 0$ for $s > 1$. There are two real parameters in this class, namely, ν and B. For the special case of $\nu = 0$, this weighting is already known as Kaiser-Bessel [3,4]; its pattern has nearly the optimum energy content within a specified bandwidth in the one-dimensional application to spectral analysis.

Substitution of (22) in (12) or (15) yields the closed form for the pattern [6; page 30, fourth integral, and 5; 9.6.3]

$$g(u) = \int_{0}^{1} ds \ s\left(\frac{s}{u}\right)^{\mu} J_{\mu}(us) \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) \quad (23A)$$

$$= \oint_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right) = \mathcal{U}_{\mu+\nu+1}\left(\sqrt{B^2-u^2}\right) \text{ for all } u, \text{ if } \mu > -1 \text{ and } \nu > -1. \quad (23B)$$

Here we have employed definitions (17) and (18) and used (20). The condition on μ guarantees convergence of integral (12) at $s = 0$, whereas the condition on ν guarantees convergence of integral (12) at $s = 1$. The first form of (23B) is more convenient computationally for $u \geq B$, whereas the second form is more convenient for $B \geq u$.

Weighting-pattern pair (22)-(23) are the fundamental results for the class of Bessel weightings under consideration. They apply to one-, two-, or three-dimensional arrays when μ is specialized to $-\frac{1}{2}$, 0, or $\frac{1}{2}$, respectively, and when u is interpreted as (4), (6), or (8), respectively. The parameter B is nonnegative and will be shown to control the ratio of mainlobe to peak sidelobe. For the case of one dimension, $\mu = -\frac{1}{2}$, the kernel of transform (23A) is a cosine (see the top line of (11)); in this special case, the pattern $\oint_{\nu+\frac{1}{2}}\left(\sqrt{u^2-B^2}\right)$ was also independently and simultaneously discovered by Roy Streit [8].

WEIGHTING CHARACTERISTICS

Weighting (22) is positive for $0 \leq s < 1$ since $\nu > -1$ and $B \geq 0$; see [5; 9.6.10]. In addition, it is zero and therefore continuous at $s = 1$ if $\nu > 0$. In fact [5; 9.6.7],

$$w(s) \sim \frac{(1-s)^\nu}{\Gamma(\nu+1)} \quad \text{as } s \to 1-. \tag{24}$$

For $\nu \geq 0$, weighting $w(s)$ is monotonically decreasing in s on $(0,1)$; see appendix C. However, for $-1 < \nu < 0$, $w(s)$ possesses an integrable singularity at $s = 1$.

Examples of weighting (22) are plotted in figures 1-6 for $\nu = 2, 1.5, 1, .5, 0, -.5$, respectively. Figure 5, for $\nu = 0$, corresponds to the Kaiser-Bessel weighting [3]. For the larger values of ν, the weighting blends smoothly to zero at $s = 1$, but for the smaller values of ν, the behavior of $w(s)$ is more irregular at $s = 1$, being discontinuous for $\nu = 0$ and infinite for $\nu < 0$. The larger values of B lead to smoother functions that are Gaussian-like; in fact, for $s < 1$ [5; 9.7.1],

$$w(s) \approx \frac{\exp(B)}{(2\pi B)^{1/2} B^\nu} \exp\left(-\frac{1}{2}B^2 s^2\right) \quad \text{as } B \to \infty. \tag{25}$$

More generally,

$$w(s) \approx (2\pi)^{-\frac{1}{2}} B^{-2\nu} \left(B\sqrt{1-s^2}\right)^{\nu-\frac{1}{2}} \exp\left(B\sqrt{1-s^2}\right) \quad \text{as } B\sqrt{1-s^2} \to \infty. \tag{26}$$

The opposite limit for small B is

$$w(s) = \left(\frac{1-s^2}{2}\right)^\nu / \Gamma(\nu+1) \quad \text{for } B = 0, \tag{27}$$

for which $g(u) = g_{\nu+1}(u)$.
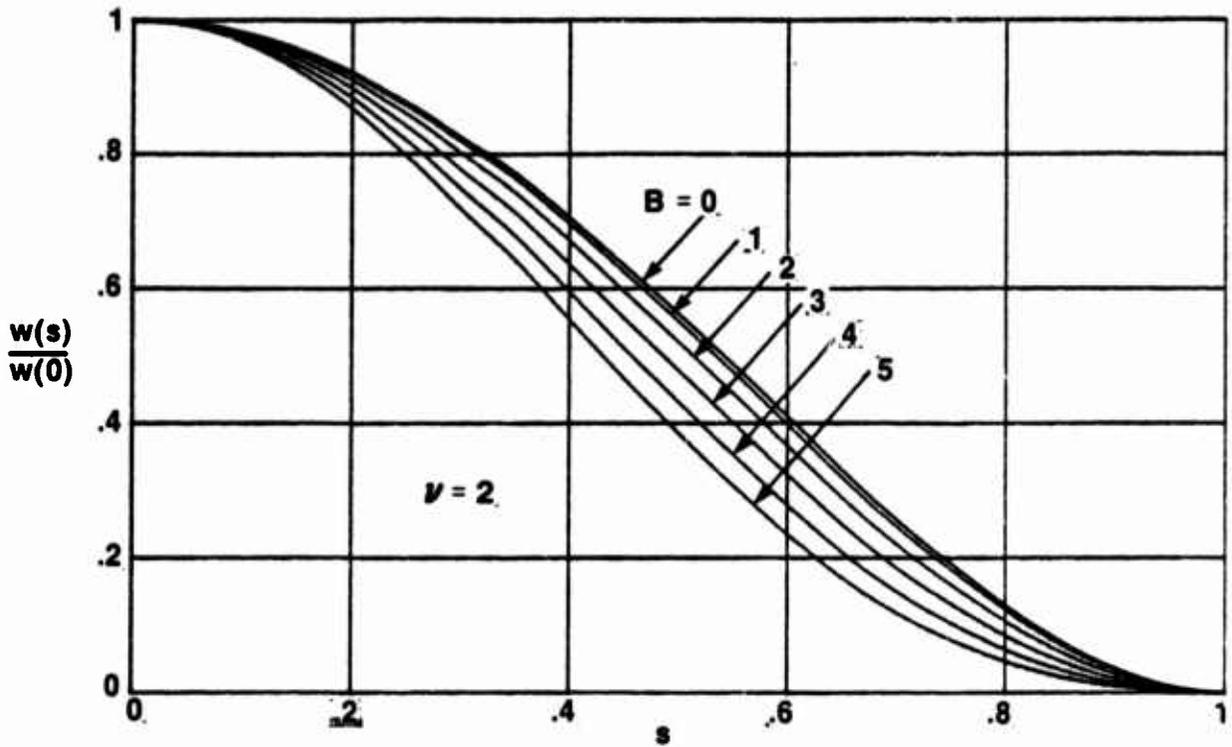
Figure 1. Normalized Weighting for $\nu = 2$
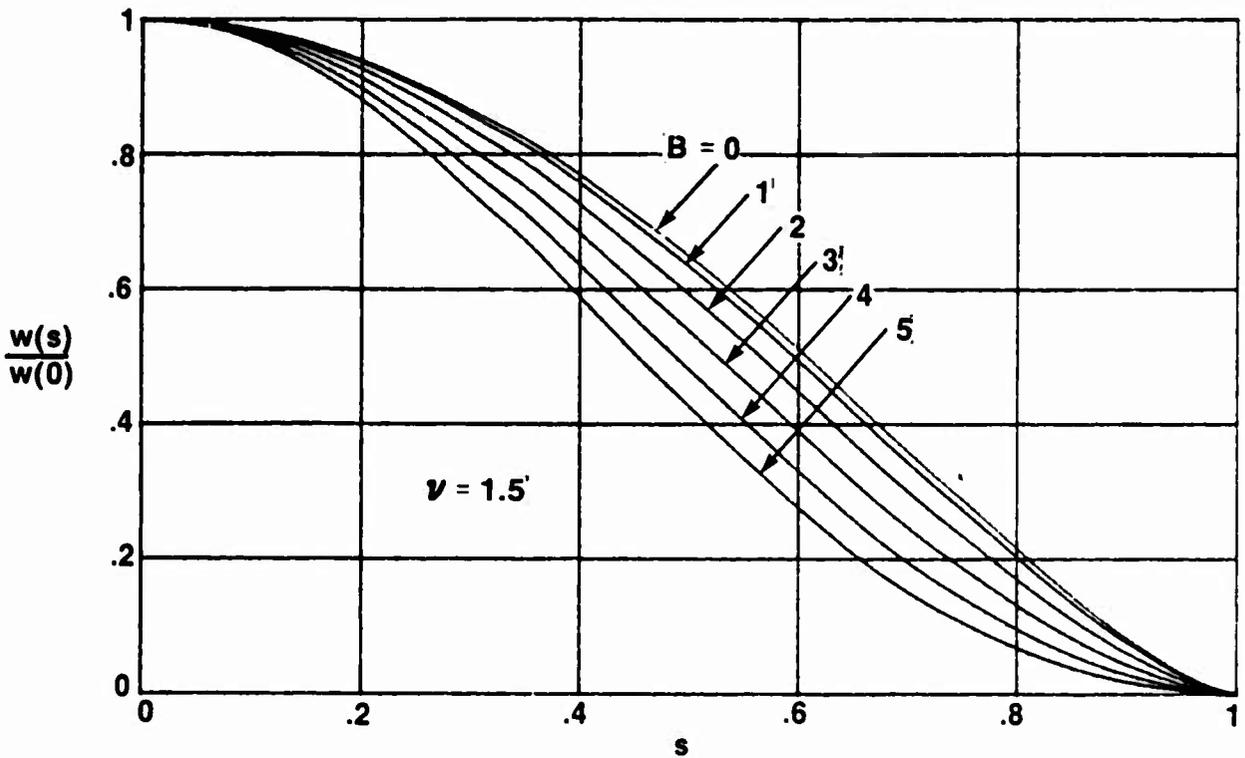
Figure 2. Normalized Weighting for $\nu = 1.5$
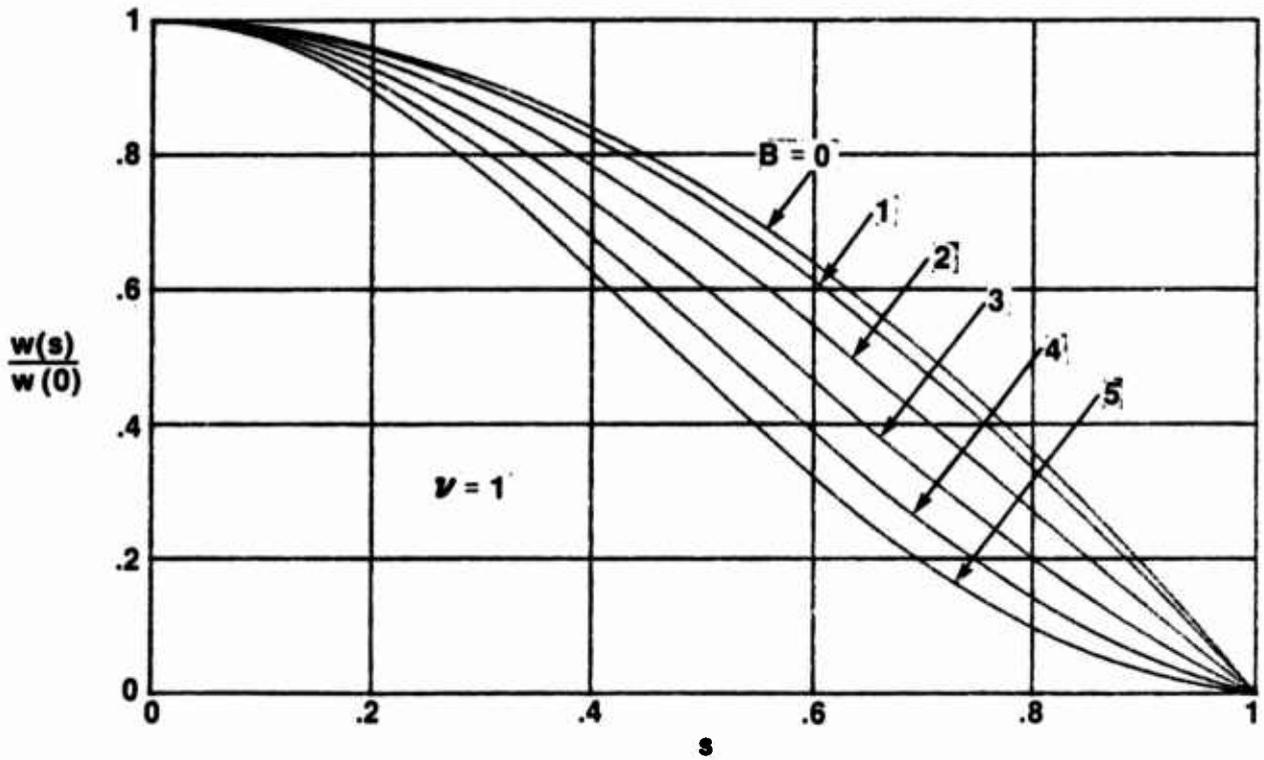
Figure 3. Normalized Weighting for $\nu = 1$
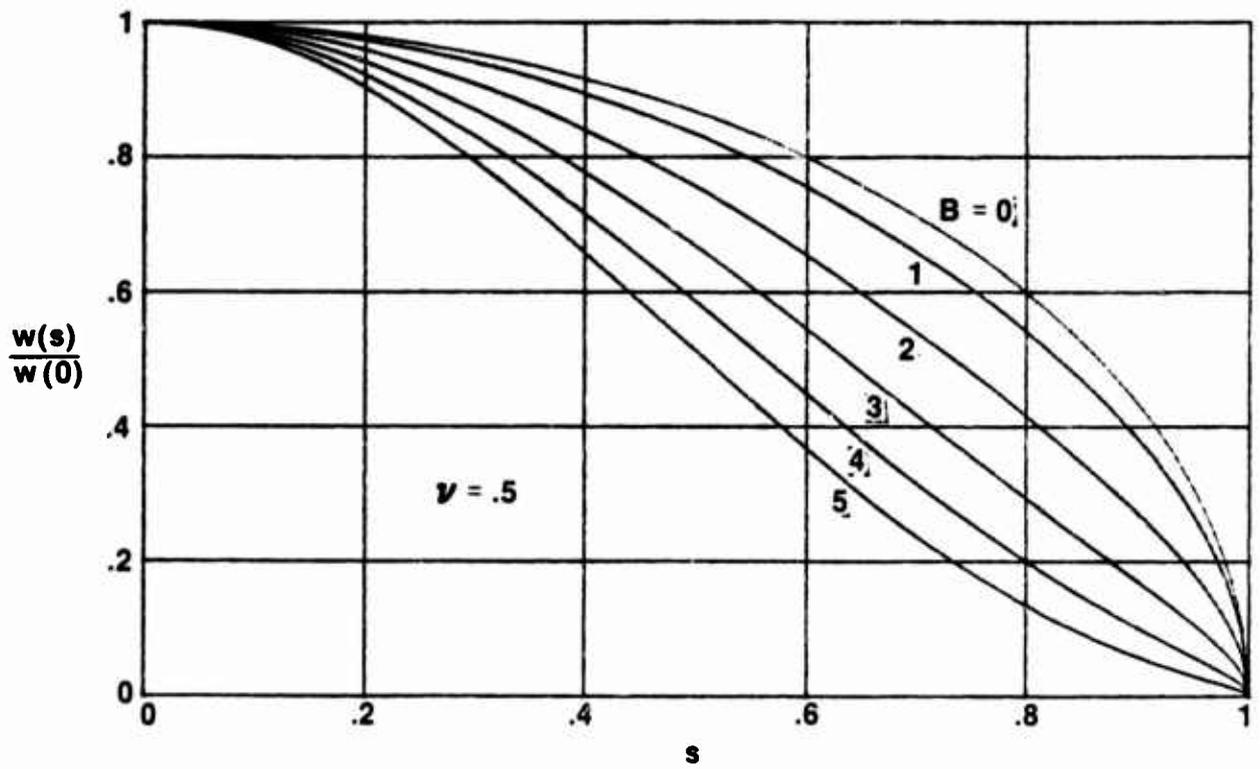
Figure 4. Normalized Weighting for $\nu = .5$
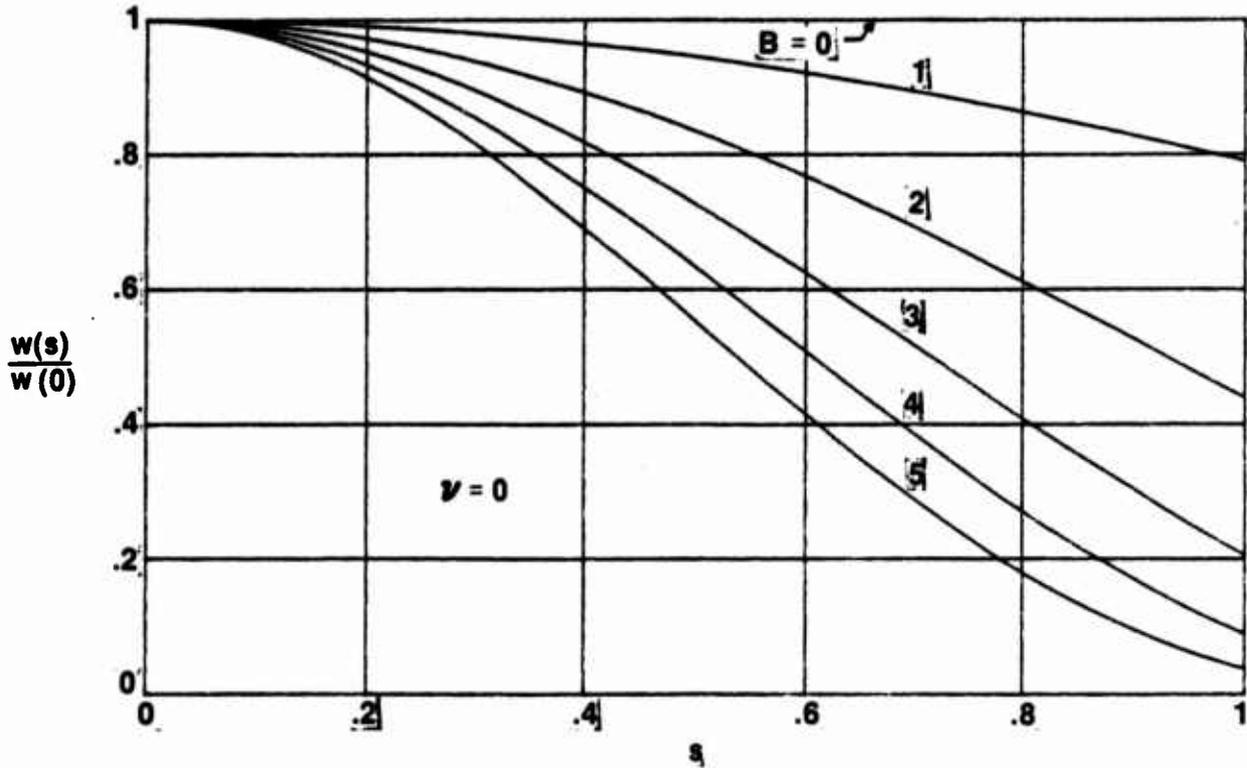
Figure 5.   Normalized Weighting for $\nu = 0$
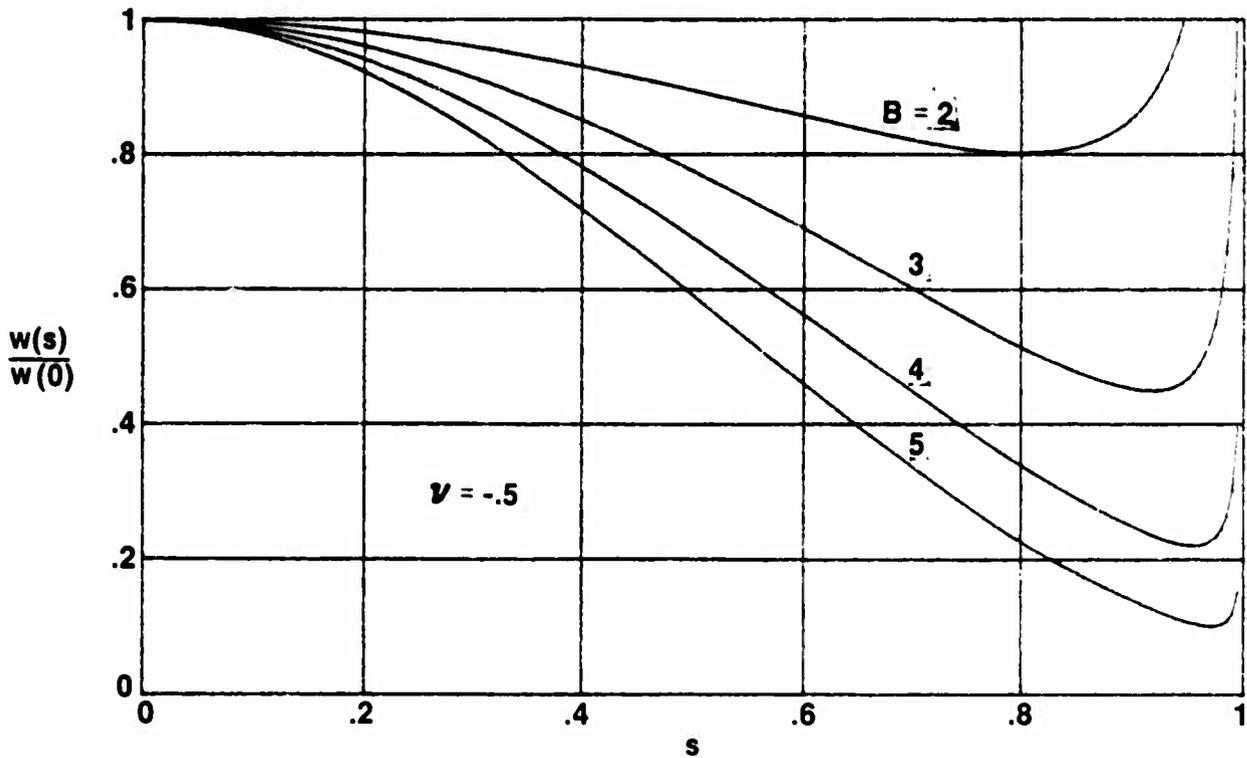
Figure 6.   Normalized Weighting for $\nu = -.5$

13

RESPONSE PATTERN CHARACTERISTICS

The pattern was given by (23) as

$$g(u) = \mathcal{J}_\alpha\left(\sqrt{u^2 - B^2}\right) = \mathcal{J}_\alpha\left(\sqrt{B^2 - u^2}\right), \tag{28}$$

where the composite order of the Bessel function is

$$\alpha = \mu + \nu + 1 . \tag{29}$$

The asymptotic behavior of the pattern (28) for large u is available from (17) and [5; 9.2.1]:

$$g(u) \sim \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{\cos\left(u - \frac{1}{2}\alpha\pi - \frac{1}{4}\pi\right)}{u^{\alpha + \frac{1}{2}}} \quad \text{as } u \to \infty. \tag{30}$$

Since g is proportional to the array voltage response, (30) corresponds to a

$$\text{decay} \sim 6\alpha + 3 \text{ dB/octave} \quad \text{as } u \to \infty. \tag{31}$$

Expressed in terms of the original dimension-parameter μ and weighting-parameter ν, this is, from (29),

$$\begin{aligned}
\text{decay} &\sim 6\mu + 6\nu + 9 \text{ dB/octave} \quad \text{as } u \to \infty \\
&= \begin{cases} 6\nu + 6 \text{ dB/octave for one dimension} \\ 6\nu + 9 \text{ dB/octave for two dimensions} \\ 6\nu + 12 \text{ dB/octave for three dimensions} \end{cases} .
\end{aligned} \tag{32}$$

Thus the greater the number of dimensions, the faster is the rate of decay of the sidelobes of the response, for a common value of weighting-parameter ν. Each additional dimension adds a 3 dB/octave decay, for a fixed ν.

Special cases of the pattern (28) are available upon use of (19) and (20); for example,

$$
g(u) = \begin{cases}
\left(\frac{2}{\pi}\right)^{1/2} \cosh\left(\sqrt{B^2-u^2}\right) = \left(\frac{2}{\pi}\right)^{1/2} \cos\left(\sqrt{u^2-B^2}\right) & \text{for } \alpha = -\frac{1}{2} \\[2ex]
I_0\left(\sqrt{B^2-u^2}\right) = J_0\left(\sqrt{u^2-B^2}\right) & \text{for } \alpha = 0 \\[2ex]
\left(\frac{2}{\pi}\right)^{1/2} \frac{\sinh\left(\sqrt{B^2-u^2}\right)}{\sqrt{B^2-u^2}} = \left(\frac{2}{\pi}\right)^{1/2} \frac{\sin\left(\sqrt{u^2-B^2}\right)}{\sqrt{u^2-B^2}} & \text{for } \alpha = \frac{1}{2} \\[2ex]
\frac{I_1\left(\sqrt{B^2-u^2}\right)}{\sqrt{B^2-u^2}} = \frac{J_1\left(\sqrt{u^2-B^2}\right)}{\sqrt{u^2-B^2}} & \text{for } \alpha = 1
\end{cases} \qquad (33)
$$

All of these relations are valid for all u, whether u is larger or smaller than B; of course, the former form in each line is more convenient for $u \lesssim B$, while the latter is more convenient for $u \geq B$. The third result in (33), for $\alpha = \frac{1}{2}$, includes the pattern in one dimension ($\mu = -\frac{1}{2}$, $\nu = 0$) for the Kaiser-Bessel weighting $I_0\left(B\sqrt{1-s^2}\right)$ for $0 < s < 1$.

The special case of $\alpha = -\frac{1}{2}$ in (33) deserves extra attention; this case will be called the ideal pattern:

$$
g_i(u) = \mathcal{I}_{-\frac{1}{2}}\left(\sqrt{B^2-u^2}\right) = \mathcal{J}_{-\frac{1}{2}}\left(\sqrt{u^2-B^2}\right)
$$

$$
= \left(\frac{2}{\pi}\right)^{1/2} \cosh\left(\sqrt{B^2-u^2}\right) = \left(\frac{2}{\pi}\right)^{1/2} \cos\left(\sqrt{u^2-B^2}\right) \quad \text{for } \alpha = -\frac{1}{2}. \quad (34)
$$

The plot in figure 7 reveals that the sidelobes are all equal, and that the

$\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\cosh{(B)}$

$$g_i(u) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\cosh\left(\sqrt{B^2 - u^2}\right) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\cos\left(\sqrt{u^2 - B^2}\right)$$

$\left(\frac{2}{\pi}\right)^{\frac{1}{2}}$

0

B

0   π 2π 3π 4π 5π 6π

u
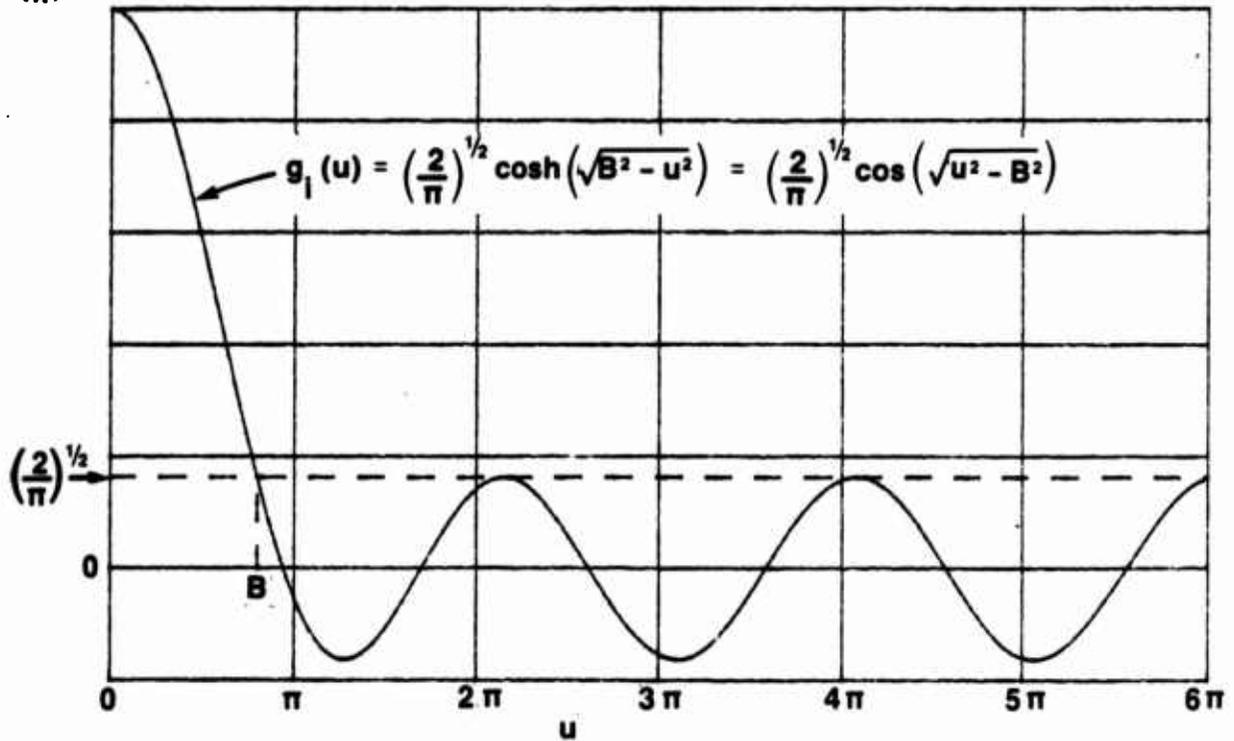
Figure 7. Ideal Pattern $g_i(u)$

voltage ratio

$$\frac{\text{mainlobe level}}{\text{sidelobe level}} = \cosh(B). \tag{35}$$

The mainlobe width, as measured to the point where the mainlobe first decays to the eventual sidelobe level, is

$$\text{mainlobe width} = B. \tag{36}$$

The abscissa u is given by (4), (6), or (8) for one, two, or three dimensions, respectively. Determination of the required weighting to realize the ideal pattern (34) in different numbers of dimensions is taken up in a later section.

16

If $\alpha < -\frac{1}{2}$, the pattern (28) has increasing sidelobes as u increases; see (30). Therefore, $\alpha \geq -\frac{1}{2}$ are the only cases of practical interest for pattern (28).

Plots of pattern (28) are given in dB in figures 8-13 for α = 2, 1.5, 1, .5, 0, -.5, respectively, for various values of B. The program is listed in appendix D. The larger values of α realize the more rapid decay of sidelobes, but, on the other hand, have wider mainlobes. Figures 8-13 indicate the necessary tradeoffs between mainlobe width, sidelobe decay, and mainlobe-to-sidelobe ratio that must be considered in any weighting selection.

A small chart in the upper right quadrant of each figure indicates some allowable values of μ and ν that apply to that figure. For example, in figure 8, the pattern for α = 2 applies to all the following:

$$\mu = -\frac{1}{2} \text{ (one dimension)} \quad \text{with } \nu = \frac{3}{2} ,$$
$$\text{or}$$
$$\mu = 0 \text{ (two dimensions)} \quad \text{with } \nu = 1,$$
$$\text{or}$$
$$\mu = \frac{1}{2} \text{ (three dimensions) with } \nu = \frac{1}{2} . \tag{37}$$

When we come to figure 11, for α = .5, however, the case of $\mu = \frac{1}{2}$, ν = -1 has an asterisk next to the ν = -1 entry. The reason for this is that the integral (23) leading to pattern g(u) was convergent only for $\nu > -1$, and now we are trying to violate that condition. A similar cautionary note is indicated in figures 12 and 13; in fact, all three cases in figure 13 violate the condition $\nu > -1$. Despite this preclusion, we shall find later that the required weighting does, in fact, have the form (22) for the corresponding ν values given in figures 11 - 13, but requires generalized functions with a singularity at the endpoint s = 1 of the interval. This extension to $\nu \leq -1$ is desirable and important because realization of the ideal pattern (figures 7 and 13) requires values for ν in this region, regardless of the number of dimensions.

17
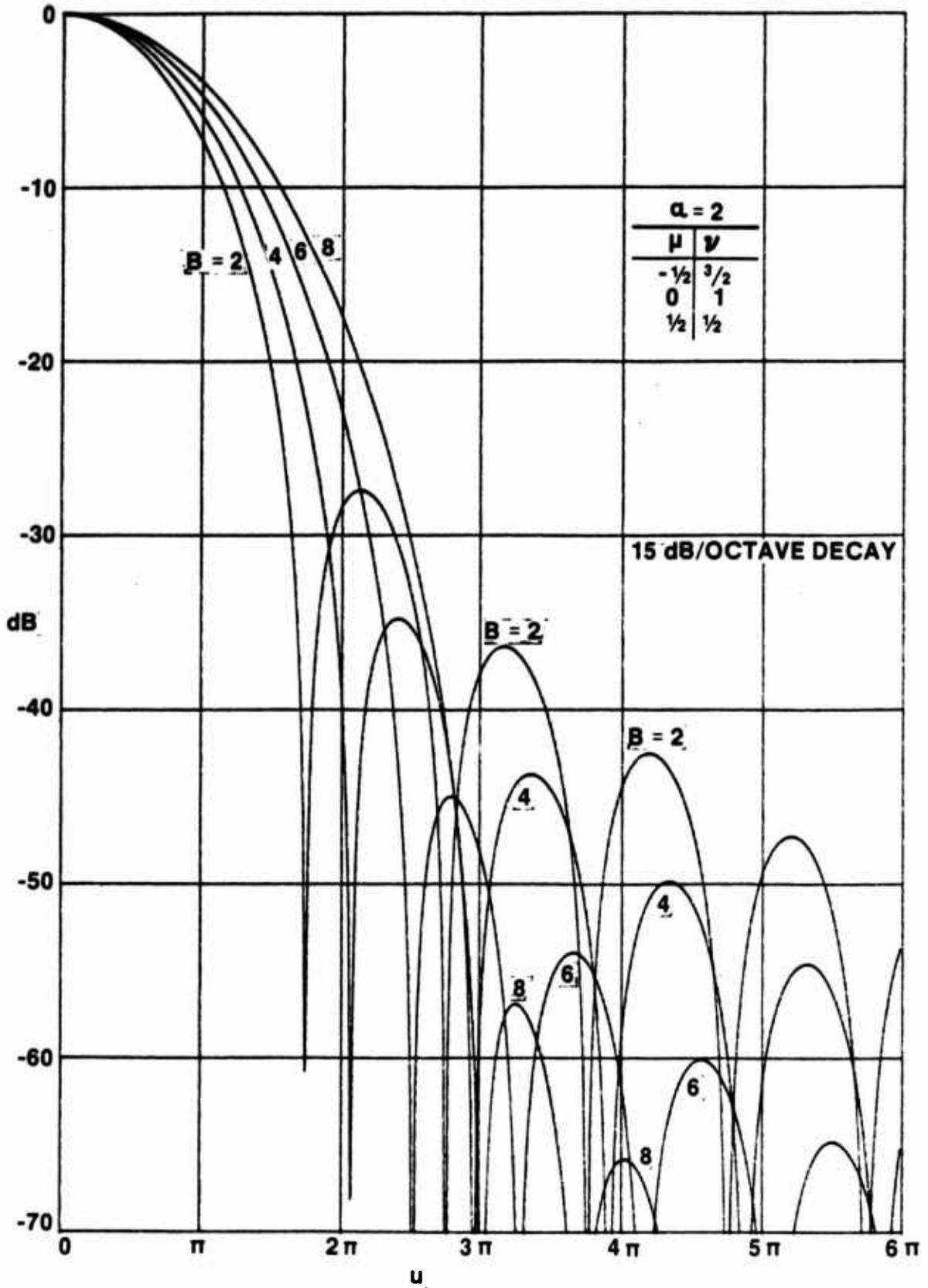
Figure 8.   Pattern in dB for α = 2
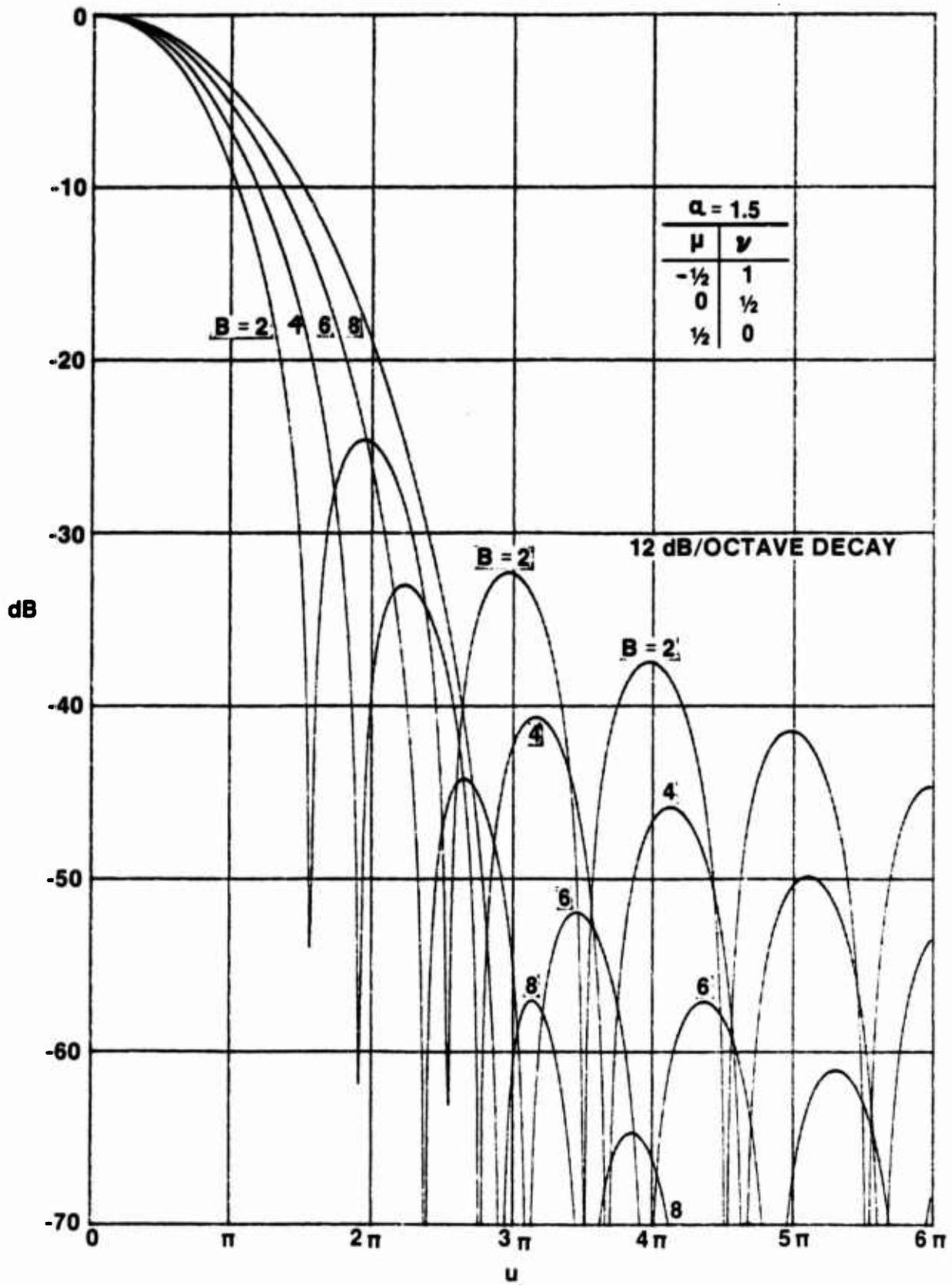
Figure 9. Pattern in dB for α = 1.5
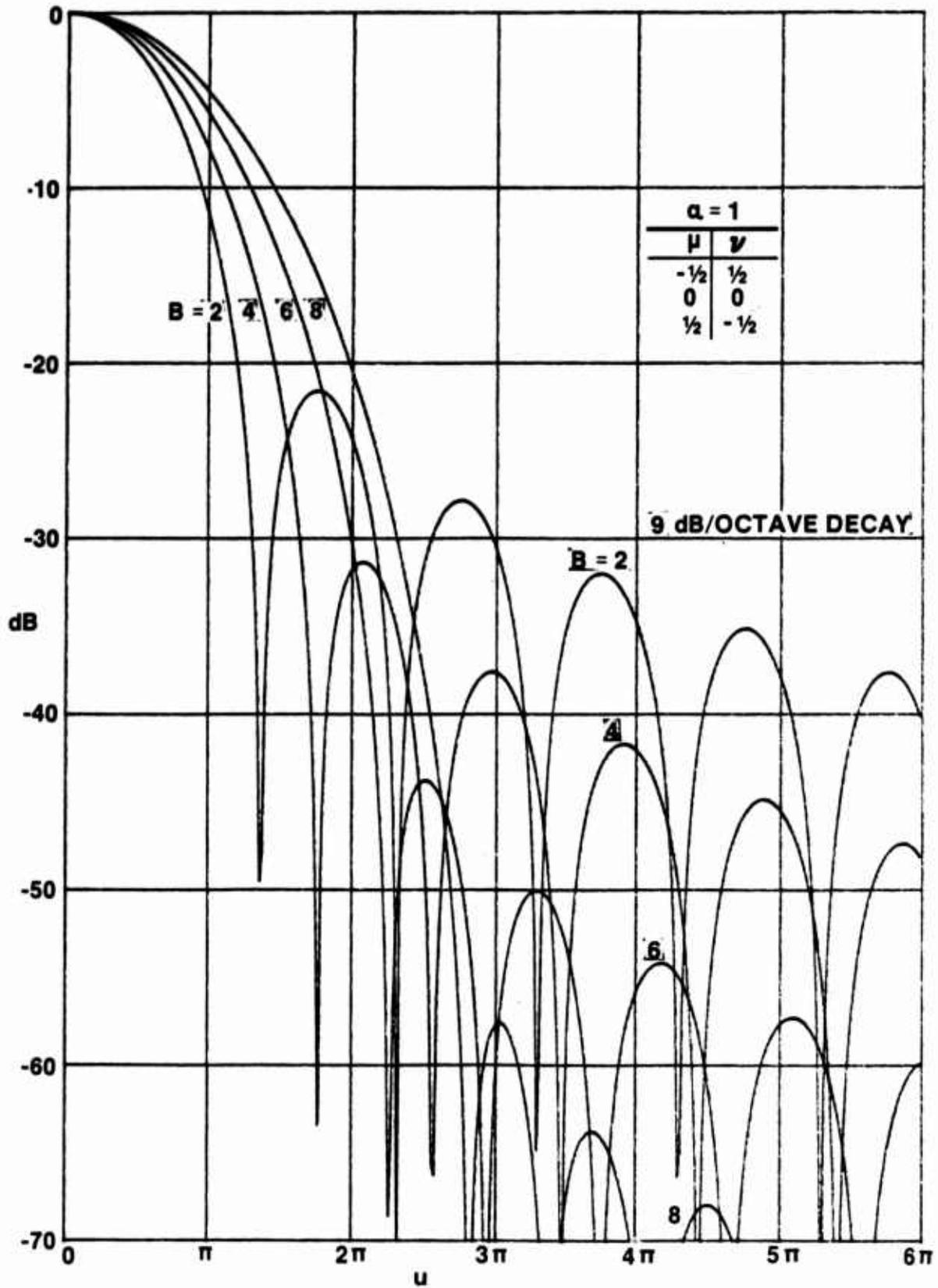
Figure 10. Pattern in dB for $\alpha = 1$
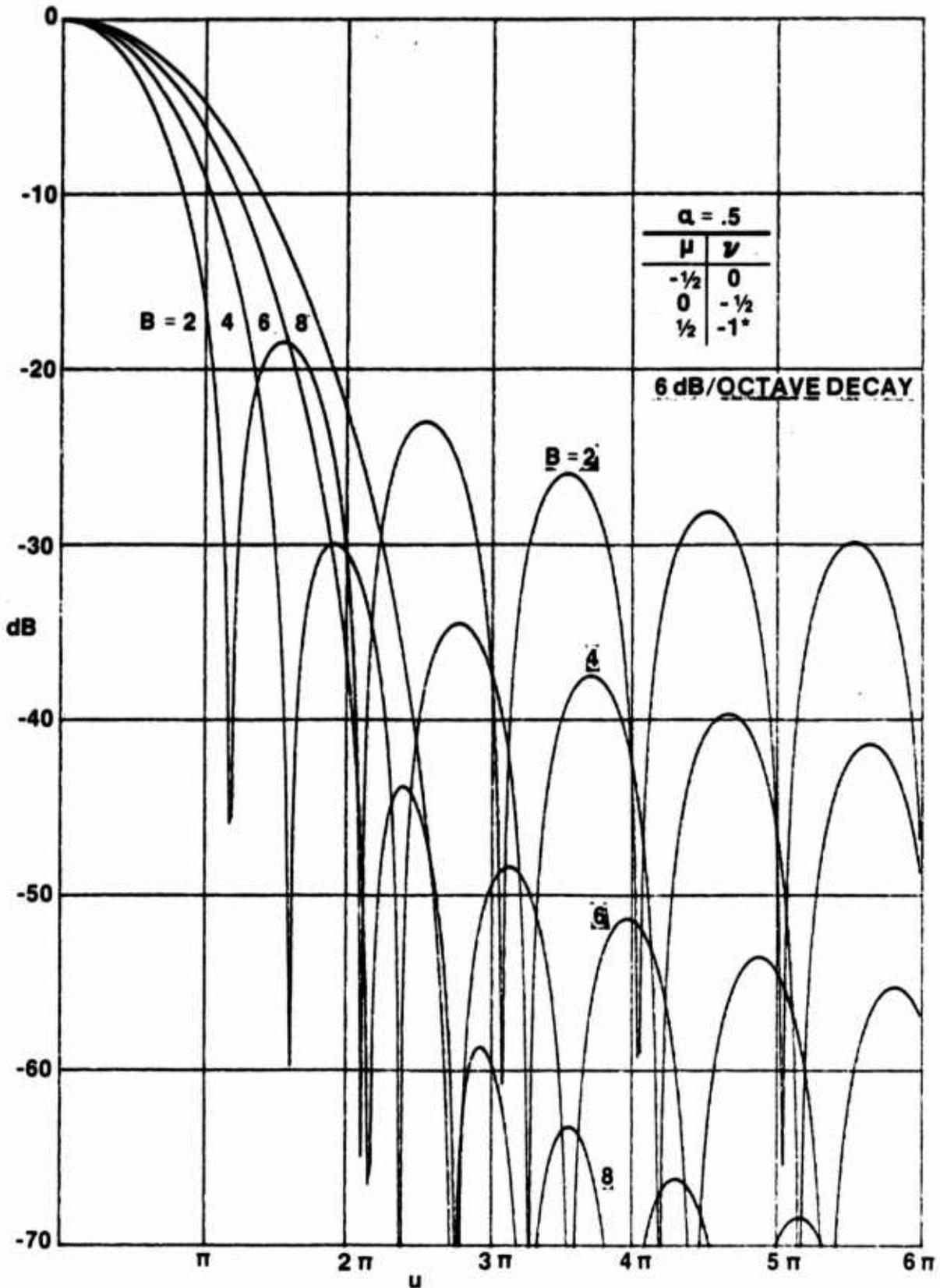
Figure 11.   Pattern in dB for α = .5
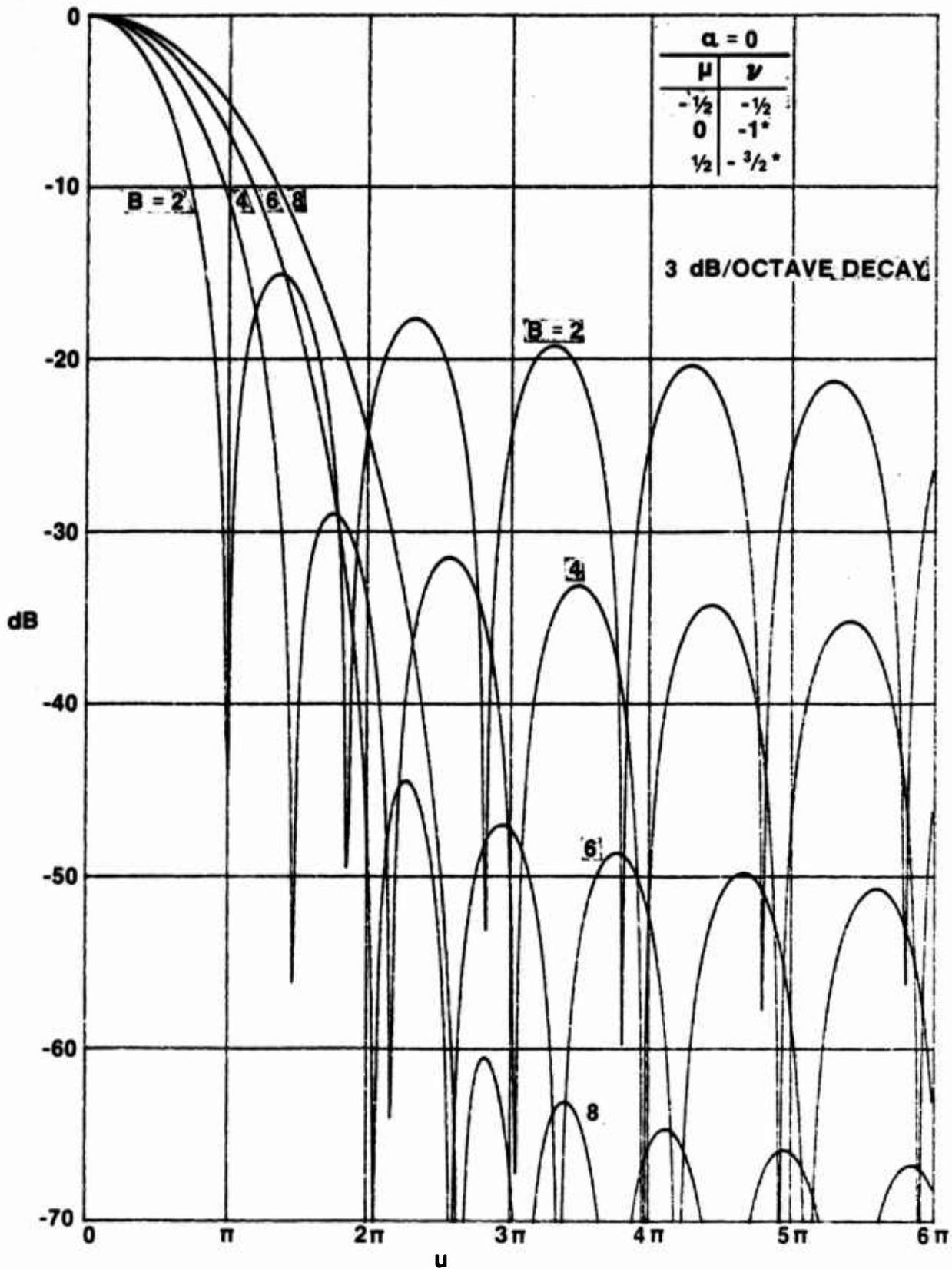
Figure 12. Pattern in dB for $\alpha = 0$
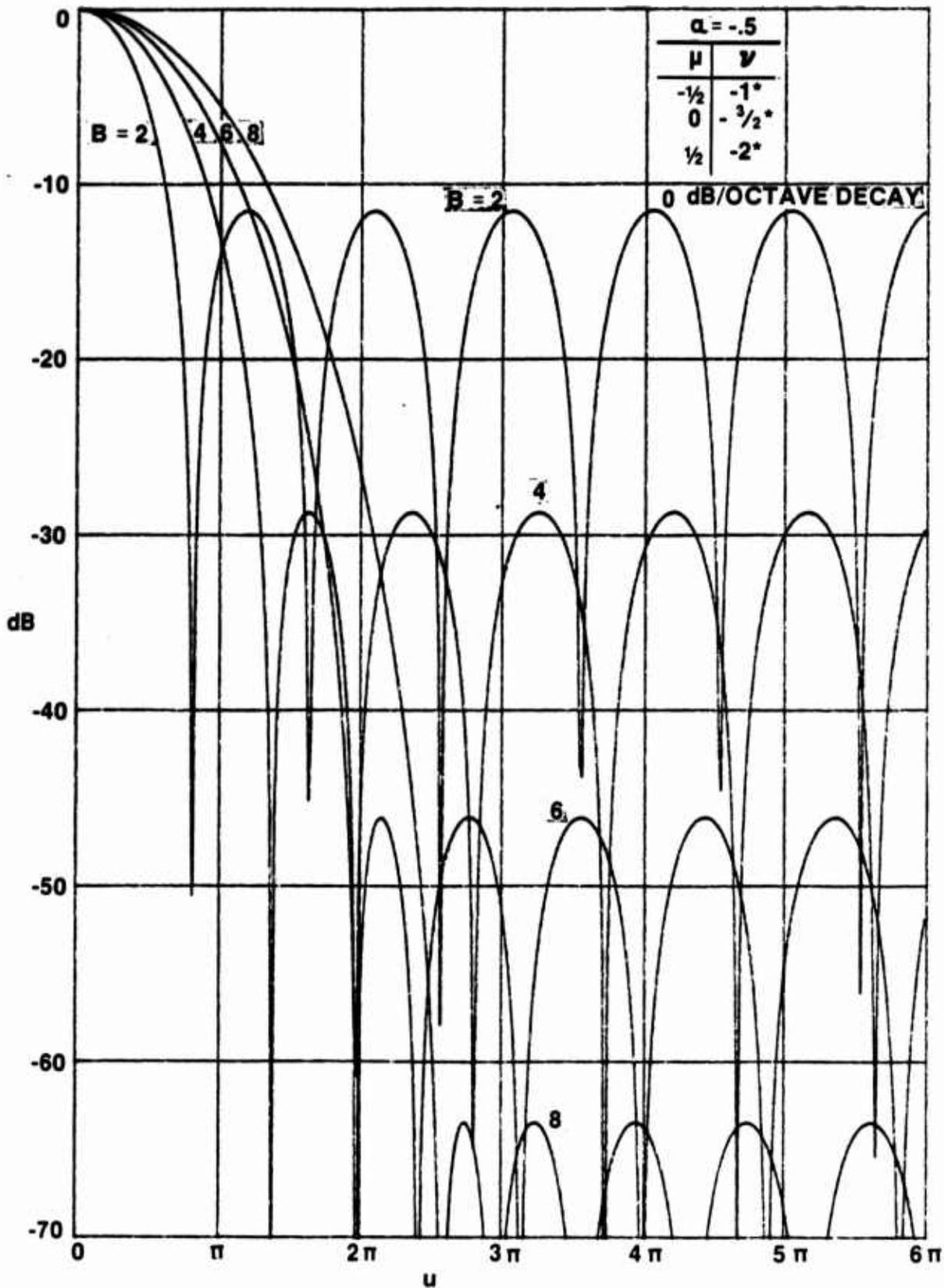
Figure 13. Pattern in dB for $\alpha = -.5$

FIRST NULL LOCATION

Let z_α be the smallest nonzero null location of $J_\alpha(z)$; i.e.,

$$J_\alpha(z_\alpha) = 0. \tag{38}$$

A short list of $\{z_\alpha\}$ is given below in table 2. Then, by use of (17), the

Table 2. First Zero of $J_\alpha(z)$

α	z_α
-0.5	1.5708
0.0	2.4048
0.5	3.1416
1.0	3.8317
1.5	4.4934
2.0	5.1356
2.5	5.7635
3.0	6.3802
3.5	6.9879
4.0	7.5883
4.5	8.1826

first null location of pattern $g(u)$ in (28) is at u_0, where

$$\left(u_0^2 - B^2\right)^{1/2} = z_\alpha, \qquad u_0 = \left(B^2 + z_\alpha^2\right)^{1/2} . \tag{39}$$

The results in figure 14 display the first null location as a function of
B, for various values of α. For large B, u_0 behaves as $B + \frac{1}{2}z_\alpha^2/B$. By the
identification of α as $\mu + \nu + 1$, this curve applies to any number of
dimensions and to whatever value of ν is selected in weighting $w(s)$ of (22).
The curves indicate that the first null location u_0 is monotonically
increasing in both B and α.
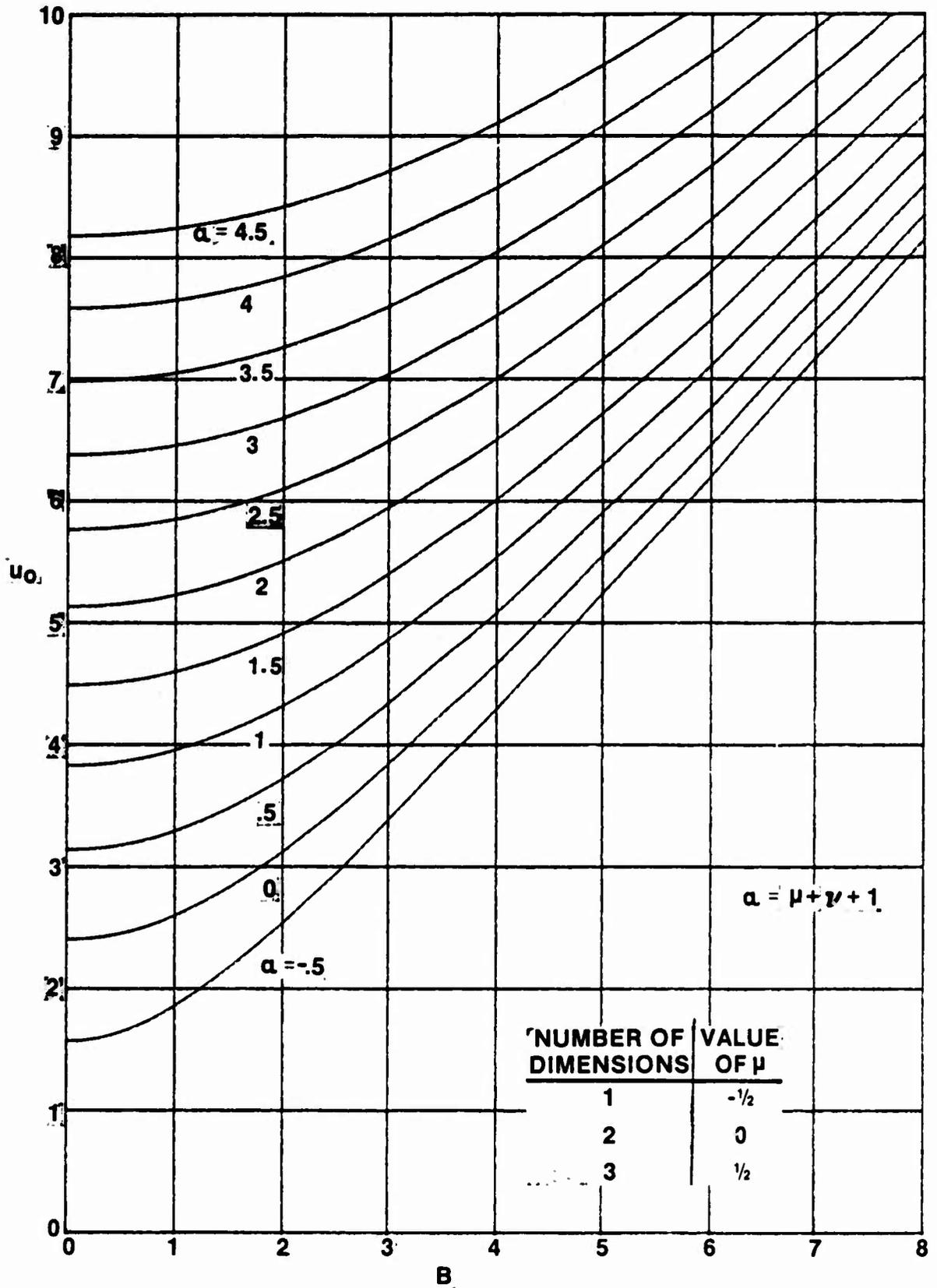
Figure 14. First Null Location of Pattern $g(u) = J_\alpha\left(\sqrt{u^2 - B^2}\right)$.

LOCATION OF PEAK OF FIRST SIDELOBE

By use of (28) and (21), we obtain derivative

$$g'(u) = \oint_\alpha' \left(\sqrt{u^2-B^2}\right) \frac{u}{\sqrt{u^2-B^2}} = -u \oint_{\alpha+1} \left(\sqrt{u^2-B^2}\right). \tag{40}$$

Therefore, reference to (38) reveals that the location of the peak of the first sidelobe of g(u) occurs where $g'(u_p) = 0$; i.e,

$$\left(u_p^2-B^2\right)^{1/2} = z_{\alpha+1}, \qquad u_p = \left(B^2+z_{\alpha+1}^2\right)^{1/2}. \tag{41}$$

If we employ more explicit notation in (39) and (41), we can express the first peak location in terms of the first null location according to

$$u_p(B,\alpha) = u_0(B,\alpha+1). \tag{42}$$

Thus all the results in figure 14 can be applied directly to the first peak location. For example, (42) yields

$$u_p(B, -\tfrac{1}{2}) = u_0(B, \tfrac{1}{2}); \tag{43}$$

thus the third curve from the bottom in figure 14 gives the location of the peak of the first sidelobe when $\alpha = -\dfrac{1}{2}$

PEAK SIDELOBE LEVEL

The value of voltage pattern g(u) at location (41) gives the level of the peak sidelobe:

$$g(u_p) = \oint_\alpha \left(\sqrt{u_p^2-B^2}\right) = \oint_\alpha \left(z_{\alpha+1}\right). \tag{44}$$

Since the origin value of the pattern is

$$g(0) = \ell_\alpha(\beta),\tag{45}$$

the voltage ratio of peak sidelobe level to mainlobe height is

$$\frac{g(u_p)}{g(0)} = \frac{\ell_\alpha(z_{\alpha+1})}{\ell_\alpha(B)}.\tag{46}$$

Relation (46) is plotted in dB in figure 15 as a function of β, for various values of α; i.e.,

$$dB = 10 \log\left[\frac{\ell_\alpha(z_{\alpha+1})}{\ell_\alpha(B)}\right]^2.\tag{47}$$

The peak sidelobe level decreases monotonically with increasing β, but has no simple behavior versus α except for very small B or very large B.

The results of these last two figures are combined in figure 16, where we plot the peak sidelobe level in dB versus the first null location u_0. These latter curves are virtually linear over a wide range. If we disregard the sidelobe decay rate, the most desirable region of this figure is in the lower left quadrant, i.e., small u_0 and very negative dB. However, the closest we can get (from our family) is the $\alpha = -.5$ curve, which is, in fact, the ideal pattern; see (34) and figure 7. Furthermore, the sidelobe decay rate is then 0 dB/octave. Higher sidelobe decay rates are attained by moving toward the upper right quadrant of the figure; for example, the $\alpha = 3.5$ curve has a $6\alpha+3$ = 24 dB/octave sidelobe decay rate. This figure furnishes a very compact display of the important interrelationships that occur between the fundamental features of peak sidelobe level, mainlobe width, and sidelobe decay rate, and allows for a quick tradeoff comparison of alternatives.
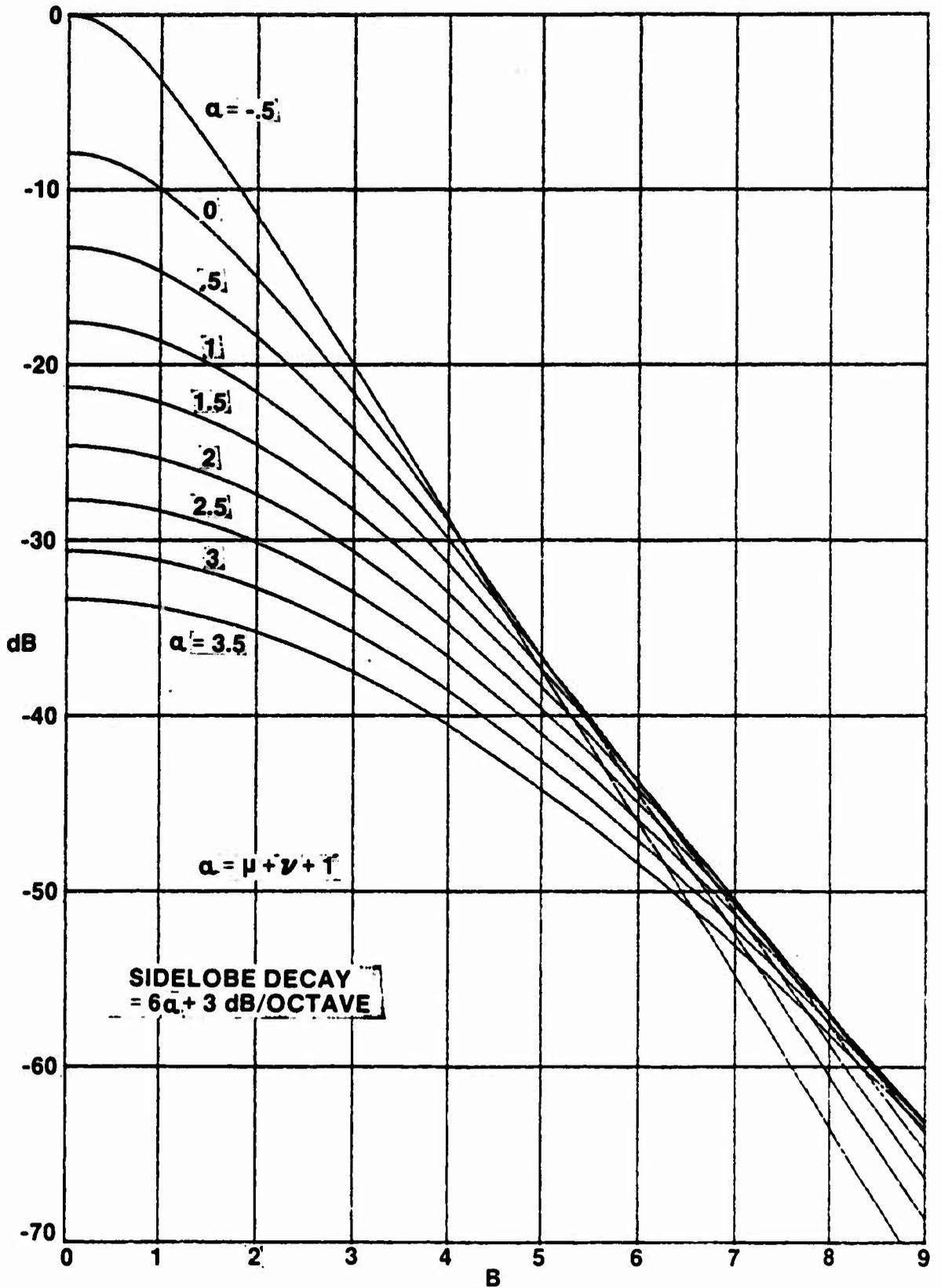
Figure 15. Peak Sidelobe Level of Pattern $g(u) = \mathcal{J}_\alpha \left(\sqrt{u^2 - B^2} \right)$
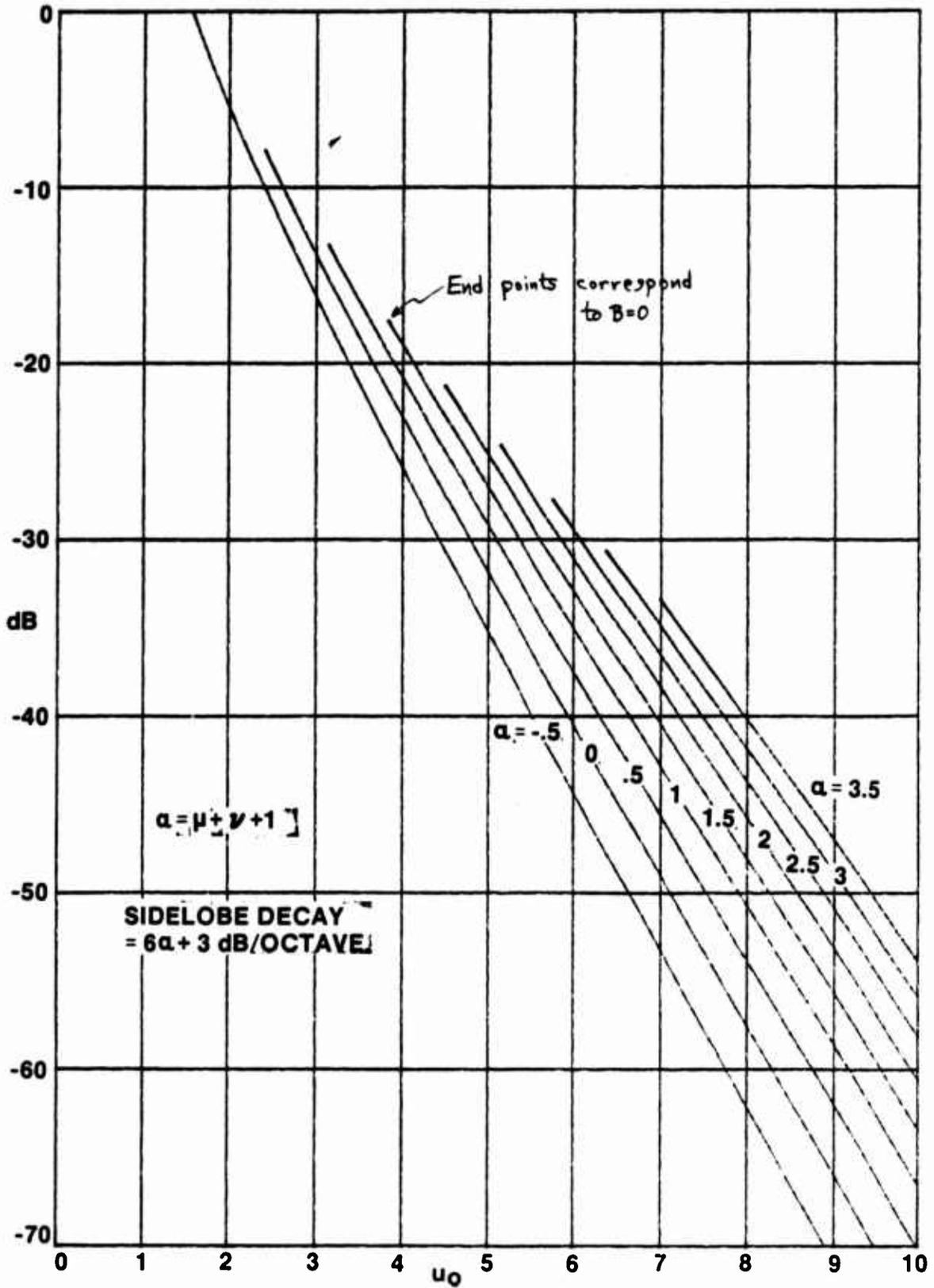
Figure 16. Peak Sidelobe Level Versus First Null Location of Pattern

$$g(u) = \mathcal{J}_\alpha \left(\sqrt{u^2 - B^2} \right)$$

IDEAL WEIGHTING AND PATTERN

It was shown in figure 7, in the previous section, that the pattern g(u) for $\alpha = -\frac{1}{2}$ takes on a particularly desirable behavior, namely, a narrow mainlobe width and a large ratio of mainlobe-to-peak-sidelobe. However, figures 11 - 13 indicated that realization of some patterns was apparently not possible in certain dimensions because we were violating the condition on parameter ν in weighting (22) that allowed for convergence of integral (23). Here we will address the more general problem of how to realize pattern (23),

$$g(u) = \mathcal{J}_{\mu+\nu+1}\left(\sqrt[4]{u^2-B^2}\right) \text{ for all } u, \tag{48}$$

for any $\mu > -1$, but without the current restriction of $\nu > -1$ in weighting (22). This procedure will of course require a different and more general weighting than (22), and will furnish solutions to the asterisked cases in figures 11 - 13.

The solution for the required weightings to realize (48) for any $\mu > -1$ is conducted in appendix E. All the weightings are zero for $s > 1$, as desired; their values for $0 \leq s \leq 1$ are listed below. From (E-14),

$$w(s) = \frac{B}{(1-s^2)^{1/2}} I_1\left(B\sqrt{1-s^2}\right) + \delta(s-1) \quad \text{for } \nu = -1; \tag{49}$$

from (E-33),

$$w(s) = \left(\frac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) - \frac{(1-s^2)^\nu}{2^\nu \Gamma(\nu+1)} + \frac{(1-s^2)^\nu}{2^\nu \Gamma(\nu+1)}\Bigg|_G \quad \text{for } -2 < \nu < -1; \tag{50}$$

and from (E-39),

$$w(s) = \frac{B^2}{1-s^2} I_2\left(B\sqrt{1-s^2}\right) + \left(\frac{B^2}{2} - 1\right)\delta(s-1) - \delta'(s-1) \quad \text{for } \nu = -2. \tag{51}$$

The extended range for $\nu < -2$ is given in (E-35) and (E-36). Weighting (49) requires a generalized function, namely, a delta function, with its singularity located at $s = 1$. Weighting (51) requires, in addition, the

derivative of a delta function, located at $s = 1$. The intermediate cases in weighting (50) require a generalized function with its singularity located at $s = 1$; interpretation and approximation of this generalized function is given in appendix F.

It can be observed from (49)-(51) that the leading term of $w(s)$ is exactly what would have been obtained from initial weighting (22) by substituting the appropriate value of ν; here we are using $I_{-n}(z) = I_n(z)$ [5; 9.6.6]. However, the price of crossing the "natural boundary" at $\nu = -1$, which was originally required for (23), is a generalized function with its singularity located at $s = 1$. And the further we go below $\nu = -1$, the more singular becomes the required generalized function; these points are elaborated upon in appendix F.

The explicit assignment of ν values in (48)-(51) leads to table 3 for the weighting-pattern pairs. With regard to application of (48)-(51) to the array

<div align="center">

Table 3. Weighting-Pattern Pairs; $\mu > -1$

</div>

Weighting	Pattern
(49)	$\mathcal{J}_{\mu}\left(\sqrt{u^2 - B^2}\right)$
(50) with $\nu = -\dfrac{3}{2}$	$\mathcal{J}_{\mu - \frac{1}{2}}\left(\sqrt{u^2 - B^2}\right)$
(51)	$\mathcal{J}_{\mu - 1}\left(\sqrt{u^2 - B^2}\right)$

processing application in various numbers of dimensions, we have table 4 for the required weightings, where we have specialized the values of μ. In all cases, the pattern realized is the ideal one of (34):

$$g_1(u) = \mathcal{J}_{-\frac{1}{2}}\left(\sqrt{u^2 - B^2}\right) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \cos\left(\sqrt{u^2 - B^2}\right) \quad \text{for all } u. \tag{52}$$

Table 4. Required Weighting for Ideal Pattern

Number of Dimensions	Value of μ	Required Weighting
1	$-\frac{1}{2}$	(49)
2	0	(50) with $\nu = -\frac{3}{2}$
3	$\frac{1}{2}$	(51)

The weighting given by (49) for one dimension, namely, $\mu = -1/2$, has already been presented by van der Maas [9]. However, the application of (49) to the realization of (48) for any $\mu > -1$ is new. Additionally, all the results in (50) and (51) for any $\mu > -1$, and their application to two- and three-dimensional array processing in table 4, are new. An approximation to the ideal pattern in two dimensions, namely, weighting (50) with $\nu = -1.5$, is treated in detail in appendix F.

SOME OTHER WEIGHTINGS

Another candidate class of weightings for consideration is

$$w(s) = \left(\frac{B}{\sqrt{1-s^2}}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) = B^{2\nu}\mathcal{U}_\nu\left(B\sqrt{1-s^2}\right) \quad \text{for } 0 < s < 1, \quad (53)$$

of which (49) and (51) are representative examples, exclusive of the generalized functions. This class is somewhat similar in form to the earlier case in (22). Substitution of (53) in (15) yields pattern

$$g(u) = \int_0^1 ds\; s\left(\frac{s}{u}\right)^\mu J_\mu(us)\; B^{2\nu}\mathcal{U}_\nu\left(B\sqrt{1-s^2}\right). \quad (54)$$

This integral converges at $s = 0$ for $\mu > -1$, but needs no restriction on ν whatsoever.

To our knowledge, evaluation of (54) is not possible in closed form for general ν; however, the following cases are evaluated in appendix G:

$$g(u) = \mathcal{J}_{\mu+1}\left(\sqrt{u^2-B^2}\right) = \mathcal{U}_{\mu+1}\left(\sqrt{B^2-u^2}\right) \qquad \text{for } \nu = 0; \quad (55)$$

$$g(u) = \mathcal{J}_\mu\left(\sqrt{u^2-B^2}\right) - \mathcal{J}_\mu(u)$$

$$= \mathcal{U}_\mu\left(\sqrt{B^2-u^2}\right) - \mathcal{J}_\mu(u) \qquad \text{for } \nu = 1; \quad (56)$$

$$g(u) = \mathcal{J}_{\mu-1}\left(\sqrt{u^2-B^2}\right) - \mathcal{J}_{\mu-1}(u) - \tfrac{1}{2}B^2\mathcal{J}_\mu(u) \quad \text{for } \nu = 2. \quad (57)$$

Numerous special cases for one, two, or three dimensions are available from (55)-(57) by setting $\mu = -\frac{1}{2}$, 0, or $\frac{1}{2}$, respectively.

As an example, if we take (54) and (56) for $\nu = 1$, we have

$$\int_0^1 ds\ s\left(\frac{s}{u}\right)^\mu J_\mu(us)\ \frac{B}{\sqrt{1-s^2}}\ I_1\left(B\sqrt{1-s^2}\right) = \vartheta_\mu\left(\sqrt{u^2-B^2}\right) - \vartheta_\mu(u). \tag{58}$$

Addition of a δ function immediately yields, for $\mu > -1$,

$$\int_0^1 ds\ s\left(\frac{s}{u}\right)^\mu J_\mu(us)\left[\frac{B}{\sqrt{1-s^2}}\ I_1\left(B\sqrt{1-s^2}\right) + \delta(s-1)\right] = \vartheta_\mu\left(\sqrt{u^2-B^2}\right), \tag{59}$$

which has already been presented in (49) and table 3. A similar combination of (54) and (57) yields the results of (51) and the bottom entry of table 3. Additional results for $\nu = 3, 4, \dots$ are possible via the method of appendix G.

One other two-parameter family of weightings that affords a closed form pattern is available from [10; 6.688 1], by identifying $\mu = 1$, $x = u$, $z = iB$, $\cos t = s$, and by using [5; 9.6.3 and 9.1.35]:

$$\int_0^1 ds\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\cos(us)\ \frac{I_\nu\left(B\sqrt{1-s^2}\right)}{\sqrt{1-s^2}} = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} J_{\frac{\nu}{2}}\left(\frac{u+\sqrt{u^2-B^2}}{2}\right) J_{\frac{\nu}{2}}\left(\frac{u-\sqrt{u^2-B^2}}{2}\right) \text{ for } \nu > -1. \tag{60}$$

This result is restricted to the line array. The weighting is continuous at $s = 1$ if $\nu \geq 1$, and the pattern (60) decays at $3+3\nu$ dB/octave. How good this pattern is has not been pursued.

All the above results have been aimed at getting closed form results for the pattern; however, this is by no means necessary. One could consider the class of weightings (53) for <u>any</u> ν, or the class

$$\exp(-B^2 s^2)\ (1-s^2)^\nu \quad \text{for } 0 < s < 1 \tag{61}$$

for example, numerically by substitution in (12) or (15) and use of some integration rule like Simpson's. Once the patterns have been numerically evaluated and plotted for a sufficiently broad range of values of B and ν, good candidates can be selected at will and the corresponding weighting, (61) or (53) for example, easily evaluated. For the line array, this numerical approach is readily accomplished by use of an FFT.

DISCUSSION

The ideal pattern was defined in (34) as $(2/\pi)^{\frac{1}{2}} \cos\left(\sqrt[4]{u^2-B^2}\right)$, and the corresponding weightings were given in (49)-(51). Now in the one-dimensional application, $\mu = -1/2$, van der Maas [9] has indeed shown, by taking the limit of a Dolph-Chebyshev discrete array design, that there is no pattern with a narrower mainlobe for a specified sidelobe level (and vice versa) than (34). However, strictly speaking, we have not proven that same result for the other values of μ, i.e., other numbers of dimensions. Instead, we have adopted (34) as an ideal pattern and shown that it can be realized by finite-support weighting functions with a generalized function whose singularity is centered at the edge of the array. Conceivably, there might be a different weighting that would realize a pattern that gets further into the left-lower quadrant of figure 16. However, we conjecture that this is not possible and that the leftmost curve in figure 16 is the ultimate attainable region for any weighting in any number of dimensions.

SUMMARY

We have presented a two-parameter class of Bessel weighting functions that have a closed form pattern with controllable mainlobe width, mainlobe-to-peak-sidelobe ratio, and sidelobe decay rate. These results have application to arrays in one, two, or three dimensions. In addition, the ideal pattern and the corresponding weightings required in various numbers of dimensions have been derived and presented. Where a generalized function is required, a method of approximating it has been presented and illustrated by a numerical example. Various weightings already extant in the literature were shown to be special cases or limiting cases of the general results given here.

Appendix A
DERIVATION OF RESPONSE OF PLANAR ARRAY

Let the receiving array lie in the x-y plane, and let a plane wave of wavelength λ arrive at polar angle ϕ_a and azimuthal angle θ_a; see figure A-1. Then if the time of arrival of the plane wave at the origin is denoted
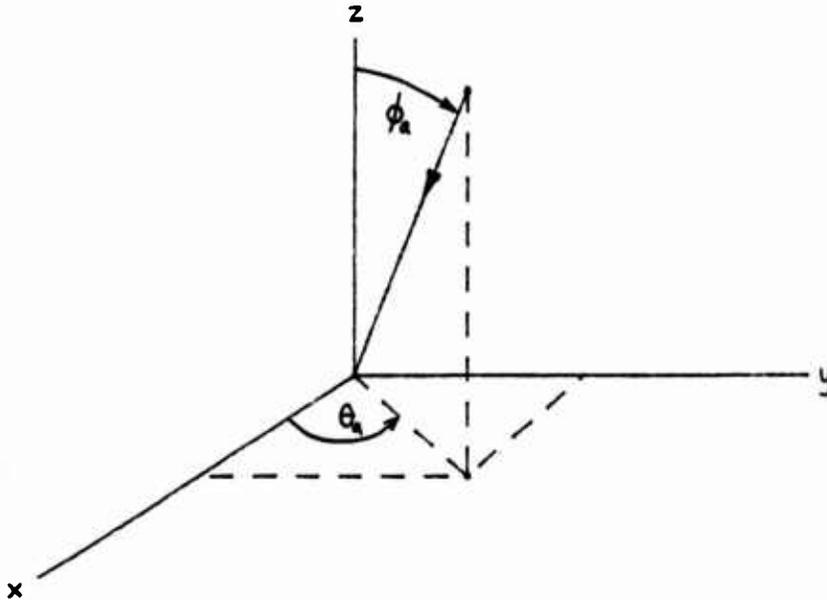
Figure A-1. Geometry for Planar Array

as 0, the time of arrival at a general point x,y in the plane of the array is

$$\tau_a = -\sin\phi_a \frac{x \cos\theta_a + y \sin\theta_a}{c} , \qquad (A-1)$$

where $\lambda = c/f_0$, c is the speed of propagation, and f_0 is the frequency of the plane-wave arrival.

To look in direction ϕ_ℓ, θ_ℓ, the receiving array should employ steering-delay

$$\tau_\ell = -\sin\phi_\ell \frac{x \cos\theta_\ell + y \sin\theta_\ell}{c} \qquad (A-2)$$

37

at the general point x,y. If, also, weighting $w_2(x,y)$ is used in the receiving array, the array output voltage response at frequency f_o is

$$g = \iint dx\ dy\ w_2(x,y)\ \exp\left[-i2\pi f_o\left(\tau_a + \tau_\ell\right)\right]$$

$$= \iint dx\ dy\ w_2(x,y)\ \exp\left[-i2\pi\ \frac{x}{\lambda}\ P_1 - i2\pi\ \frac{y}{\lambda}\ P_2\right],\tag{A-3}$$

where we define angle functions

$$P_1 = \sin\phi_\ell \cos\theta_\ell \cdot \sin\phi_a \cos\theta_a,\quad P_2 = \sin\phi_\ell \sin\theta_\ell - \sin\phi_a \sin\theta_a,\tag{A-4}$$

and where the integration is carried out over all x,y where weighting $w_2 \neq 0$. Thus the planar array can have arbitrary geometry in the x,y plane; equations (A-3) and (A-4) are general results for the array response.

If weighting w_2 contains an impulse at x_0, y_0, then we have

$$w_2(x,y) = \delta(x-x_0)\,\delta(y-y_0),\tag{A-5}$$

with array response

$$g = \exp\left[-i2\pi\ \frac{x_0}{\lambda}\ P_1 - i2\pi\ \frac{y_0}{\lambda}\ P_2\right],\tag{A-6}$$

which never decays in amplitude with increasing angle.

SPECIALIZATION TO CIRCULAR ARRAY

As a special case of the above, consider a planar-circular array of radius R with weighting independent of angle; i.e.,

$$w_2(x,y) = \left\{ \begin{matrix} w_1\left(\sqrt{x^2+y^2}\right) & \text{for } x^2+y^2 < R^2 \\ 0 & \text{otherwise} \end{matrix} \right\}.\tag{A-7}$$

Then response (A-3) becomes

$$g = \int_0^R dr\, r \int_{-\pi}^{\pi} d\theta\, w_1(r) \exp\left[-12\pi\frac{r}{\lambda}\left(P_1 \cos\theta + P_2 \sin\theta\right)\right]$$

$$= 2\pi \int_0^R dr\, r\, w_1(r)\, J_0\left(2\pi\frac{r}{\lambda}\sqrt{P_1^2 + P_2^2}\right). \tag{A-8}$$

And from (A-4), we have

$$\sqrt{P_1^2 + P_2^2} = \left[\sin^2\phi_\ell + \sin^2\phi_a - 2\sin\phi_\ell \sin\phi_a \cos(\theta_\ell - \theta_a)\right]^{1/2}$$

$$= \left|\sin\phi_\ell\, e^{i\theta_\ell} - \sin\phi_a\, e^{i\theta_a}\right|. \tag{A-9}$$

Now let

$$r = Rs \tag{A-10}$$

in (A-8). There follows, for response (A-8),

$$g(u) = \int_0^1 ds\, s\, J_0(us)\, w(s), \tag{A-11}$$

where

$$u \equiv 2\pi\frac{R}{\lambda}\left|\sin\phi_\ell\, e^{i\theta_\ell} - \sin\phi_a\, e^{i\theta_a}\right|, \tag{A-12}$$

and

$$w(s) \equiv 2\pi R^2 w_1(Rs) \quad \text{for } 0 < s < 1. \tag{A-13}$$

Thus the voltage response g(u) is given by a zero-th order Bessel transform over (0,1) of normalized weighting w(s). Dimensionless parameter u incorporates the received wavelength λ, the array radius R, and the various look and arrival angles.

Appendix B
DERIVATION OF RESPONSE OF VOLUMETRIC ARRAY

Consider a plane wave of wavelength λ arriving at angles ϕ_a, θ_a; see figure A-1. The time of arrival (relative to the origin) at a general point x,y,z in the volumetric array is [11; eq. 3]

$$\tau_a = -\frac{1}{c}\left[x \cos\theta_a \sin\phi_a + y \sin\theta_a \sin\phi_a + z \cos\phi_a\right]. \quad (B-1)$$

To look in direction ϕ_ℓ, θ_ℓ, the delay used at location x,y,z should be

$$\tau_\ell = \frac{1}{c}\left[x \cos\theta_\ell \sin\phi_\ell + y \sin\theta_\ell \sin\phi_\ell + z \cos\phi_\ell\right]. \quad (B-2)$$

The response of a weighted array, at frequency f_0, is then

$$g = \iiint dx\, dy\, dz\, w_3(x,y,z) \exp\left[-i 2\pi f_0(\tau_a + \tau_\ell)\right]$$

$$= \iiint dx\, dy\, dz\, w_3(x,y,z) \exp\left[-i\frac{2\pi}{\lambda}\left(x P_1 + y P_2 + z P_3\right)\right], \quad (B-3)$$

where w_3 is the weighting and

$$P_1 = \cos\theta_\ell \sin\phi_\ell - \cos\theta_a \sin\phi_a,$$

$$P_2 = \sin\theta_\ell \sin\phi_\ell - \sin\theta_a \sin\phi_a,$$

$$P_3 = \cos\phi_\ell - \cos\phi_a. \quad (B-4)$$

The integration in (B-3) is over all x,y,z where $w_3 \neq 0$. (B-3) is the general result for any time-delay steered volumetric array.

Let weighting w_3 depend only on the distance from the array center; that is,

$$w_3(x,y,z) = w_1\left(\sqrt{x^2+y^2+z^2}\right) \text{ for } x^2+y^2+z^2 < R^2, \tag{B-5}$$

where R is the sphere radius. Then the voltage response of the array is, from (B-3) and (B-5),

$$g = \iiint_R dx \; dy \; dz \; w_1\left(\sqrt{x^2+y^2+z^2}\right) \exp\left[-i\frac{2\pi}{\lambda}\left(x P_1 + y P_2 + z P_3\right)\right]$$

$$= \int_0^R dr \; r^2 \int_0^\pi d\phi \; \sin\phi \int_{-\pi}^\pi d\theta \; w_1(r) \exp\left[-i2\pi\frac{r}{\lambda}\left(P_1\cos\theta \; \sin\phi + P_2\sin\theta\sin\phi + P_3\cos\phi\right)\right]$$

$$= 2\pi \int_0^R dr \; r^2 w_1(r) \int_0^\pi d\phi \; \sin\phi \; \exp\left(-i2\pi\frac{r}{\lambda} P_3\cos\phi\right) J_0\left(2\pi\frac{r}{\lambda}\sqrt{P_1^2+P_2^2} \; \sin\phi\right), \tag{B-6}$$

where we have changed to polar coordinates.

In the integral on ϕ, let $t = \cos\phi$; the inner integral in (B-6) becomes, by use of [10; 6.677 6],

$$\int_{-1}^1 dt \; \exp\left(-i2\pi\frac{r}{\lambda}P_3 t\right) \; J_0\left(2\pi\frac{r}{\lambda}\sqrt{P_1^2+P_2^2}\sqrt{1-t^2}\right)$$

$$= 2 \int_0^1 dt \; \cos\left(2\pi\frac{r}{\lambda}P_3 t\right) \; J_0\left(2\pi\frac{r}{\lambda}\sqrt{P_1^2+P_2^2}\sqrt{1-t^2}\right)$$

$$= 2 \; \frac{\sin\left(2\pi\frac{r}{\lambda}Q\right)}{2\pi\frac{r}{\lambda}Q}, \tag{B-7}$$

where we define

$$Q \equiv \sqrt{P_1{}^2 + P_2{}^2 + P_3{}^2}.$$ (B-8)

Then the response is, upon substitution of (B-7) in (B-6),

$$g = 4\pi \int_0^R dr\ r^2\ w_1(r)\ \frac{\sin\left(2\pi\frac{r}{\lambda}Q\right)}{2\pi\frac{r}{\lambda}Q}.$$ (B-9)

Now let $r = Rs$; then the response can be written as

$$g(u) = \int_0^1 ds \left(\frac{2}{\pi}\right)^{\frac{1}{2}} s\ \frac{\sin(us)}{u}\ w(s),$$ (B-10)

where

$$w(s) \equiv (2\pi)^{3/2}\ R^3 w_1(Rs),$$ (B-11)

and dimensionless parameter

$$u \equiv 2\pi\frac{R}{\lambda}Q.$$ (B-12)

The quantity Q, involving the look and arrival angles, can be expressed as

$$Q = \left[2 - 2\cos\phi_\ell\cos\phi_a - 2\sin\phi_\ell\ \sin\phi_a\cos(\theta_\ell - \theta_a)\right]^{\frac{1}{2}}$$ (B-13)

Thus the general result, (B-10), for the volumetric array is given by a sine transform of the normalized weighting w.

Appendix C
MONOTONICITY OF BESSEL WEIGHTING

We investigate here the behavior of weighting (22):

$$w(s) = \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) \quad \text{for } 0 < s < 1. \tag{C-1}$$

Let $y = B\sqrt{1-s^2}$; then (C-1) becomes

$$w(s) = B^{-2\nu} y^{\nu} I_{\nu}(y). \tag{C-2}$$

Now [5; 9.6.28]

$$\frac{d}{dy}\left\{y^{\nu}I_{\nu}(y)\right\} = y^{\nu}I_{\nu-1}(y), \tag{C-3}$$

which is positive for $\nu \geq 0$, $y > 0$; see, for example, [5; 9.6.10]. Thus (C-2) is monotonically increasing in y if $\nu \geq 0$; therefore (C-1) is monotonically decreasing in s if $\nu \geq 0$.

Appendix D
PROGRAM FOR CALCULATION OF PATTERN (28)

```
10      Alpha=-.5
20      PLOTTER IS "9872A"
30      LIMIT 25,175,35,245
40      OUTPUT 705;"VS5"
50      SCALE 0,6*PI,-70,0
60      GRID PI,10
70      PENUP
80      FOR B=2 TO 8 STEP 2
90      B2=B*B
100     F=FNInuxnu(Alpha,B)
110     FOR U=0 TO B STEP .05
120     Y=FNInuxnu(Alpha,SQR(B2-U*U))
130     PLOT U,20*LGT(ABS(Y/F))
140     NEXT U
150     FOR U=B TO 6*PI STEP .05
160     Y=FNJnuxnu(Alpha,SQR(U*U-B2))
170     PLOT U,20*LGT(ABS(Y/F))
180     NEXT U
190     PENUP
200     NEXT B
210     END
220     !
230     DEF FNGamma(X) !  GAMMA(X)  via HART, page 275, #5231
240     N=INT(X)
250     R=X-N
260     IF (N>0) OR (R<>0) THEN 290
270     PRINT "FNGamma(X) IS NOT DEFINED FOR X = ";X
280     STOP
290     IF R>0 THEN 320
300     Gamma2=1
310     GOTO 360
320     P=3.36954359131+R*(1.09850630453+R*(.142928007949+R*3.93013464186E-2))
330     P=43.9410209189+R*(22.9680800836+R*(12.8021598111+R*P))
340     Q=43.9410209191+R*(4.39050474596-R*(7.15075063294-R))
350     Gamma2=P/Q                    ! GAMMA(2+R) for 0 < R < 1
360     IF N>2 THEN 400
370     IF N<2 THEN 450
380     Gamma=Gamma2
390     GOTO 500
400     Gamma=Gamma2
410     FOR K=1 TO N-2
420     Gamma=Gamma*(X-K)
430     NEXT K
440     GOTO 500
450     R=1
460     FOR K=0 TO 1-N
470     R=R*(X+K)
480     NEXT K
490     Gamma=Gamma2/R
500     RETURN Gamma
510     FNEND
520     !
```

```
530    DEF FNJnuxnu(Nu,X)            !  Jnu(x)/x^nu
540    IF ABS(X)<1 THEN 650
550    A=.797884560803
560    IF Nu=0 THEN RETURN FNJo(X)
570    IF Nu=.5 THEN RETURN A*SIN(X)/X
580    IF Nu=-.5 THEN RETURN A*COS(X)
590    IF Nu=1 THEN RETURN FNJ1(X)/X
600    IF Nu=-1 THEN RETURN -FNJ1(X)*X
610    IF Nu=1.5 THEN RETURN A*(SIN(X)-X*COS(X))/X^3
620    IF Nu=-1.5 THEN RETURN -A*(X*SIN(X)+COS(X))
630    IF Nu=2 THEN RETURN (2*FNJ1(X)-X*FNJo(X))/X^3
640    IF Nu=-2 THEN RETURN (2*FNJ1(X)-X*FNJo(X))*X
650    A=Nu
660    IF (INT(A)<>A) OR (A>=0) THEN 680
670    K=A=-Nu
680    S=T=1/(2^A*FNGamma(A+1))
690    R=-.25*X*X
700    Big=ABS(S)
710    FOR N=1 TO 100
720    T=T*R/(N*(N+A))
730    S=S+T
740    Big=MAX(Big,ABS(S))
750    IF ABS(T)<=1E-11*ABS(S) THEN 790
760    NEXT N
770    PRINT "100 TERMS IN FNJnuxnu(Nu,X) AT ";Nu;X
780    PAUSE
790    D=12-LGT(ABS(Big/S))   ! NO. OF SIGNIF. DIGITS
800    IF K>0 THEN S=S+(4*R)^K
810    RETURN S
820    FNEND
830    !
840    DEF FNJo(X)      ! Jo(X) via 9.4.1 & 9.4.3 of AMS 55
850    Y=ABS(X)
860    IF Y>3 THEN 910
870    T=Y*Y/9
880    Jo=.0444479-T*(.0039444-T*.00021)
890    Jo=1-T*(2.2499997-T*(1.2656208-T*(.3163866-T*Jo)))
900    GOTO 970
910    T=3/Y
920    Jo=9.512E-5-T*(.00137237-T*(.00072805-T*.00014476))
930    Jo=.79788456-T*(7.7E-7+T*(.00552740+T*Jo))
940    S=.00262573-T*(.00054125+T*(.00029333-T*.00013558))
950    T=Y-.78539816-T*(.04166397+T*(3.954E-5-T*S))
960    Jo=Jo*COS(T)/SQR(Y)
970    RETURN Jo
980    FNEND
990    !
1000   DEF FNJ1(X)      ! J1(X) via 9.4.4 & 9.4.6 of AMS 55
1010   Y=ABS(X)
1020   IF Y>3 THEN 1070
1030   T=Y*Y/9
1040   J1=.00443319-T*(.00031761-T*.00001109)
1050   J1=X*(.5-T*(.56249985-T*(.21093573-T*(.03954289-T*J1))))
1060   GOTO 1130
1070   T=3/Y
1080   J1=.00017105-T*(.00249511-T*(.00113653-T*.00020033))
1090   J1=.79788456+T*(1.56E-6+T*(.01659667+T*J1))
1100   S=.00637879-T*(.00074348+T*(.00079824-T*.00029166))
1110   T=Y-2.35619449+T*(.12499612+T*(5.65E-5-T*S))
1120   J1=SGN(X)*J1*COS(T)/SQR(Y)
1130   RETURN J1
1140   FNEND
1150   !
48
```

```
1160    DEF FNInuxnu(Nu,X)            !   Inu(x)/x^nu
1170    IF ABS(X)<1 THEN 1290
1180    A=.398942280401
1190    E=EXP(X)
1200    IF Nu=0 THEN RETURN FNIo(X)
1210    IF Nu=.5 THEN RETURN A*(E-1/E)/X
1220    IF Nu=-.5 THEN RETURN A*(E+1/E)
1230    IF Nu=1 THEN RETURN FNI1(X)/X
1240    IF Nu=-1 THEN RETURN FNI1(X)*X
1250    IF Nu=1.5 THEN RETURN A*((X-1)*E+(X+1)/E)/X^3
1260    IF Nu=-1.5 THEN RETURN A*((X-1)*E-(X+1)/E)
1270    IF Nu=2 THEN RETURN (X*FNIo(X)-2*FNI1(X))/X^3
1280    IF Nu=-2 THEN RETURN (X*FNIo(X)-2*FNI1(X))*X
1290    A=Nu
1300    IF (INT(A)<>A) OR (A>=0) THEN 1320
1310    K=A=-Nu
1320    S=T=1/(2^A*FNGamma(A+1))
1330    R=.25*X*X
1340    Big=ABS(S)
1350    FOR N=1 TO 100
1360    T=T*R/(N*(N+A))
1370    S=S+T
1380    Big=MAX(Big,ABS(S))
1390    IF ABS(T)<=1E-11*ABS(S) THEN 1430
1400    NEXT N
1410    PRINT "100 TERMS IN FNInuxnu(Nu,X) AT ";Nu;X
1420    PAUSE
1430    D=12-LGT(ABS(Big/S))  ! NO. OF SIGNIF. DIGITS
1440    IF K>0 THEN S=S*(4*R)^K
1450    RETURN S
1460    FNEND
1470    !
1480    DEF FNIo(X)       ! Io(X) via 9.8.1 & 9.8.2
1490    Y=ABS(X)
1500    IF Y>3.75 THEN 1550
1510    T=Y*Y/14.0625
1520    Io=.2659732+T*(.0360768+T*.0045813)
1530    Io=1+T*(3.5156229+T*(3.0899424+T*(1.2067+T*(T-Io))))
1540    GOTO 1590
1550    T=3.75/Y
1560    Io=.00916281-T*(.02057706-T*(.02635537-T*(.01647633-T*.00392377)))
1570    Io=.39894228+T*(.01328592+T*(.00225319-T*(.00157565-T*Io)))
1580    Io=Io*EXP(Y)/SQR(Y)
1590    RETURN Io
1600    FNEND
1610    !
1620    DEF FNI1(X)       ! I1(X) via 9.8.3 & 9.8.4
1630    Y=ABS(X)
1640    IF Y>3.75 THEN 1690
1650    T=Y*Y/14.0625
1660    I1=.02658733+T*(.00301532+T*.00032411)
1670    I1=X*(.5+T*(.87890594+T*(.51498869+T*(.15084934-T*I1))))
1680    GOTO 1730
1690    T=3.75/Y
1700    I1=.01031555-T*(.02282967-T*(.02895312-T*(.01787654-T*.00420059)))
1710    I1=.39894228-T*(.03988024+T*(.00362018-T*(.00163801-T*I1)))
1720    I1=SGN(X)*I1*EXP(Y)/SQR(Y)
1730    RETURN I1
1740    FNEND
```

49/50
Reverse Blank

Appendix E
DERIVATION OF WEIGHTING FOR IDEAL PATTERN

We want to realize pattern (23),

$$g(u) = \mathcal{J}_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right) \quad \text{for all } u, \tag{E-1}$$

for any $\mu > -1$, but without the restriction $\nu > -1$, which was required for convergence of integral (22). $\mathcal{J}_\alpha(z)$ was defined in (17) et seq.

We begin the derivation for the required weighting to realize (E-1) by substituting (E-1) in (16):

$$w(s) = \int_0^\infty du \, u\left(\frac{u}{s}\right)^\mu J_\mu(su) \mathcal{J}_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right) \quad \text{for } s > 0. \tag{E-2}$$

It is important to observe that we must allow all $s > 0$ in (E-2); hopefully, when we evaluate $w(s)$ from (E-2), it will be zero for $s > 1$.

Now we already know from (22) and (23) that (E-2) yields

$$w(s) = \begin{cases} \left(\frac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1 \\ \\ 0 & \text{for } 1 < s \end{cases} \quad \text{if } \mu > -1, \, \nu > -1. \tag{E-3}$$

Letting $\alpha = \mu+\nu+1$ in (E-2) and (E-3), and eliminating ν, we have the useful integral identity

$$\int_0^\infty du \, u\left(\frac{u}{s}\right)^\mu J_\mu(su) \mathcal{J}_\alpha\left(\sqrt{u^2-B^2}\right) = \begin{cases} \left(\frac{\sqrt{1-s^2}}{B}\right)^{\alpha-\mu-1} I_{\alpha-\mu-1}\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1 \\ \\ 0 & \text{for } 1 < s \end{cases}. \tag{E-4}$$

For convergence of this integral at $u = 0$, we require $\mu > -1$, whereas for convergence at $u = \infty$, we must have $\alpha > \mu$; i.e., $-1 < \mu < \alpha$.

51

Now we have the relation [5; 9.1.30]

$$u\left(\frac{u}{s}\right)^{\mu} J_{\mu}(su) = -\frac{1}{s}\frac{d}{ds}\left\{u\left(\frac{u}{s}\right)^{\mu-1} J_{\mu-1}(su)\right\} \quad \text{for } s > 0. \qquad (E-5)$$

Thus (E-2) can be expressed as

$$w(s) = -\frac{1}{s}\frac{d}{ds}\left\{\int_{0}^{\infty}du\, u\left(\frac{u}{s}\right)^{\mu-1} J_{\mu-1}(su)\, \mathcal{J}_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right)\right\} \quad \text{for } s > 0. \quad (E-6)$$

Appeal to (E-4) now reveals that

$$w(s) = \left\{\begin{array}{c} -\dfrac{1}{s}\dfrac{d}{ds}\left\{\left(\dfrac{\sqrt{1-s^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{1-s^2}\right)\right\}\text{for } 0 < s < 1 \\[4mm] 0 \qquad \text{for } 1 < s \end{array}\right\}, \qquad (E-7)$$

provided that $\mu > 0$, $\nu > -2$.

We have succeeded in extending the range of ν from $\nu > -1$ to $\nu > -2$, as desired, but have apparently altered and restricted the range of μ from $\mu > -1$ to $\mu > 0$. However, this last restriction is due solely to the method of derivation, and may be restored to $\mu > -1$, by observing that the right-hand side of (E-7) is analytic in μ (in fact, constant), and that the function $w(s)$ defined by (E-2) is analytic* in μ for $\mu > -1$. Thus (E-7) gives the required weighting to realize pattern (E-1), provided that

$$\mu > -1, \quad \nu > -2. \qquad (E-8)$$

However, care must be taken, in the evaluation of the derivative in (E-7), to account for any generalized functions that may be generated.

---

* We are using the fact that $\mathcal{J}_{\alpha}(z)$ is an entire function of α, regardless of the value of z; see (17) et seq.

In (E-7), define the function

$$w_1(s) = \begin{cases} \left(\dfrac{\sqrt{1-s^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1 \\ \\ 0 & \text{for } 1 < s \end{cases} . \tag{E-9}$$

Then there follows

$$w(s) = -\frac{1}{s}\frac{d}{ds} w_1(s) \quad \text{for all } s, \quad \text{if } \mu > -1, \ \nu > -2. \tag{E-10}$$

We now break the region $\nu > -2$ into the three subcases (i) $\nu > -1$, (ii) $\nu = -1$, (iii) $-2 < \nu < -1$.

(i) $\nu > -1$

As $s \to 1-$, $w_1(s)$ in (E-9) approaches 0, since $\nu+1 > 0$. Thus $w_1(s)$ in (E-9) is continuous for all s, and we find, by the use of [5; 9.6.28] in (E-7), that

$$w(s) = \begin{cases} \left(\dfrac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1 \\ \\ 0 & \text{for } 1 < s \end{cases} , \tag{E-11}$$

which checks (22), as it must of course, for $\nu > -1$.

We observe that

$$w(s) \sim \frac{(1-s)^{\nu}}{\Gamma(\nu+1)} \quad \text{as } s \to 1-. \tag{E-12}$$

Thus for $-1 < \nu < 0$, there is an integrable singularity in $w(s)$ at $s = 1$. For $\nu > 0$, $w(s)$ is continuous at $s = 1$, while for $\nu = 0$, $w(s)$ has a discontinuous step of value -1 at $s = 1$.

(ii) $\nu = -1$

Now, from (E-9),

$$w_1(s) = \begin{cases} I_0\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1 \\ 0 & \text{for } 1 < s \end{cases} . \tag{E-13}$$

This function has a discontinuous step of value -1 at $s = 1$. Thus, from (E-10) and [5; 9.6.28 and 9.6.6],

$$w(s) = \begin{cases} \dfrac{B}{\sqrt{1-s^2}} I_{-1}\left(B\sqrt{1-s^2}\right) + \delta(s-1) & \text{for } 0 \leq s \leq 1 \\ 0 & \text{for } 1 < s \end{cases}$$

$$= \begin{cases} \dfrac{B}{\sqrt{1-s^2}} I_1\left(B\sqrt{1-s^2}\right) + \delta(s-1) & \text{for } 0 \leq s \leq 1 \\ 0 & \text{for } 1 < s \end{cases} . \tag{E-14}$$

The Bessel function portion of this weighting approaches the finite value $B^2/2$ as $s \to 1-$. The integrable singularity at $s = 1$ of the previous subcase for $\nu > -1$ has evolved here into a δ-function for $\nu = -1$ at $s = 1$.

This weighting $w(s)$ in (E-14) realizes the desired pattern in (E-1) for $\nu = -1$, namely,

$$J_\mu\left(\sqrt{u^2-B^2}\right) \text{ for } \mu > -1. \tag{E-15}$$

The special case of a linear array, $\mu = -1/2$, yields the ideal pattern

$$J_{-\frac{1}{2}}\left(\sqrt{u^2-B^2}\right) = \left(\frac{2}{\pi}\right)^{1/2} \cos\left(\sqrt{u^2-B^2}\right). \tag{E-16}$$

This weighting-pattern pair, (E-14) and (E-16), is already known for the line array under the name of van der Maas [9].

An alternative way of obtaining the result (E-14) is by taking the limit of (E-11) as $\nu \to -1$. First we observe that

$$\lim_{\nu \to -1} \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) = \frac{B}{\sqrt{1-s^2}} I_{-1}\left(B\sqrt{1-s^2}\right) \text{ for } s < 1, \qquad \text{(E-17)}$$

since $I_{\nu}(z)$ is an entire function of ν when $z \neq 0$ [5; 9.6.1]. We then define a difference or remainder function (for $\nu > -1$) as

$$R_{\nu}(s) \equiv \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) - \frac{B}{\sqrt{1-s^2}} I_{1}\left(B\sqrt{1-s^2}\right) \text{ for } s < 1, \qquad \text{(E-18)}$$

where we used [5; 9.6.6]. The area under the remainder function $s\, R_{\nu}(s)$ in a small region near $s = 1$ is

$$\int_{1-\epsilon}^{1} ds\, s\, R_{\nu}(s) = \int_{0}^{(2\epsilon-\epsilon^2)^{1/2}} dt\left[t\left(\frac{t}{B}\right)^{\nu} I_{\nu}(Bt) - B\, I_{1}(Bt)\right]$$

$$= \left(\frac{\sqrt{2\epsilon-\epsilon^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{2\epsilon-\epsilon^2}\right) - I_{0}\left(B\sqrt{2\epsilon-\epsilon^2}\right) + 1 \quad \text{for } \nu > -1, \qquad \text{(E-19)}$$

where we used [5; 9.6.28]. Therefore,

$$\lim_{\nu \to -1} \int_{1-\epsilon}^{1} ds\, s\, R_{\nu}(s) = 1, \quad \text{regardless of } \epsilon\ (>0). \qquad \text{(E-20)}$$

Thus since factor s is 1 at the upper limit of integration,

$$\lim_{\nu \to -1} R_{\nu}(s) = \delta(s-1). \qquad \text{(E-21)}$$

Therefore, combining (E-17) and (E-21), we obtain

$$\lim_{\nu \to -1} \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) = \frac{B}{\sqrt{1-s^2}} I_1\left(B\sqrt{1-s^2}\right) + \int(s-1) \quad \text{for } s \leq 1, \quad \text{(E-22)}$$

in agreement with (E-14).

An alternative and simpler equivalent form of (E-22) is

$$\lim_{\nu \to -1} \left(\sqrt{s}\right)^{\nu} I_{\nu}\left(\sqrt{s}\right) = \frac{I_1\left(\sqrt{s}\right)}{\sqrt{s}} + 2\int(s) \quad \text{for } s \geq 0. \quad \text{(E-23)}$$

The derivation of (E-23) is similar to that given above in (E-18)-(E-21).

(iii) $-2 < \nu < -1$

We return to (E-7) and (E-9). Observe that [5; 9.6.7]

$$w_1(s) \sim \frac{\left(1-s^2\right)^{\nu+1}}{2^{\nu+1}\Gamma(\nu+2)} \quad \text{as } s \to 1-. \quad \text{(E-24)}$$

Since we now have

$$-1 < \nu+1 < 0, \quad \text{(E-25)}$$

there is an integrable singularity in $w_1(s)$ at $s = 1$. Thus the derivative in (E-10) will generate a generalized function with a singularity located at $s = 1$. We handle this case by defining an auxiliary function

$$A(s) \equiv \begin{cases} \dfrac{\left(1-s^2\right)^{\nu+1}}{2^{\nu+1}\Gamma(\nu+2)} & \text{for } 0 < s < 1 \\ \\ 0 & \text{for } 1 < s \end{cases}. \quad \text{(E-26)}$$

Then using (E-9), the difference

$$w_1(s) - A(s) = \begin{cases} \left(\dfrac{\sqrt{1-s^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{1-s^2}\right) - \dfrac{\left(1-s^2\right)^{\nu+1}}{2^{\nu+1}\Gamma(\nu+2)} & \text{for } 0 < s < 1 \\ \\ 0 & \text{for } 1 < s \end{cases} \quad (E\text{-}27)$$

is a continuous function of s; in fact,

$$w_1(s) - A(s) \sim \frac{B^2\left(1-s^2\right)^{\nu+2}}{2^{\nu+3}\Gamma(\nu+3)} \quad \text{as } s\to 1\text{-},\qquad (E\text{-}28)$$

which approaches zero since $\nu > -2$. Therefore, the required weighting $w(s)$ in (E-10) can be expressed as

$$w(s) = -\frac{1}{s}\frac{d}{ds}\left\{w_1(s) - A(s) + A(s)\right\}$$

$$= \left[\left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-s^2}\right) - \frac{\left(1-s^2\right)^{\nu}}{2^{\nu}\Gamma(\nu+1)}\right] - \frac{1}{s}\frac{d}{ds}A(s)$$

$$\equiv D(s) + G(s),\qquad (E\text{-}29)$$

where both $D(s)$ and $G(s)$ are zero for $1 < s$. The difference function $D(s)$ possesses an integrable singularity at $s = 1$; in fact (recall (E-25)),

$$D(s) \sim \frac{B^2\left(1-s^2\right)^{\nu+1}}{2^{\nu+2}\Gamma(\nu+2)} \quad \text{as } s\to 1\text{-}.\qquad (E\text{-}30)$$

The last term in $w(s)$ in (E-29) is a generalized function; from (E-26),

57

$$G(s) = -\frac{1}{s}\frac{d}{ds}A(s)$$

$$= -\frac{1}{2^{\nu+1}\Gamma(\nu+2)}\frac{1}{s}\frac{d}{ds}\left\{\left(1-s^2\right)^{\nu+1}U\left(1-s^2\right)\right\}$$

$$= \frac{\left(1-s^2\right)^{\nu}}{2^{\nu}\Gamma(\nu+1)}U\left(1-s^2\right)\Bigg|_G \qquad \text{for } s > 0, \qquad (E-31)$$

where the sub G denotes a generalized function. Here step function

$$U(t) = \begin{cases} 0 \text{ for } t < 0 \\ 1 \text{ for } t > 0 \end{cases} . \qquad (E-32)$$

Combining (E-29) and (E-31), the required weighting is

$$w(s) = \begin{cases} \left(\frac{\sqrt{1-s^2}}{B}\right)^{\nu} I_\nu\left(B\sqrt{1-s^2}\right) - \frac{\left(1-s^2\right)^{\nu}}{2^{\nu}\Gamma(\nu+1)} + \frac{\left(1-s^2\right)^{\nu}}{2^{\nu}\Gamma(\nu+1)}\Bigg|_G & \text{for } 0 < s < 1 \\ 0 & \text{for } 1 < s \end{cases} .$$

$$(E-33)$$

We have now completed the consideration of the three subcases delineated under (E-10). We now wish to extend ν to values that are equal to and less than -2, so that we can handle the volumetric array discussed in (37) et seq.

We return to (E-6) and employ (E-5) again:

$$w(s) = -\frac{1}{s}\frac{d}{ds}\left\{\int_0^\infty du\left[-\frac{1}{s}\frac{d}{ds}\left\{u\left(\frac{u}{s}\right)^{\mu-2}J_{\mu-2}(su)\right\}\right]\mathcal{J}_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right)\right\}$$

$$= -\frac{1}{s}\frac{d}{ds}\left\{-\frac{1}{s}\frac{d}{ds}\left\{\int_0^\infty du\, u\left(\frac{u}{s}\right)^{\mu-2}J_{\mu-2}(su)\mathcal{J}_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right)\right\}\right\} \quad \text{for } s > 0.$$

$$(E-34)$$

58

Recourse to (E-4) yields

$$
w(s) = \begin{cases} -\dfrac{1}{s}\dfrac{d}{ds}\left\{-\dfrac{1}{s}\dfrac{d}{ds}\left\{\left(\dfrac{\sqrt{1-s^2}}{B}\right)^{\nu+2} I_{\nu+2}\left(B\sqrt{1-s^2}\right)\right\}\right\} & \text{for } 0 < s < 1 \\[4mm] 0 & \text{for } 1 < s \end{cases} \tag{E-35}
$$

provided that $\nu > -3$, $\mu > 1$. However, the last restriction on μ may be restored to $\mu > -1$, by the argument under (E-7). Thus (E-35) gives the required weighting to realize pattern (E-1), provided that

$$
\mu > -1, \quad \nu > -3. \tag{E-36}
$$

The only subcase of (E-35) that we consider in detail is:

<u>(iv) $\nu = -2$</u>

We now can write (E-35) as

$$
w(s) = -\frac{1}{s}\frac{d}{ds}\left\{-\frac{1}{s}\frac{d}{ds}\left\{I_0\left(B\sqrt{1-s^2}\right)U(1-s)\right\}\right\}. \tag{E-37}
$$

Then [5; 9.6.27]

$$
-\frac{1}{s}\frac{d}{ds}\left\{I_0\left(B\sqrt{1-s^2}\right)U(1-s)\right\}
$$
$$
= -\frac{1}{s}I_1\left(B\sqrt{1-s^2}\right)B\frac{1}{2}\left(1-s^2\right)^{-\frac{1}{2}}(-2s)U(1-s) - \frac{1}{s}I_0(0)\left[-\delta(s-1)\right]
$$
$$
= \frac{B}{\sqrt{1-s^2}}I_1\left(B\sqrt{1-s^2}\right)U(1-s) + \delta(s-1). \tag{E-38}
$$

Therefore [5; 9.6.28],

$$w(s) = -\frac{1}{s}\frac{d}{ds}\left\{\frac{B}{\sqrt{1-s^2}}\, I_1\left(B\sqrt{1-s^2}\right)U(1-s) + \delta(s-1)\right\}$$

$$= -\frac{B^2}{s}\frac{d}{ds}\left\{\frac{I_1\left(B\sqrt{1-s^2}\right)}{B\sqrt{1-s^2}}\,U(1-s)\right\} - \delta(s-1) - \delta'(s-1)$$

$$= \frac{B^2}{1-s^2}\, I_2\left(B\sqrt{1-s^2}\right)U(1-s) + \left(\frac{B^2}{2}-1\right)\delta(s-1) - \delta'(s-1). \qquad (E-39)$$

Here we have used (21) and

$$\frac{1}{s}\delta'(s-1) = \delta(s-1) + \delta'(s-1), \qquad (E-40)$$

which follows from

$$f(x)\delta'(x-a) = \left[f(a) + f'(a)(x-a) + 0\left\{(x-a)^2\right\}\right]\delta'(x-a)$$

$$= f(a)\delta'(x-a) - f'(a)\delta(x-a). \qquad (E-41)$$

This required weighting, (E-39) for $\nu = -2$, has both a δ function and a δ' function at $s = 1$.

Summary

The required weightings to realize pattern (E-1) are given by

$$\left.\begin{cases}
(E-11) \text{ for } \nu > -1 \\
(E-14) \text{ for } \nu = -1 \\
(E-33) \text{ for } -2 < \nu < -1 \\
(E-39) \text{ for } \nu = -2 \\
(E-35) \text{ for } -3 < \nu < -2
\end{cases}\right\} \text{ for } \mu > -1. \qquad (E-42)$$

60

APPROXIMATION OF A GENERALIZED FUNCTION

Equation (50) or (E-33) gives the required weighting, for $-2 < \nu < -1$, to realize pattern (48) in terms of a generalized function which is difficult to interpret. Here we address this intrepretation by means of an approximation to the generalized function (E-31). We begin by approximating the (singular) auxiliary function $A(s)$ defined in (E-26) by an ordinary function $A_\epsilon(s)$, and then derive an approximation to generalized function $G(s)$ of (E-31) according to the same rule, namely,

$$G_\epsilon(s) = -\frac{1}{s}\frac{d}{ds}A_\epsilon(s).$$

(F-1)

In particular, consider $A_\epsilon(s)$ as shown in figure F-1; that is, $A_\epsilon(s)$ is still given by (E-26) for $0 \leq s \leq 1-\epsilon$, but then tapers linearly to zero at $s = 1$ in order to be everywhere continuous. The height H of $A_\epsilon(s)$ at $s = 1-\epsilon$ is, from (E-26),

$$H = \frac{\epsilon^{\nu+1}(1-\epsilon/2)^{\nu+1}}{\Gamma(\nu+2)} \qquad (> 0).$$

(F-2)

For small ϵ, we have

$$H \sim \frac{\epsilon^{\nu+1}}{\Gamma(\nu+2)} \quad \text{as } \epsilon \to 0+,$$

(F-3)

which tends to $+\infty$ as $\epsilon \to 0+$, since we have, from (E-25),

$$-1 < \nu+1 < 0.$$

(F-4)

The result of applying (F-1) to figure F-1 is shown in figure F-2. The large positive pulse in $(1-\epsilon, 1)$ has height proportional to

$$\frac{H}{\epsilon} = \frac{\epsilon^\nu(1-\epsilon/2)^{\nu+1}}{\Gamma(\nu+2)} \sim \frac{\epsilon^\nu}{\Gamma(\nu+2)} \quad \text{as } \epsilon \to 0+,$$

(F-5)
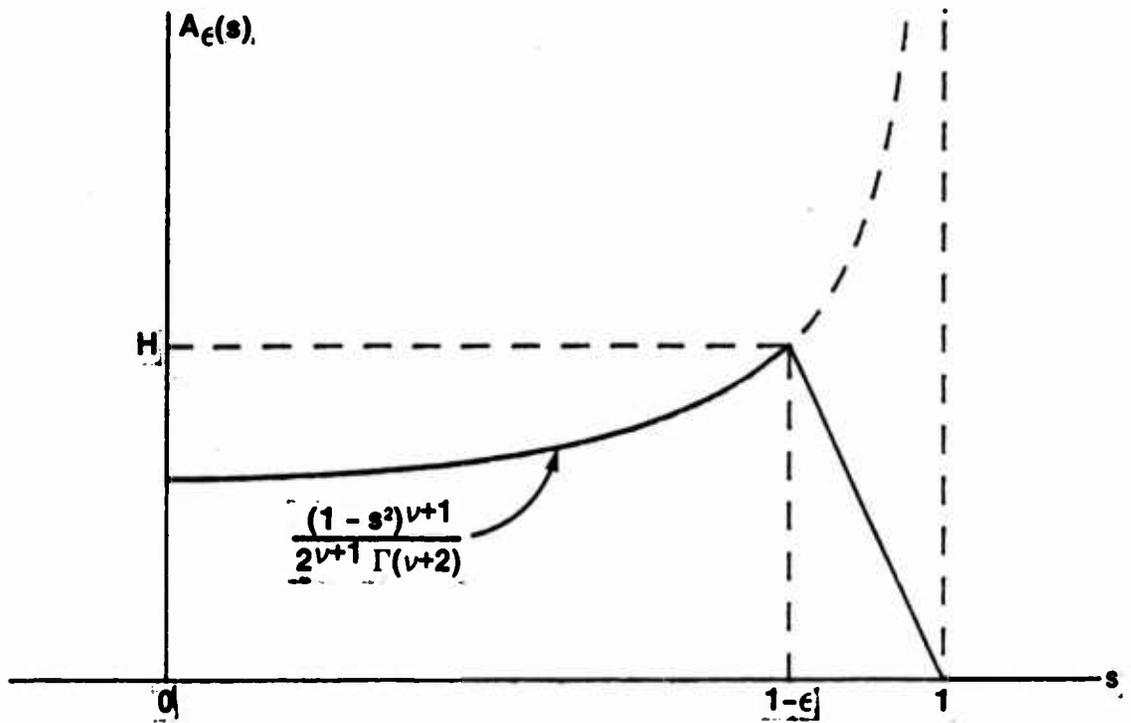
Figure F-1. Approximation to Auxiliary Function A(s)
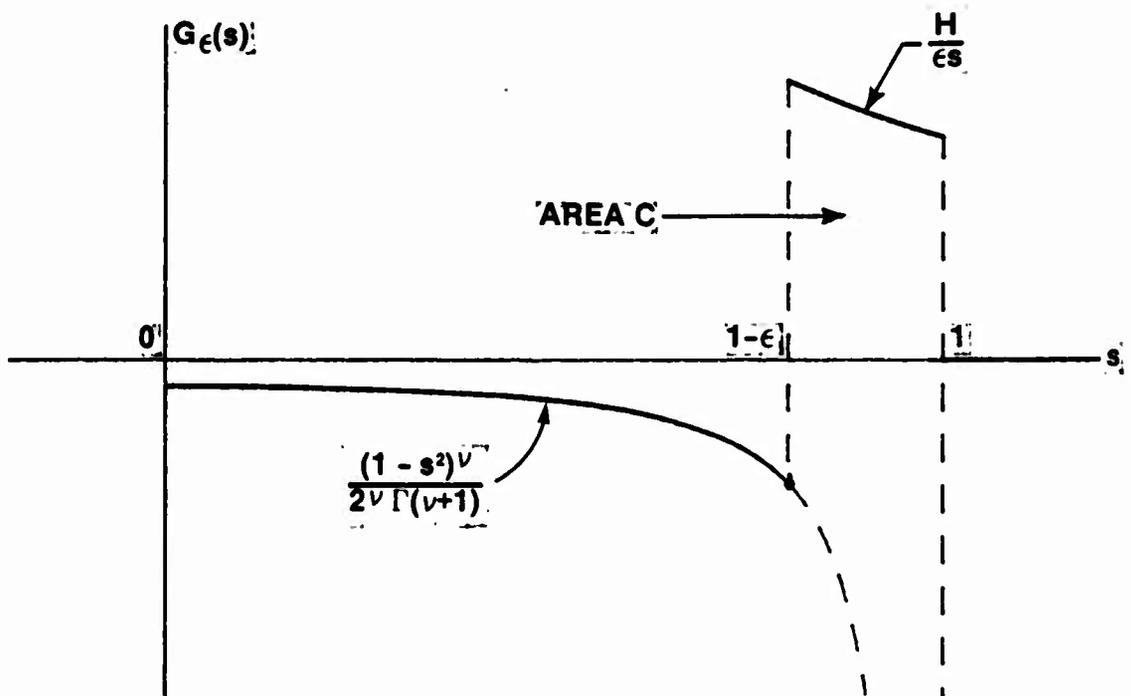
Figure F-2. Approximation to Generalized Function G(s)

which is tending to infinity since $-2 < \nu < -1$. The area of this positive pulse is

$$C = \int_{1-\epsilon}^{1} ds \frac{H}{\epsilon s} = - \frac{H}{\epsilon} \ln(1-\epsilon) \sim \frac{\epsilon^{\nu+1}}{\Gamma(\nu+2)} \quad \text{as } \epsilon \to 0+. \tag{F-6}$$

This area is also tending to $+\infty$ as $\epsilon \to 0+$; recall (F-4). The rate of increase of C is greater, the closer ν is to -2. (For $\nu \to -1$, $H \to 1$, $C \to \frac{-1}{\epsilon} \ln(1-\epsilon)$; thus area $C \sim 1$ as $\epsilon \to 0+$. This is the unit area impulse presented in (49) for $\nu = -1$.)

Figure F-2 is one approximation to generalized function $G(s)$ defined in (E-31). Its most important feature is the impulsive-like positive pulse near $s = 1$. An alternative approximation is afforded in figure F-3, where the area $\tilde{C}$ of the impulse at $s = 1$ is given by (recall (F-6))

$$\tilde{C} = \frac{\epsilon^{\nu+1}}{\Gamma(\nu+2)}. \tag{F-7}$$

The notation used in (E-31) for the generalized function,

$$G(s) = \left. \frac{(1-s^2)^\nu}{2^\nu \Gamma(\nu+1)} \right|_G, \tag{F-8}$$

conceals the positive impulsive behavior at $s = 1$ that the series of approximations in figures F-1 through F-3 indicate must be present. In fact, (F-8) is negative for $0 \le s < 1$, by reference to (F-4).

The alternative approximation we obtain to weighting (50) is then, from (E-29) and figure F-3,
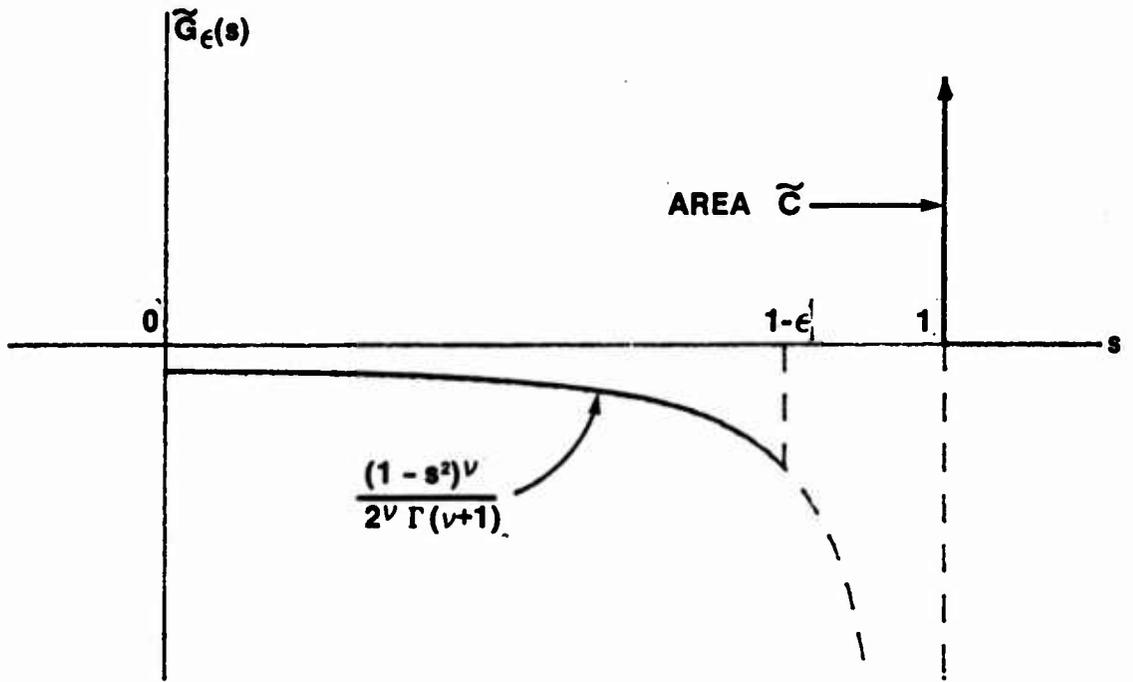
Figure F-3. An Alternative Approximation to Generalized Function G(s)
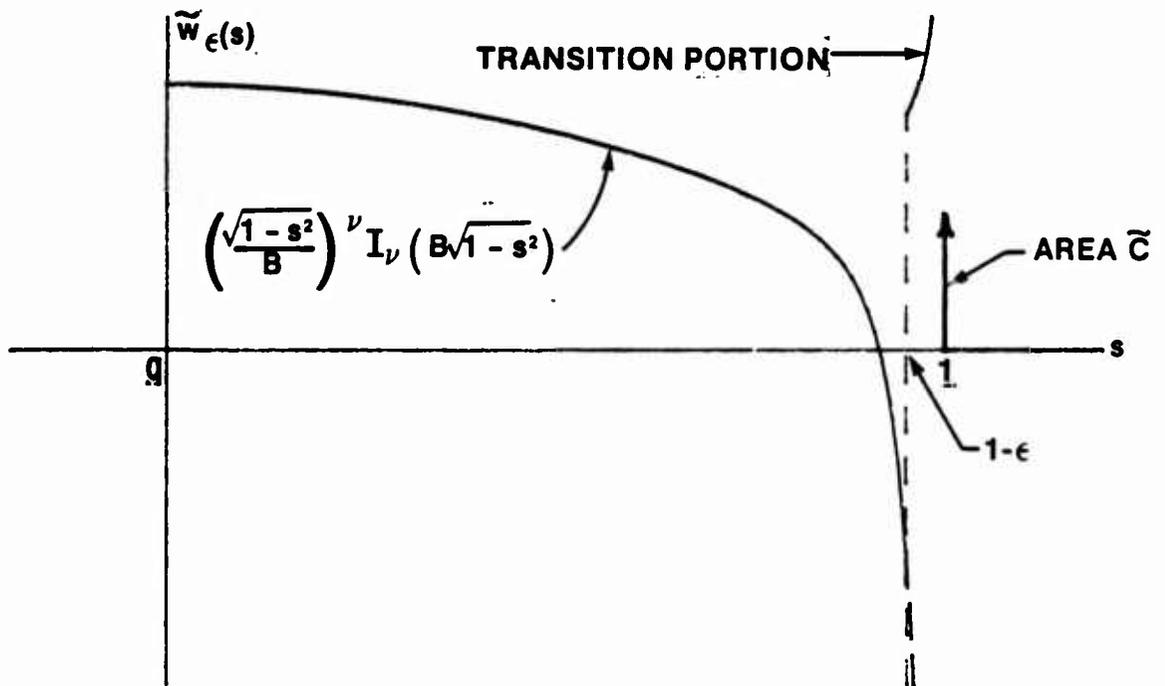
Figure F-4. An Approximation to Weighting (50)

$$\tilde{w}_e(s) = D(s) + \tilde{G}_\varepsilon(s)$$

$$= \begin{cases} \left(\dfrac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1-\varepsilon \\[2em] \left(\dfrac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) - \dfrac{(1-s^2)^\nu}{2^\nu \Gamma(\nu+1)} & \text{for } 1-\varepsilon < s < 1 \end{cases} + \tilde{C}\,\delta(s-1). \qquad (\text{F-9})$$

The plot in figure F-4 illustrates this approximation. The "transition portion" in $(1-\varepsilon, 1)$, which is the bottom line of (F-9), is singular at $s = 1-$; however, this is an integrable singularity, as may be seen by reference to (E-29) and (E-30). The impulse at $s = 1$ is of finite area $\tilde{C}$ given by (F-7).

As $\varepsilon \to 0+$, area $\tilde{C}$ of the impulse tends to infinity; see (F-7) and (F-4). However, the area under the main portion of the approximation $\tilde{w}_\varepsilon(s)$ precisely <u>cancels</u> this singular behavior; that is, as $\varepsilon \to 0+$,

$$\int_a^{1-\varepsilon} ds \left(\frac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) \sim \int_a^{1-\varepsilon} ds\, \frac{(1-s)^\nu}{\Gamma(\nu+1)}$$

$$= \frac{(1-a)^{\nu+1} - \varepsilon^{\nu+1}}{\Gamma(\nu+2)} \sim -\frac{\varepsilon^{\nu+1}}{\Gamma(\nu+2)}, \qquad (\text{F-10})$$

which is $-\tilde{C}$. Since the area under the transition portion is

$$\int_{1-\varepsilon}^1 ds\, D(s) \sim \int_{1-\varepsilon}^1 ds\, \frac{B^2(1-s)^{\nu+1}}{2\Gamma(\nu+2)} = \frac{B^2 \varepsilon^{\nu+2}}{2\Gamma(\nu+3)} \quad \text{as } \varepsilon \to 0+, \qquad (\text{F-11})$$

which tends to 0 as $\varepsilon \to 0+$, the area under approximation $\tilde{w}_\varepsilon(s)$ remains <u>finite</u> as $\varepsilon \to 0+$. Indeed it must remain finite, because Hankel transform (12) or (15) must remain finite in order to realize pattern (48), which is entire in μ, ν, u, and B.

Thus our final approximation to $w(s)$ is simply

$$\hat{w}_\epsilon(s) = \begin{cases} \left(\frac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1-\epsilon \\ 0 & \text{for } 1-\epsilon < s < 1 \end{cases} + \tilde{C}\delta(s-1), \qquad (F-12)$$

and is shown in figure F-5. It is very important to observe that the simple expedient of approximating $w(s)$ by

$$\begin{cases} \left(\frac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) & \text{for } 0 < s < 1-\epsilon \\ 0 & \text{otherwise} \end{cases} \qquad (F-13)$$

is totally inadequate because, as $\epsilon \to 0+$, the area under the cusp at $s = 1-\epsilon$ tends to infinity, and cannot possibly yield an entire function for the pattern. The impulse is <u>necessary</u> to compensate for the singular behavior of (F-13) near $s = 1$; it allows us to realize the "finite part" of the Hankel transform of (F-13) for $\epsilon = 0$.

(As $\nu \to -1$, the value of $\tilde{C}$ in (F-7) tends to the finite value 1, which is the impulse in (49). And as $\nu \to -2$, the doublet of (51) could probably be extracted as a limit from (F-12); this procedure has not been pursued.)

The result of using approximation (F-12) with (F-7), for $\nu = -1.5$ and $B = 4$, is displayed as the patterns in figures F-6 through F-8 for $\epsilon = .1$, .01, .001, respectively. We have selected $\mu = 0$, that is, two dimensions, and are approximating the ideal pattern for $B = 4$ shown in figure 13. It is seen that the approximations become progressively better as ϵ decreases, and that the result in figure F-8 is indeed very close to figure 13.

The approximation $\hat{w}_\epsilon(s)$ in (F-12) and figure F-5 used, for the impulse area, the value $\tilde{C}$ given by (F-7) as a limit of (F-6) for small ϵ. A better approach is not to use the limiting value, but to use the actual value of the pertinent function, since we would like good approximations for moderate values of ϵ, not just very small ϵ. This procedure is considered in detail in [12], with the result

66

$$\bar{w}_\epsilon(s) = \left(\frac{\sqrt{1-s^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right) U(1-\epsilon-s)$$

$$+ \left(\frac{\sqrt{2\epsilon-\epsilon^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{2\epsilon-\epsilon^2}\right)\delta(s-1). \qquad (F-14)$$

The patterns for this approximation are depicted in figures F-9 through F-11. The result in figure F-11 for ϵ = .1 is now <u>better</u> than the result for ϵ = .001 in figure F-8; and all we have done is to modify the area of the impulse at s = 1. The result for ϵ = .15 in figure F-10 indicates a modest change from the ideal and would be acceptable in some cases. The program for the pattern evaluation is listed at the end of this appendix.

Another possibility is to relocate the impulse in $(1-\epsilon, 1)$ to best approximate the ideal pattern in figure 13. More generally, a shaped narrow pulse, which is concentrated toward the boundary at s = 1, could be used; these possibilities are discussed further in [12].

Some additional results involving delta functions and Bessel transforms are given here in appendix H.
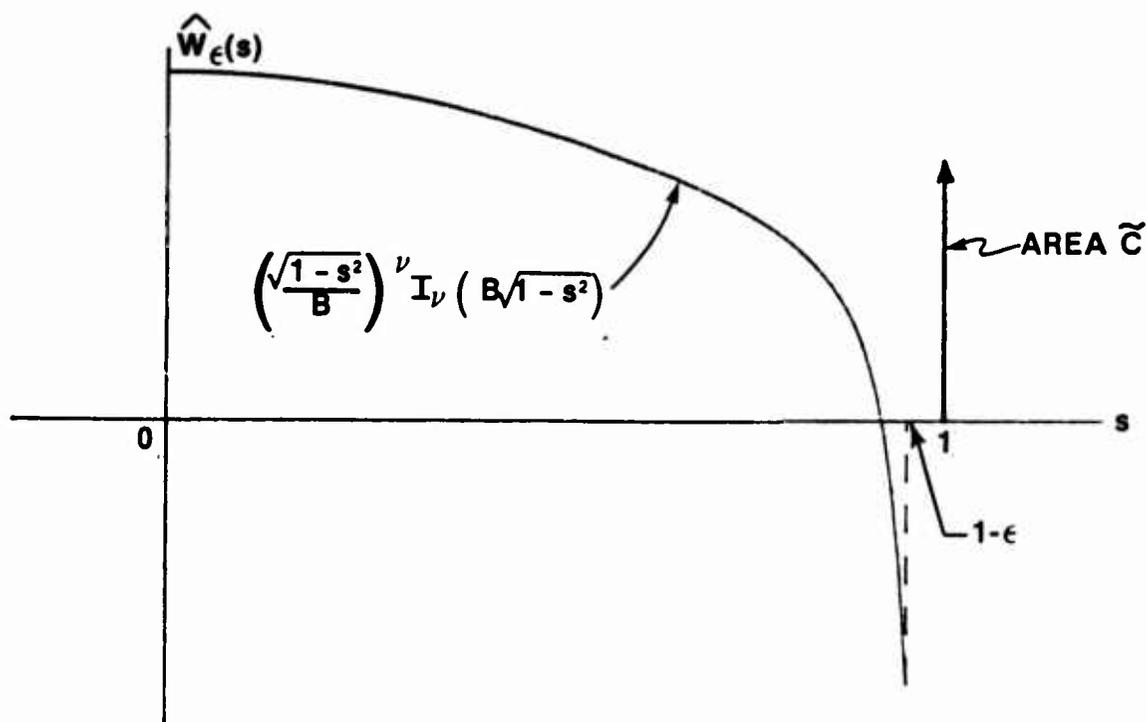
Figure F-5. Final Approximation to Weighting (50)
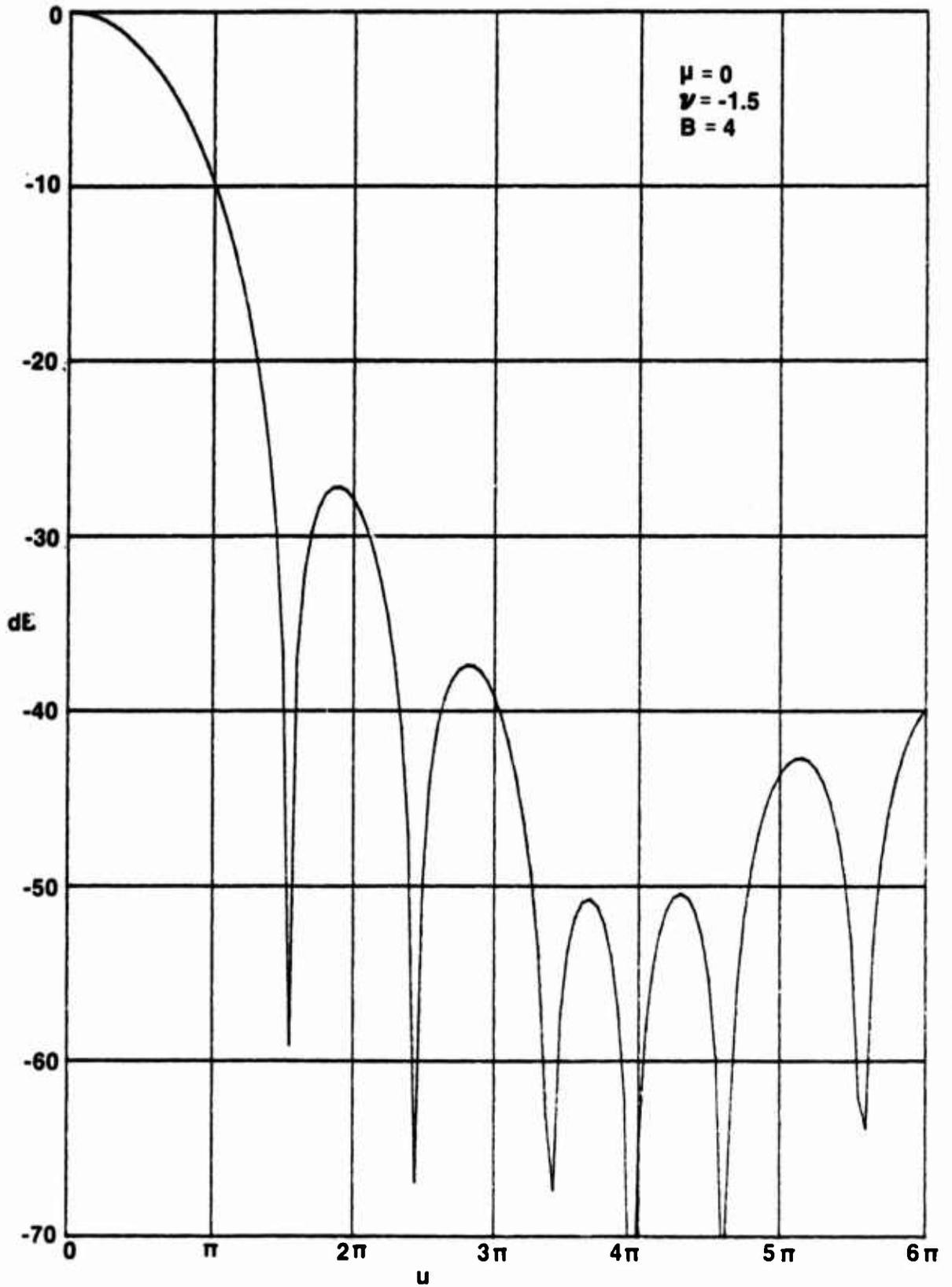
Figure F-6.   Pattern of Approximation (F-12) for $\epsilon = .1$
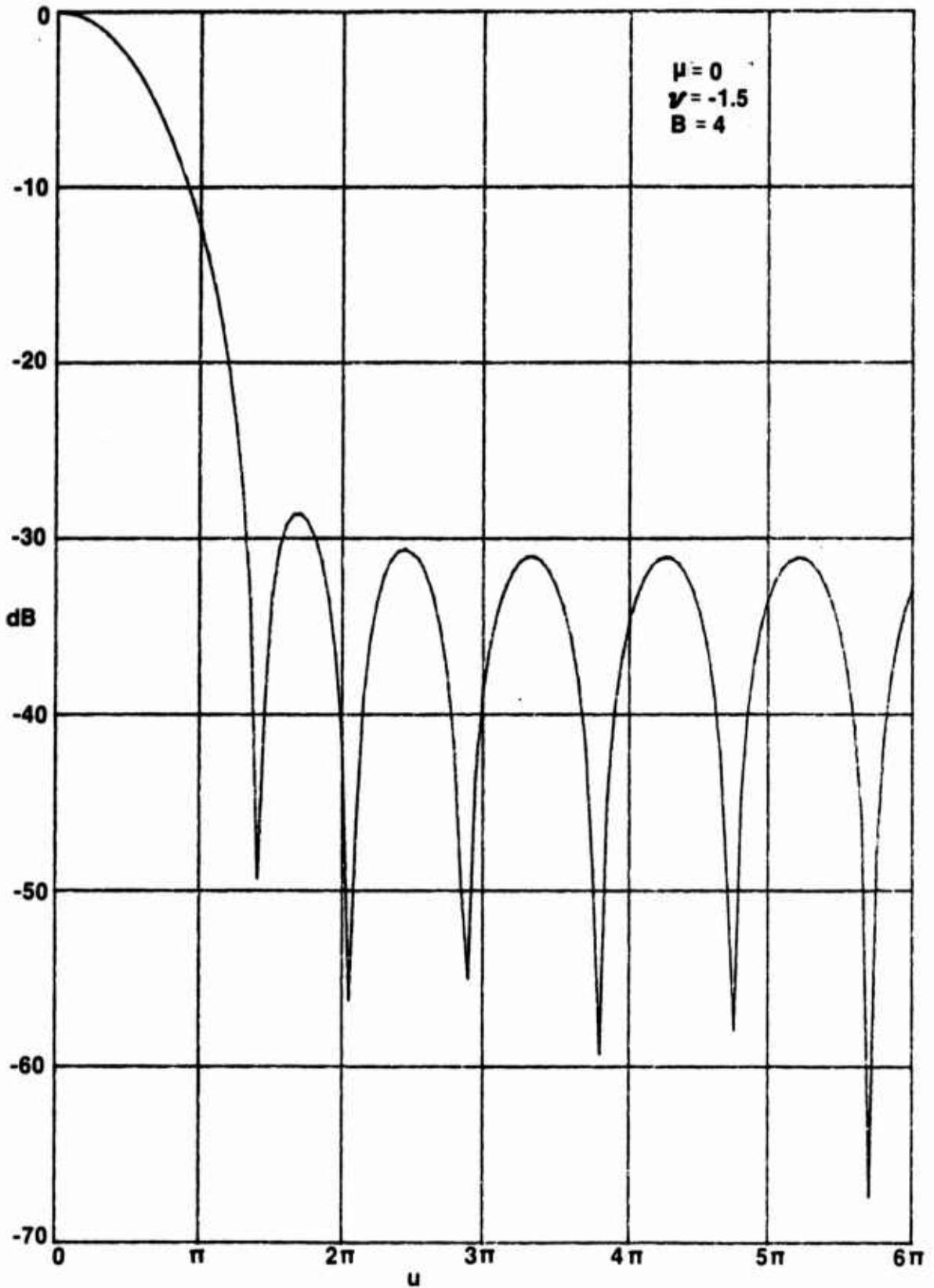
Figure F-7. Pattern of Approximation (F-12) for $\varepsilon = .01$
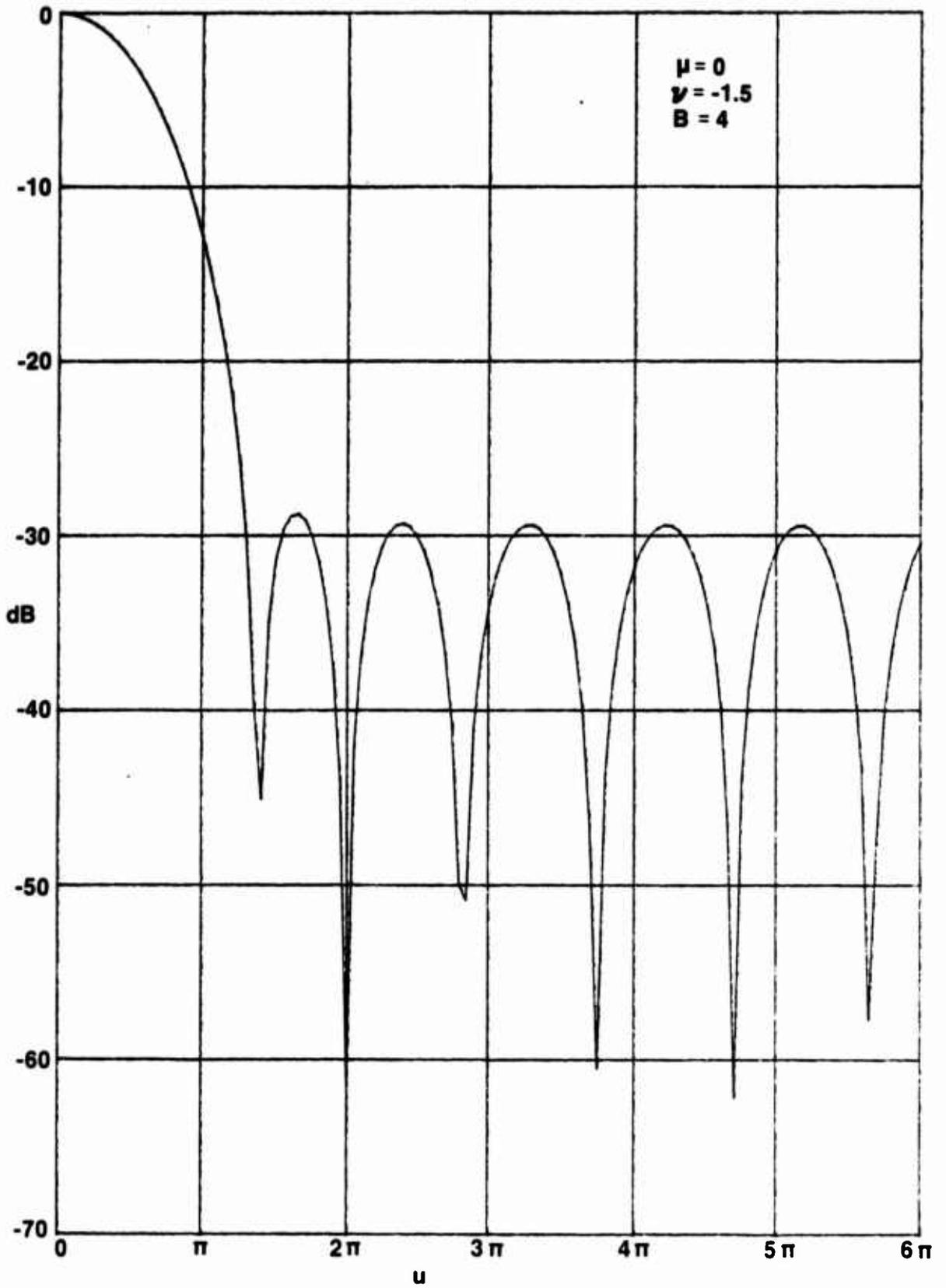
Figure F-8.   Pattern of Approximation (F-12) for ε = .001
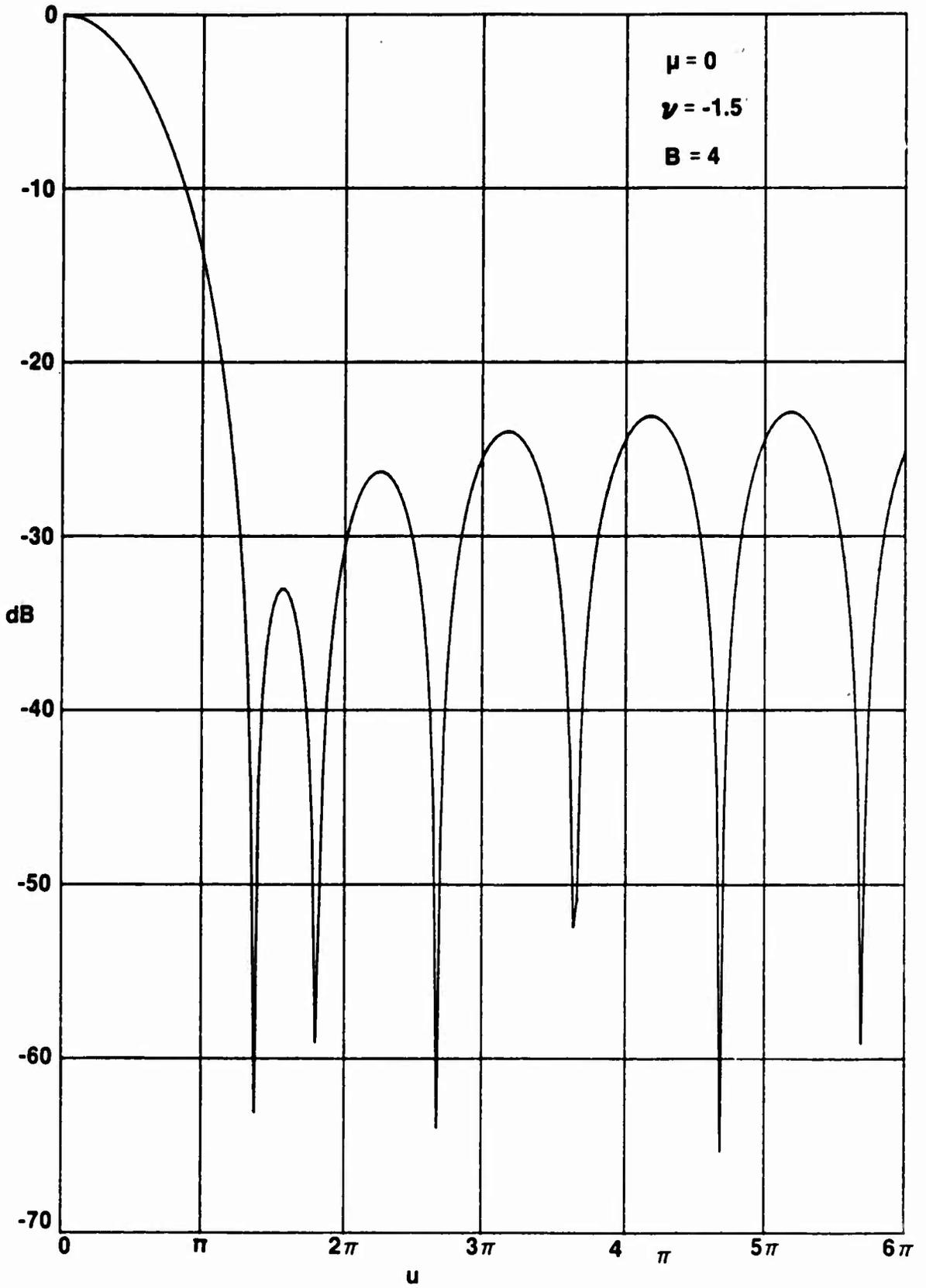
Figure F-9.   Pattern of Approximation (F-14) for ε = .2
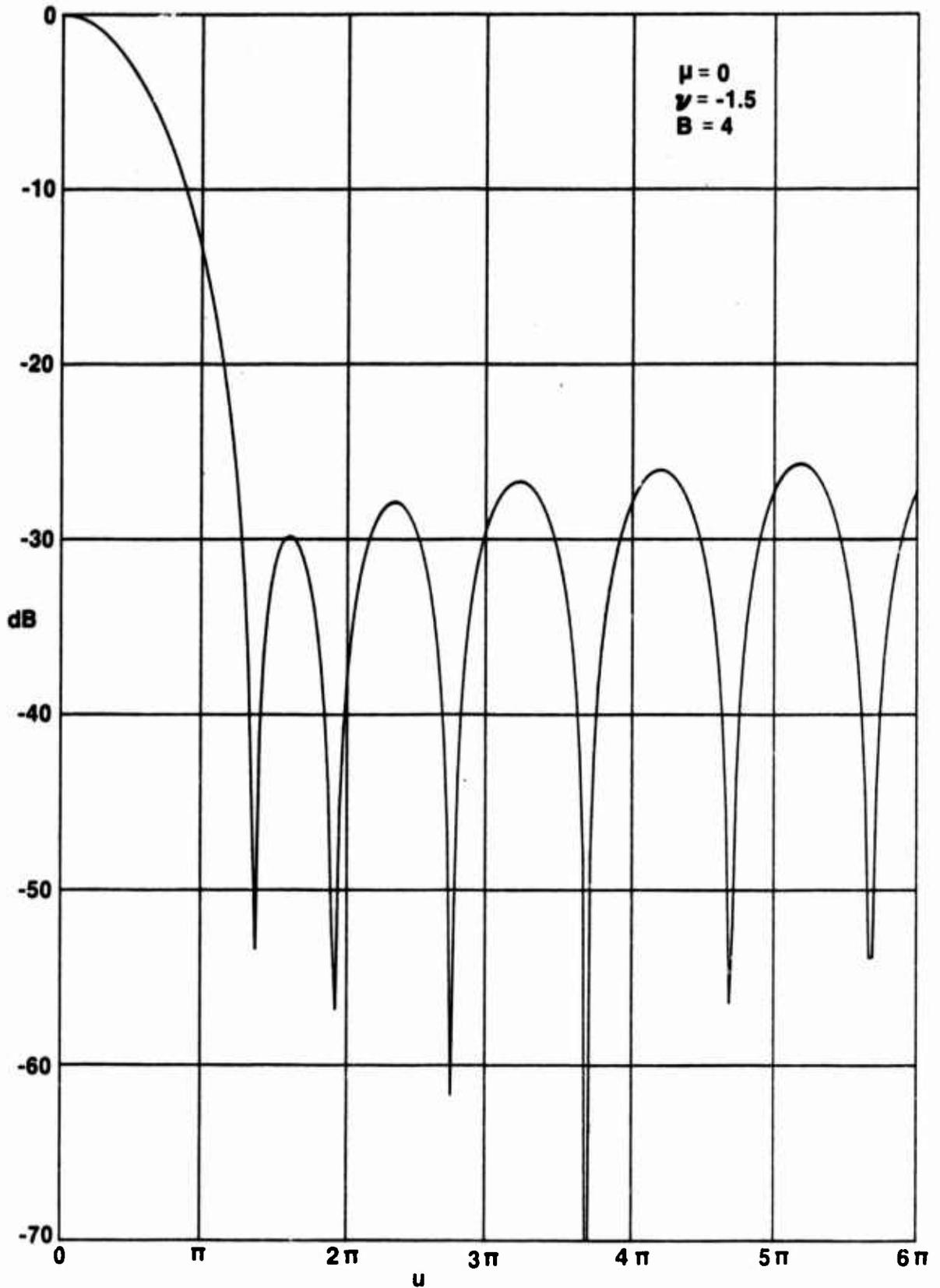
Figure F-10. Pattern of Approximation (F-14) for ε = .15
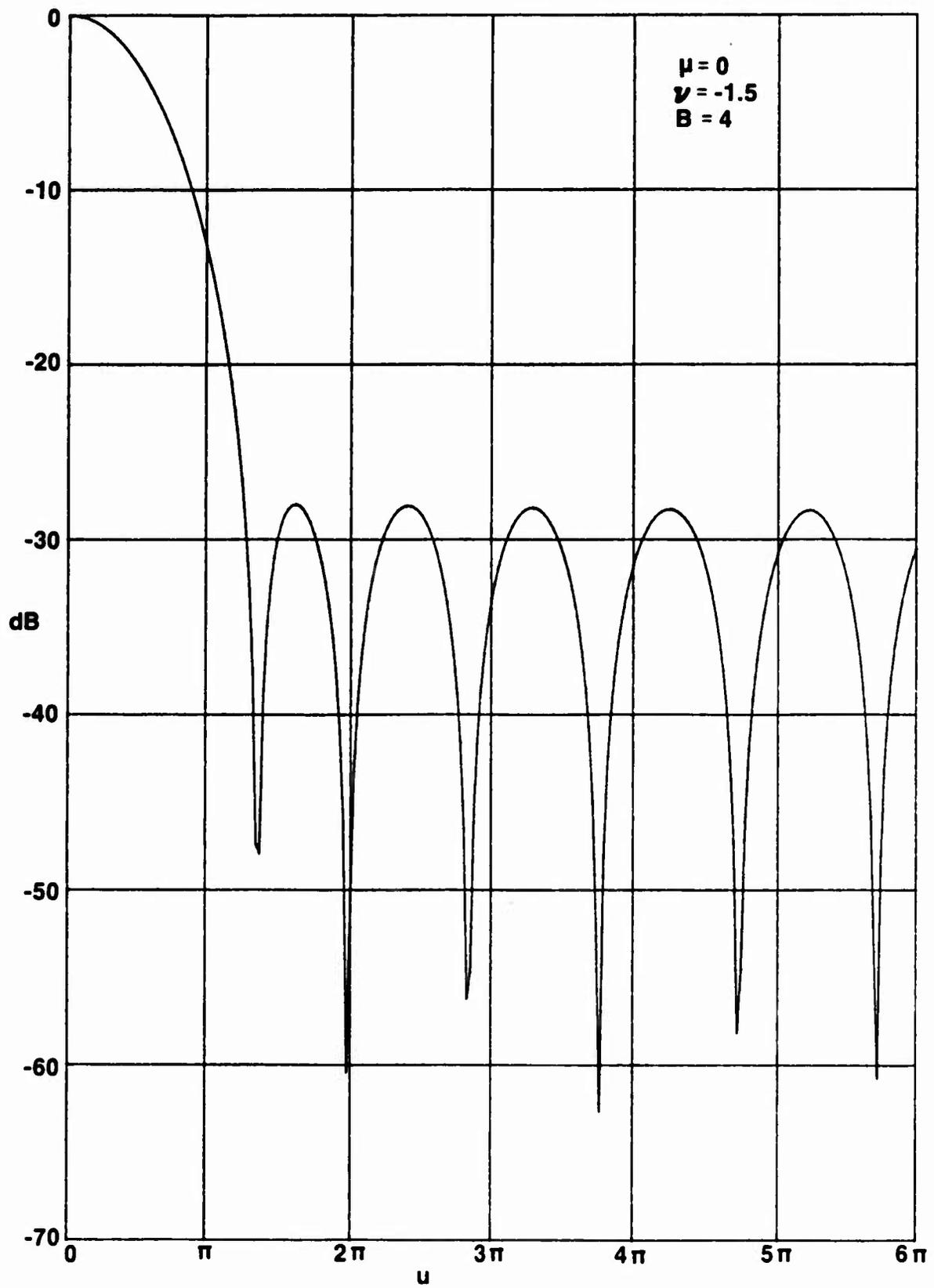
Figure F-11. Pattern of Approximation (F-14) for ϵ = .1

PROGRAM FOR PATTERN EVALUATION

```
10      Eps=.1       ! Pattern for Weighting (F-14)
20      Mu=0
30      Nu=-1.5
40      Bc=4
50      DIM G(0:240)
60      COM U,Bc,Mu,Nu,M21
70      M21=Mu*2+1
80      T=2*Eps-Eps^2
90      F1e=T^(Nu+1)*FNInuxnu(Nu+1,Bc*SQR(T))
100     A=0
110     B=1-Eps
120     FOR Iu=0 TO 240
130     U=Iu*PI/40
140     S=(FNS(A)+FNS(B))*.5
150     N=2
160     H=(B-A)*.5
170     F=(B-A)/3
180     Vo=9E99
190     T=0
200     FOR k=1 TO N-1 STEP 2
210     T=T+FNS(A+H*K)
220     NEXT K
230     S=S+T
240     V=(S+T)*F
250     IF ABS(V-Vo)<=ABS(V)*1E-4 THEN 310
260     Vo=V
270     N=N*2
280     H=H*.5
290     F=F*.5
300     GOTO 190
310     G(Iu)=F1e*FNJnuxnu(Mu,U)+V    ! Voltage Response
320     PRINT Iu,G(Iu)
330     NEXT Iu
340     PLOTTER IS "9872A"
350     LIMIT 25,175,35,245
360     OUTPUT 705;"VS5"
370     SCALE 0,240,-70,0
380     GRID 40,10
390     PENUP
400     FOR Iu=0 TO 240
410     PLOT Iu,20*LGT(ABS(G(Iu)/G(0)))
420     NEXT Iu
430     PENUP
440     END
450     !
460     DEF FNS(S)
470     COM U,Bc,Mu,Nu,M21
480     T=1-S*S
490     T1=FNJnuxnu(Mu,U*S)
500     T2=FNInuxnu(Nu,Bc*SQR(T))
510     RETURN S^M21*T^Nu*T1*T2
520     FNEND
530     !
```

FNJnu nu and FNInu nu are listed in Appendix D

Appendix G
EVALUATION OF A BESSEL INTEGRAL VIA RECURSION

The integral of interest is

$$g_\nu(u,B) \equiv \int_0^1 ds\, K(u,s)\left(\frac{B}{\sqrt{1-s^2}}\right)^\nu I_\nu\left(B\sqrt{1-s^2}\right), \qquad (G\text{-}1)$$

where the kernel K is the Bessel function as given in (11). We have, via [5; 9.6.28],

$$\frac{\partial}{\partial B}\, g_\nu(u,B) = \int_0^1 ds\, K(u,s)\, B\left(\frac{B}{\sqrt{1-s^2}}\right)^{\nu-1} I_{\nu-1}\left(B\sqrt{1-s^2}\right)$$

$$= B\, g_{\nu-1}(u,B). \qquad (G\text{-}2)$$

(This relation actually holds for any kernel K, not just (11).) Since, from (G-1),

$$g_\nu(u,0) = 0 \quad \text{if } \nu > 0,$$

we have the integral recursion

$$g_\nu(u,B) = \int_0^B dt\, t\, g_{\nu-1}(u,t) \quad \text{for } \nu > 0. \qquad (G\text{-}3)$$

We already know the starting case of

$$g_0(u,B) = \int_0^1 ds\, s\left(\frac{s}{u}\right)^\mu J_\mu(us)\, I_0\left(B\sqrt{1-s^2}\right) = \int_{\mu+1}\left(\sqrt{u^2-B^2}\right); \qquad (G\text{-}4)$$

see (23). Substitution in (G-3) yields

$$g_1(u,B) = \int_0^B dt\, t\, \mathcal{J}_{\mu+1}\left(\sqrt{u^2-t^2}\right) = \int_{\sqrt{u^2-B^2}}^u dx\, x\, \mathcal{J}_{\mu+1}(x)$$

$$= \mathcal{J}_\mu\left(\sqrt{u^2-B^2}\right) - \mathcal{J}_\mu(u), \tag{G-5}$$

where we employed (21).

Now we employ (G-3) and (G-5):

$$g_2(u,B) = \int_0^B dt\, t\left[\mathcal{J}_\mu\left(\sqrt{u^2-t^2}\right) - \mathcal{J}_\mu(u)\right]$$

$$= \mathcal{J}_{\mu-1}\left(\sqrt{u^2-B^2}\right) - \mathcal{J}_{\mu-1}(u) - \frac{B^2}{2}\mathcal{J}_\mu(u); \tag{G-6}$$

the integral evaluation follows from direct comparison with (G-5).

The last case for $\nu = 3$ follows in similar fashion:

$$g_3(u,B) = \mathcal{J}_{\mu-2}\left(\sqrt{u^2-B^2}\right) - \mathcal{J}_{\mu-2}(u) - \frac{1}{2}B^2\mathcal{J}_{\mu-1}(u) - \frac{1}{8}B^4\mathcal{J}_\mu(u). \tag{G-7}$$

Appendix H
TWO BESSEL INTEGRALS THAT YIELD GENERALIZED FUNCTIONS

The starting point is the Hankel transform pair in (15) and (16). If we let $w(s) = \delta(s-a)$ (where $a > 0$) in (15), we get

$$g(u) = a\left(\frac{a}{u}\right)^{\mu} J_{\mu}(au) \quad \text{for } u > 0 \quad . \tag{H-1}$$

Substituting (H-1) in (16) then yields the useful relation

$$\int\limits_{0}^{\infty} du\, u\, J_{\mu}(su)\, J_{\mu}(au) = \frac{1}{a}\, \delta(s-a). \tag{H-2}$$

On the other hand, if we let the weighting be a doublet,

$$w(s) = \frac{1}{s}\delta'(s-a), \quad \text{then} \quad g(u) = -u\left(\frac{a}{u}\right)^{\mu} J_{\mu-1}(au). \tag{H-3}$$

The inverse relation (16) yields

$$-\left(\frac{s}{a}\right)^{\mu} \frac{1}{s}\, \delta'(s-a) = \int\limits_{0}^{\infty} du\, u^2\, J_{\mu}(su)\, J_{\mu-1}(au). \tag{H-4}$$

However, since

$$f(s)\delta'(s-a) = \left[f(a) + f'(a)(s-a) + \ldots\right]\delta'(s-a)$$
$$= f(a)\delta'(s-a) - f'(a)\delta(s-a), \tag{H-5}$$

then

$$\int\limits_{0}^{\infty} du\, u^2\, J_{\mu}(su)\, J_{\mu-1}(au) = -\frac{1}{a}\delta'(s-a) + \frac{\mu-1}{a^2}\, \delta(s-a). \tag{H-6}$$

Equations (H-2) and (H-6) are the desired results.

REFERENCES

1. F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, vol. 66, no. 1, January 1978, pp. 51-83.

2. A. H. Nuttall, "Some Windows with Very Good Sidelobe Behavior," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 1, February 1981, pp. 84-91; also NUSC Technical Report 6239, 9 April 1980.

3. J. F. Kaiser, "Digital Filters" in *System Analysis by Digital Computer*, F. F. Kuo and J. F. Kaiser, editors, J. Wiley and Sons, NY, 1966, pp. 218-285.

4. J. F. Kaiser and R. W. Schafer, "On the Use of the I_0-Sinh Window for Spectrum Analysis," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 1, February 1980, pp. 105-107.

5. *Handbook of Mathematical Functions*, U. S. Dept. of Commerce, National Bureau of Standards, Applied Math. Series 55, U.S. Govt. Printing Office, June 1964.

6. W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics*, Chelsea Publishing Co., N.Y., 1954.

7. G. N. Watson, *Theory of Bessel Functions*, Cambridge University Press, London, England, Second Edition, 1958.

8. R. L. Streit, "A Two-Parameter Family of Weights for Nonrecursive Digital Filters and Antennas," submitted to *IEEE Trans. on Antennas*, 21 April 1982.

9. G. J. van der Maas, "A Simplified Calculation for Dolph-Chebyshev Arrays," *Journal of Applied Physics*, vol. 25, no. 1, January 1954, pp. 121-124.

10. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, Inc., N.Y., 1980.

11. A. H. Nuttall, <u>Probability Distribution of Array Response for Randomly Perturbed Element Gains</u>, NUSC Technical Report 5687, 29 September 1977.

12. <u>A. H. Nuttall, Approximations to Some Generalized Functions; Application to Array Processing</u>, NUSC Technical Report 6767, 19 July 1982

Approximations to Some Generalized Functions; Application to Array Processing

A. H. Nuttall

ABSTRACT

Approximations to some generalized functions generated by the process of differentiation are developed and used to interpret some integrals involving generalized functions. Application to the realization of the ideal voltage response pattern of an array is made and illustrated for both an impulse and a smooth approximation to the required generalized weighting function.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF SYMBOLS

$f(x)$	ordinary function
$g(x)$	generalized function
$h(x)$	good function
ϵ	parameter controlling the transition region
$f_\epsilon(x)$	approximation to $f(x)$
$g_\epsilon(x)$	approximation to $g(x)$
$U(x)$	unit step function
$\delta(x)$	delta function
I_ϵ	approximation to integral I
ν, B	weighting parameters; (32)
μ	dimensionality parameter of array; (38)
$w(x)$	required array weighting; (32), (37)
$w_\epsilon(x)$	approximate array weighting; (36), (53)
u	array parameter containing geometry, wavelength, and look and steering angles; (38)
$v(u)$	array voltage response; (38)
$\mathcal{Q}_\alpha(z)$	Bessel function ratio; (40)
$v_i(u)$	ideal voltage pattern; (41)
$v_\epsilon(u)$	approximate voltage pattern; (59)
$n_\epsilon(x)$	neutralizer; figure 7
y	linear transformation of x; (54)

INTRODUCTION

Evaluation of integrals involving generalized functions is often accomplished via an integration by parts, without regard to an interpretation of the behavior of the function at its singular points. See, for example, ref. 1, eqs. (I-17), (I-18), and (I-32); thus, there follows an integral such as

$$\int_{-\infty}^{\infty} dx \, \frac{\cos(wx)}{x^2} = -\pi |w| .$$
(1)

It is difficult to interpret and attach physical significance to this integral; in fact, the major contribution to the integral in (1) comes from the neighborhood of $x = 0$, where the integrand appears to be positive and not integrable, yet the right-hand side of (1) is negative and finite. We would like to approximate the generalized function $1/x^2$ in (1) and get a physical interpretation that is consistent with the result given by (1).

In a recent study of the ideal patterns for arrays in one, two, and three dimensions, it was found that the required weightings were impulsive or more singular than an impulse, depending on the dimensionality of the application; see ref. 2. In order to make these results of practical utility, it is necessary to approximate these singular weightings (generalized functions) by finite-valued functions and thereby realize approximations to the ideal patterns. This approximation procedure and its performance will be detailed here.

APPROXIMATION PROCEDURE

We take as given the possibility of approximating a delta function $\delta(x)$ by finite-valued functions; see, for example, ref. 1, page 279, where a one-sided rectangular pulse and a two-sided Gaussian pulse are used for illustration purposes. Also, on page 280, a one-sided approximation to the doublet $\delta'(x)$ is given. Alternative approximations are presented in ref. 3, pages 11-12.

Let us suppose that $f(x)$ is an ordinary function which is integrable at $x = a+$, but that generalized function (ref. 3, page 30)

$$g(x) \equiv f'(x) \tag{2}$$

is not integrable at $x = a+$. For example, with $a = 0$,

$$f(x) = \begin{cases} -2x^{-\frac{1}{2}} & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}, \tag{3}$$

$$g(x) = \begin{cases} x^{-\frac{3}{2}} & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}, \tag{4}$$

is such a pair; see ref. 3, definition 8. Yet integration by parts yields

$$\int_{-\infty}^{\infty} dx\, g(x)\, h(x) = -\int_{-\infty}^{\infty} dx\, f(x)\, h'(x) = 2\int_{0}^{\infty} dx\, x^{-\frac{1}{2}}\, h'(x), \tag{5}$$

which is integrable for any good function $h(x)$.

To make sense of this situation, we approximate function $f(x)$ by function $f_\epsilon(x)$ as shown in figure 1; that is, $f_\epsilon(x)$ has constant value $f(a+\epsilon)$ in the neighborhood ϵ of a. The approximation to generalized function $g(x)$ that we adopt is

$$g_\epsilon(x) = f'_\epsilon(x) = f(a+\epsilon)\, \delta(x-a) + f'(x)\, U(x-a-\epsilon) , \qquad (6)$$

as depicted in figure 2, where $U(x)$ is the unit step function; that is, $g_\epsilon(x)$ has an impulse of area $f(a+\epsilon)$ at $x = a$, and is equal to $f'(x)$ for $x > a+\epsilon$. (We can approximate the impulse as discussed at the beginning of this section.)
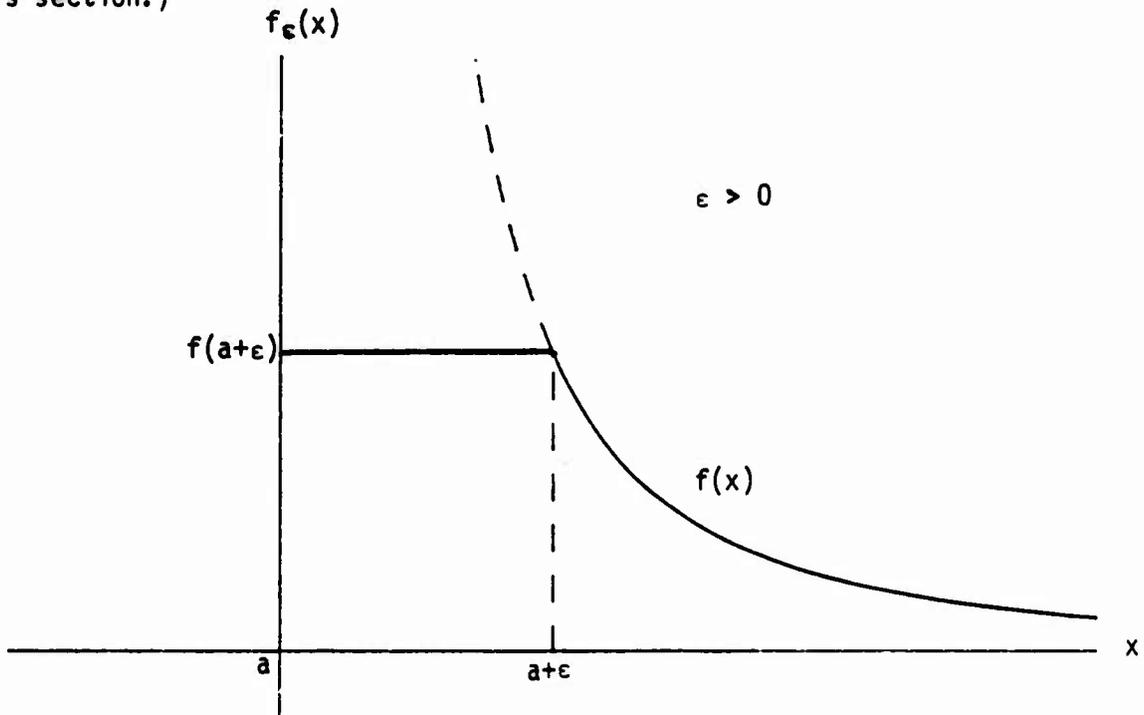
Figure 1. Approximation $f_\epsilon(x)$ to $f(x)$.
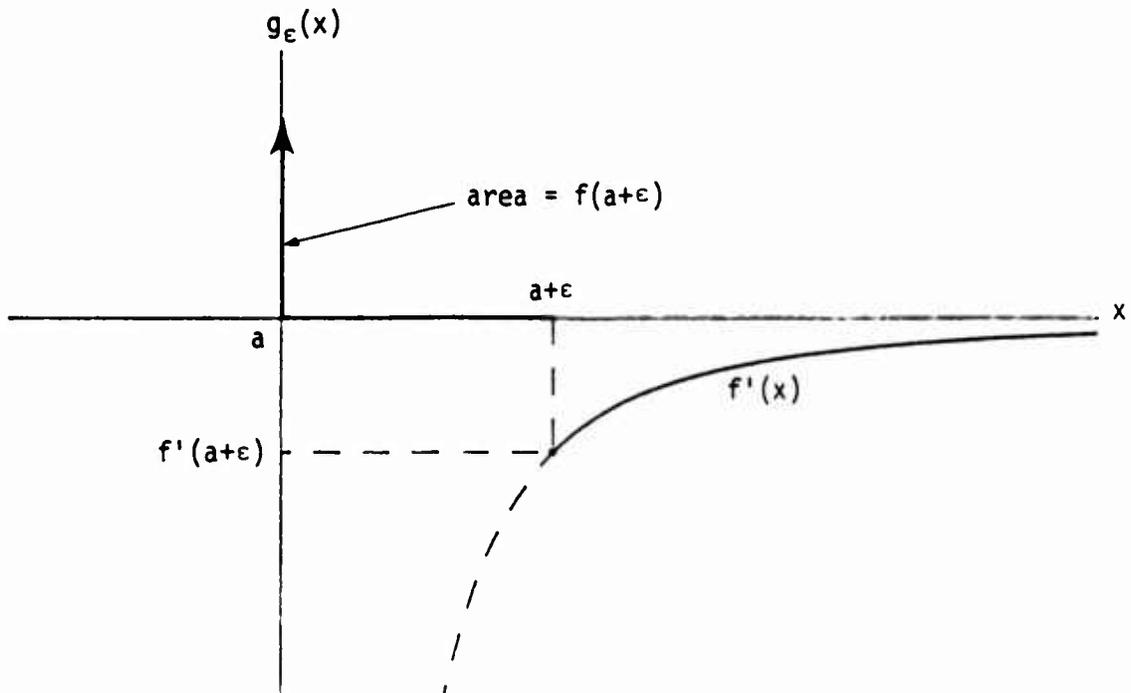
Figure 2. Approximation $g_\epsilon(x)$ to $g(x)$.

3

It is important to observe that if f(x) is singular (ref. 3, definition 21) at x = a, as for example (3), the area f(a+ε) of the impulse at x = a becomes progressively <u>larger</u> as ε→0. However, the area under the remaining portion of $g_\epsilon(x)$ also becomes larger, but in such a fashion that the total area under $g_\epsilon(x)$ remains <u>finite</u> as ε→0; this will be shown below in (11).

To see the effect of this approximation on an integral of a generalized function g(x), consider integral

$$I = \int_a^b dx\ g(x)\ h(x)\ .\tag{7}$$

The approximation to I is:

$$I_\epsilon = \int_a^b dx\ g_\epsilon(x)\ h(x)$$

$$= f(a+\epsilon)\ h(a) + \int_{a+\epsilon}^b dx\ f'(x)\ h(x)\ ,\tag{8}$$

where we employed (6). Integration by parts on (8) yields two alternative forms:

$$I_\epsilon = f(a+\epsilon)\Big[h(a) - h(a+\epsilon)\Big] + f(b)\ h(b) - \int_{a+\epsilon}^b dx\ f(x)\ h'(x)\tag{9A}$$

$$= f(a+\epsilon)\Big[h(a) - h(a+\epsilon)\Big] + \int_a^{a+\epsilon} dx\ f(x)\ h'(x)$$

$$+ f(b)\ h(b) - \int_a^b dx\ f(x)\ h'(x)\ .\tag{9B}$$

Since f(x) is integrable at x = a+, the leading term in (9A) and the two leading terms in (9B) both approach zero as ε→0, yielding the limit

$$I = \lim_{\varepsilon \to 0} I_\varepsilon = f(b)\, h(b) - \int_a^b dx\, f(x)\, h'(x) \ . \tag{10}$$

This is the value of (7) expressed in terms of integrable functions.

The area under approximation $g_\varepsilon(x)$, for $\varepsilon > 0$, is available by substituting h(x) = 1 in (8) and (9):

$$\int_a^b dx\, g_\varepsilon(x) = f(b) \ , \tag{11}$$

which is finite and independent of ε. Thus, the impulse in figure 2 is necessary in order to compensate for the increasing area that develops under f'(x) near x = a, when ε approaches zero.

SINGULARITY AT x = b-

Suppose, instead, that f(x) is integrable at x = b-, but that generalized function (2) is not. By employing a procedure similar to that above, we have approximation

$$g_\varepsilon(x) = -f(b-\varepsilon)\, \delta(x-b) + f'(x)\, U(b-\varepsilon-x) \tag{12}$$

to g(x). The integral I_ε can then be expressed as

$$I_\epsilon = \int_a^b dx \; g_\epsilon(x) \; h(x)$$

$$= -f(b-\epsilon) \; h(b) + \int_a^{b-\epsilon} dx \; f'(x) \; h(x)$$

$$= -f(b-\epsilon)\left[h(b) - h(b-\epsilon)\right] - f(a) \; h(a) - \int_a^{b-\epsilon} dx \; f(x) \; h'(x) \; . \quad (13)$$

There follows

$$\int_a^b dx \; g(x) \; h(x) = I = \lim_{\epsilon \to 0} I_\epsilon = -f(a) \; h(a) - \int_a^b dx \; f(x) \; h'(x) \; , \quad (14)$$

in terms of integrable functions.

The area under approximation $g_\epsilon(x)$ is finite and independent of ϵ: let $h(x) = 1$ in (13) and get directly

$$\int_a^b dx \; g_\epsilon(x) = -f(a) \; . \quad (15)$$

RELATION TO "FINITE PART" OF INTEGRAL

If we start with integral (7) and integrate by parts, there follows

$$I = \left[f(x) \; h(x)\right]_a^b - \int_a^b dx \; f(x) \; h'(x) \; . \quad (16)$$

6

Now if $f(x)$ is singular at $x = a$, the finite part (ref. 3, page 32) of (16) is obtained by dropping the term involving $f(a)$, thereby obtaining identically (10). Conversely, if $f(x)$ is singular at $x = b$, the finite part of (16) is just (14). Thus, the limit of the approximation procedure developed here is exactly what is yielded by the finite part procedure.

EXAMPLES

EXAMPLE 1

$$f(x) = \frac{x^{\nu+1}}{\nu+1} \quad \text{for } x > 0 \quad \text{where } -2 < \nu < -1 . \tag{17}$$

This function is singular but integrable at $x = 0$. In figure 1, we identify

$$a = 0 \tag{18}$$

and get

$$f'(x) = x^{\nu} \quad \text{for } x > 0 . \tag{19}$$

Then, from (6), the approximation to the generalized function $x^{\nu} U(x)$ is

$$g_{\epsilon}(x) = \frac{\epsilon^{\nu+1}}{\nu+1} \delta(x) + x^{\nu} U(x-\epsilon) \tag{20}$$

and is depicted in figure 3.
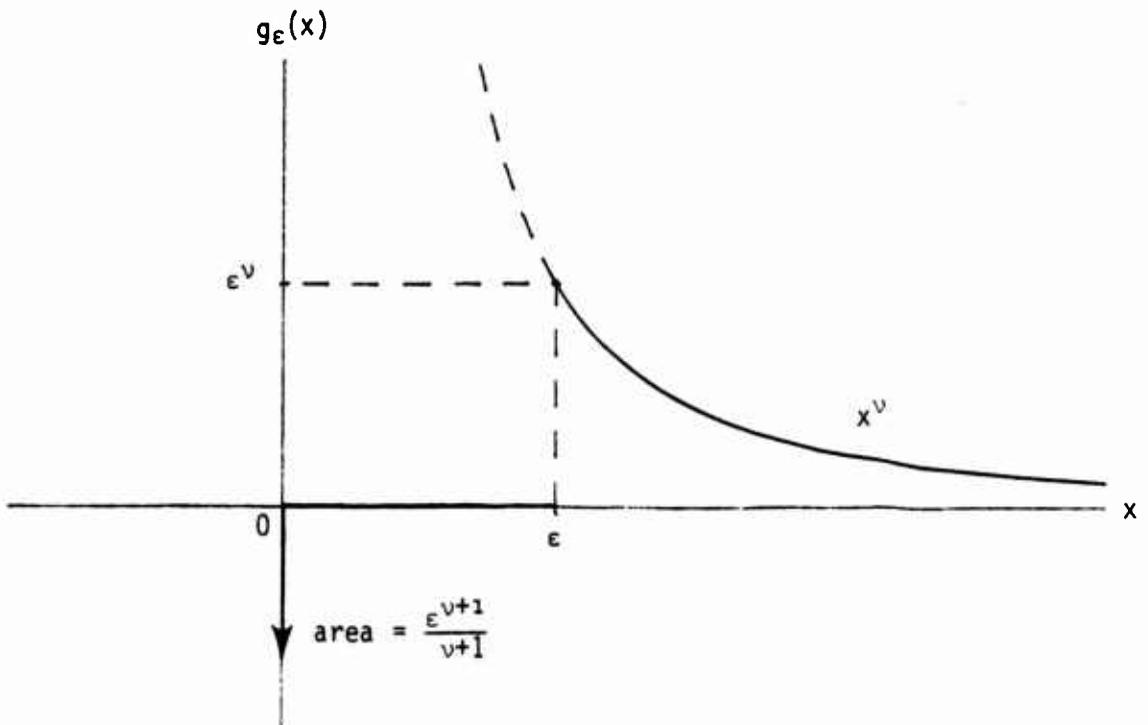
Figure 3. Approximation to Generalized Function $x^{\nu} U(x)$.

8

The result of applying (20) to a good function $h(x)$ is available from (8) and (17),

$$\int_0^b dx\, g_\epsilon(x)\, h(x) = \frac{\epsilon^{\nu+1}}{\nu+1}\, h(0) + \int_\epsilon^b dx\, x^\nu\, h(x)\,, \tag{21}$$

or from either form in (9). The limit follows from (10) and (7):

$$I = \frac{b^{\nu+1}}{\nu+1}\, h(b) - \int_0^b dx\, \frac{x^{\nu+1}}{\nu+1}\, h'(x) = \int_0^b dx\, x^\nu\, h(x)\,. \tag{22}$$

The following examples are derived in similar fashion; just the results are listed.

EXAMPLE 2

$$f(x) = \ln x \quad \text{for } x > 0$$

$$a = 0$$

$$f'(x) = \frac{1}{x} \quad \text{for } x > 0$$

$$g_\epsilon(x) = \ln \epsilon\, \delta(x) + \frac{1}{x}\, U(x-\epsilon)$$

$$I = \ln b\, h(b) - \int_0^b dx\, \ln x\, h'(x) = \int_0^b dx\, \frac{1}{x}\, h(x)\,. \tag{23}$$

EXAMPLE 3

$$f(x) = \frac{1}{2}(\ln x)^2 \quad \text{for } x > 0$$

$$a = 0$$

$$f'(x) = \frac{\ln x}{x} \qquad \text{for } x > 0$$

$$g_\epsilon(x) = \frac{1}{2}(\ln \epsilon)^2 \delta(x) + \frac{\ln x}{x} U(x-\epsilon)$$

$$I = \frac{1}{2}(\ln b)^2 h(b) - \int_0^b dx \, \frac{1}{2}(\ln x)^2 h'(x) = \int_0^b dx \, \frac{\ln x}{x} h(x) \ . \quad (24)$$

EXAMPLE 4

$$f(x) = \frac{x^{\nu+1}}{\nu+1}\left[\ln x - \frac{1}{\nu+1}\right] \quad \text{for } x > 0 \ ; \ \nu > -2$$

$$a = 0$$

$$f'(x) = x^\nu \ln x \quad \text{for } x > 0$$

$$g_\epsilon(x) = \frac{\epsilon^{\nu+1}}{\nu+1}\left[\ln \epsilon - \frac{1}{\nu+1}\right] \delta(x) + x^\nu \ln x \, U(x-\epsilon)$$

$$I = \frac{b^{\nu+1}}{\nu+1}\left[\ln b - \frac{1}{\nu+1}\right] h(b) - \int_0^b dx \, \frac{x^{\nu+1}}{\nu+1}\left[\ln x - \frac{1}{\nu+1}\right] h'(x)$$

$$= \int_0^b dx \, x^\nu \ln x \, h(x) \ . \quad (25)$$

10

EXAMPLE 5

Here we consider the result for the generalized function $1/x^2$ as presented in the integral (1). Consider the ordinary function

$$f_\epsilon(x) = \begin{cases} -1/x & \text{for } |x| > \epsilon \\ -\frac{1}{\epsilon}\,\text{sgn}(x) & \text{for } |x| < \epsilon \end{cases} \quad , \tag{26}$$

and its derivative

$$g_\epsilon(x) = f'_\epsilon(x) = \begin{cases} 1/x^2 & \text{for } |x| > \epsilon \\ -\frac{2}{\epsilon}\,\delta(x) & \text{for } |x| < \epsilon \end{cases} \quad ; \tag{27}$$

see figure 4. The function $g_\epsilon(x)$ is an approximation to generalized function
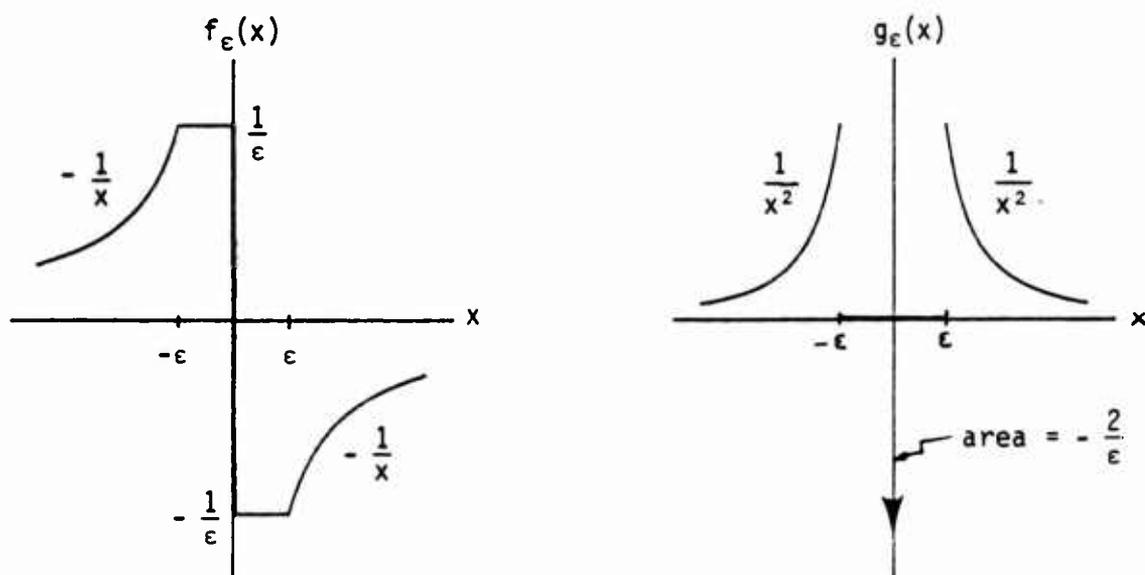
Figure 4. Approximation to Generalized Function $\frac{1}{x^2}$

$1/x^2$; it contains an impulse of area $-2/\epsilon$ (at $x = 0$) that tends to $-\infty$ as $\epsilon \to 0$.

The approximation to integral (7) is as defined in (8); namely, for $a = -\infty$, $b = +\infty$,

$$I_\epsilon = \int_{-\infty}^{+\infty} dx \ g_\epsilon(x) \cos(wx) = 2 \int_\epsilon^{+\infty} dx \ \frac{\cos(wx)}{x^2} - \frac{2}{\epsilon}$$

$$= -\pi |w| - 2 \frac{1 - \cos(w\epsilon)}{\epsilon} + 2w \ Si(w\epsilon), \tag{28}$$

where Si is the sine integral (ref. 4, eq. 5.2.1). The limit as $\epsilon \to 0$ gives the result for the generalized function $1/x^2$, namely,

$$\int_{-\infty}^{+\infty} dx \ \frac{1}{x^2} \cos(wx) = -\pi |w|. \tag{29}$$

This result is in agreement with ref. 1, eq. I-32, and with ref. 3, page 43, for x^{-m} with $m = 2$.

An alternative approximation to generalized function $1/x^2$ that uses finite functions is

$$g_\epsilon(x) = \begin{cases} 1/x^2 & \text{for } |x| > \epsilon \\ \\ 0 & \text{for } |x| < \epsilon \end{cases} - \frac{2}{\epsilon} \frac{1}{\sqrt{2\pi\epsilon}} \ \exp\left(-\frac{x^2}{2\epsilon^2}\right). \tag{30}$$

There follows

$$I_\epsilon = \int_{-\infty}^{+\infty} dx \, g_\epsilon(x) \, \cos(wx)$$

$$= -\pi |w| + \frac{2}{\epsilon} \left[\cos(w\epsilon) - \exp\left(- \frac{1}{2} w^2 \epsilon^2\right) \right] + 2w \, \text{Si}(w\epsilon) \quad . \quad (31)$$

The limit as $\epsilon \to 0$ is again exactly (29). The essential feature of approximations (27) and (30) is the large, sharp negative pulse of area $-2/\epsilon$ that develops about $x = 0$ as ϵ gets small.

APPLICATION TO ARRAY WEIGHTING

This example relies heavily on material presented in ref. 2 on array weighting for good voltage response patterns in one, two, and three dimensions. In particular, from ref. 2, eq. E-7: for $-2 < \nu < -1$, we have, for the required array weighting $w(x)$ to realize pattern $\mathcal{J}_{\mu+\nu+1}\left(\sqrt{u^2-B^2}\right)$, in one, two, or three dimensions $\left(\mu = -\frac{1}{2}, 0, \text{ or } \frac{1}{2}\right)$, the relation

$$-x\, w(x) = \frac{d}{dx}\left\{\left(\frac{\sqrt{1-x^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{1-x^2}\right) U(1-x)\right\} \quad \text{for } 0 < x, \qquad (32)$$

and zero otherwise. To match (2), we identify

$$g(x) = -x\, w(x), \qquad (33)$$

and

$$f(x) = \left(\frac{\sqrt{1-x^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{1-x^2}\right) U(1-x) \quad \text{for } 0 < x. \qquad (34)$$

The function $f(x)$ is singular but integrable at $x = 1-$, since $-1 < \nu+1 < 0$; thus the derivative in (32) will generate a generalized function (ref. 3, page 30) for the weighting $w(x)$.

We now identify $b = 1$ in (12) and obtain, via (34) and ref. 4, eq. 9.6.28, the approximation

$$g_\epsilon(x) = -\left(\frac{\sqrt{2\epsilon-\epsilon^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt{2\epsilon-\epsilon^2}\right)\delta(x-1)$$

$$-x\left(\frac{\sqrt{1-x^2}}{B}\right)^{\nu} I_{\nu}\left(B\sqrt{1-x^2}\right) U(1-\epsilon-x) \quad \text{for } 0 < x. \qquad (35)$$

Then there follows from (33), the approximate weighting

$$w_\epsilon(x) = \left(\frac{\sqrt[\nu]{2\epsilon-\epsilon^2}}{B}\right)^{\nu+1} I_{\nu+1}\left(B\sqrt[\nu]{2\epsilon-\epsilon^2}\right) \delta(x-1)$$

$$+ \left(\frac{\sqrt[\nu]{1-x^2}}{B}\right)^{\nu} I_\nu\left(B\sqrt[\nu]{1-x^2}\right) U(1-\epsilon-x) \quad \text{for } 0 < x . \tag{36}$$

This is the result presented in ref. 2, eq. F-14. It is an approximation to the generalized function

$$w(x) = \left(\frac{\sqrt[\nu]{1-x^2}}{B}\right)^{\nu} I_\nu\left(B\sqrt[\nu]{1-x^2}\right) U(1-x)\, U(x) , \tag{37}$$

which is the required array weighting according to ref. 2, eq. E-7.

We will carry this example further than the previous ones, by determining the voltage response pattern that is actually realized by an array employing approximation weighting (36) rather than (37). The array voltage response pattern is given by ref. 2, eqs. 1-12, for any μ, as

$$v(u) = \int_0^1 dx\ x\left(\frac{x}{u}\right)^\mu J_\mu(ux)\, w(x) , \tag{38}$$

where the parameter μ determines the array dimensionality, and u contains the array geometry, the plane-wave arrival wavelength, and the various look and steering angles. Substitution of generalized function (37) in (38) yields pattern (ref. 2, eq. E-1)

$$v(u) = \mathcal{J}_{\mu+\nu+1}\left(\sqrt[\nu]{u^2-B^2}\right) \quad \text{for all } u , \tag{39}$$

where we define entire function (ref. 2, eqs. 17-21)

$$\mathcal{J}_\alpha(z) \equiv \frac{J_\alpha(z)}{z^\alpha} . \tag{40}$$

The particular case of $\mu + \nu = -1.5$ in (39) yields ideal voltage pattern

$$v_i(u) = \mathcal{J}_{-\frac{1}{2}}\left(\sqrt[4]{u^2-B^2}\right) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\cos\left(\sqrt[4]{u^2-B^2}\right) = \left(\frac{2}{\pi}\right)^{\frac{1}{4}}\cosh\left(\sqrt[4]{B^2-u^2}\right) . \quad (41)$$

This relation is true for all u, whether larger or smaller than B. It indicates a mainlobe at $u = 0$ of amplitude $\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\cosh(B)$, and sidelobes for $u > B$, all of equal amplitude $\pm\left(\frac{2}{\pi}\right)^{\frac{1}{2}}$. Plots of this ideal pattern are available in ref. 2, figures 13-16.

When approximation $w_\epsilon(x)$ in (36) is substituted, instead, in (38), we get several equivalent expressions for the corresponding pattern $v_\epsilon(u)$, as given in appendix A. Plots of typical results are presented in figures 5 and 6 for $\nu = -1.5$. In both figures, the curve labeled $\epsilon = 0$ is the desired pattern (39). Figure 5 corresponds to a volumetric-spherical array ($\mu = .5$), and the desired pattern is, from (39),

$$\mathcal{J}_0\left(\sqrt[4]{u^2-B^2}\right) = J_0\left(\sqrt[4]{u^2-B^2}\right) = I_0\left(\sqrt[4]{B^2-u^2}\right), \quad (42)$$

which decays at a 3 dB/octave rate for large u. Figure 6 corresponds to a planar-circular array ($\mu = 0$) with desired pattern equal to ideal pattern (41). The approximations in both of these figures for $\epsilon = .1$ are quite good, but those for $\epsilon = .2$ have undergone significant degradation. The possibility of replacing the delta function in (36) by a narrow pulse is considered in the next section.

Figure 5 furnishes an approximation to the bottom asterisked case in ref. 2, figure 12; figure 6 does the same for the middle asterisked case in ref. 2, figure 13. The three asterisked cases for $\nu = -1$ in ref. 2, figures 11-13, merely require delta functions at $x = 1$ and are considered solved. The last remaining asterisked case is the bottom one in ref. 2, figure 13, for $\nu = -2$. But this has already been shown in ref. 2, eqs. E-37 - E-39, to involve a delta function and its derivative, both of which are easily approximated; see, for example, ref. 1, pages 279-280, or ref. 2, pages 11-12.
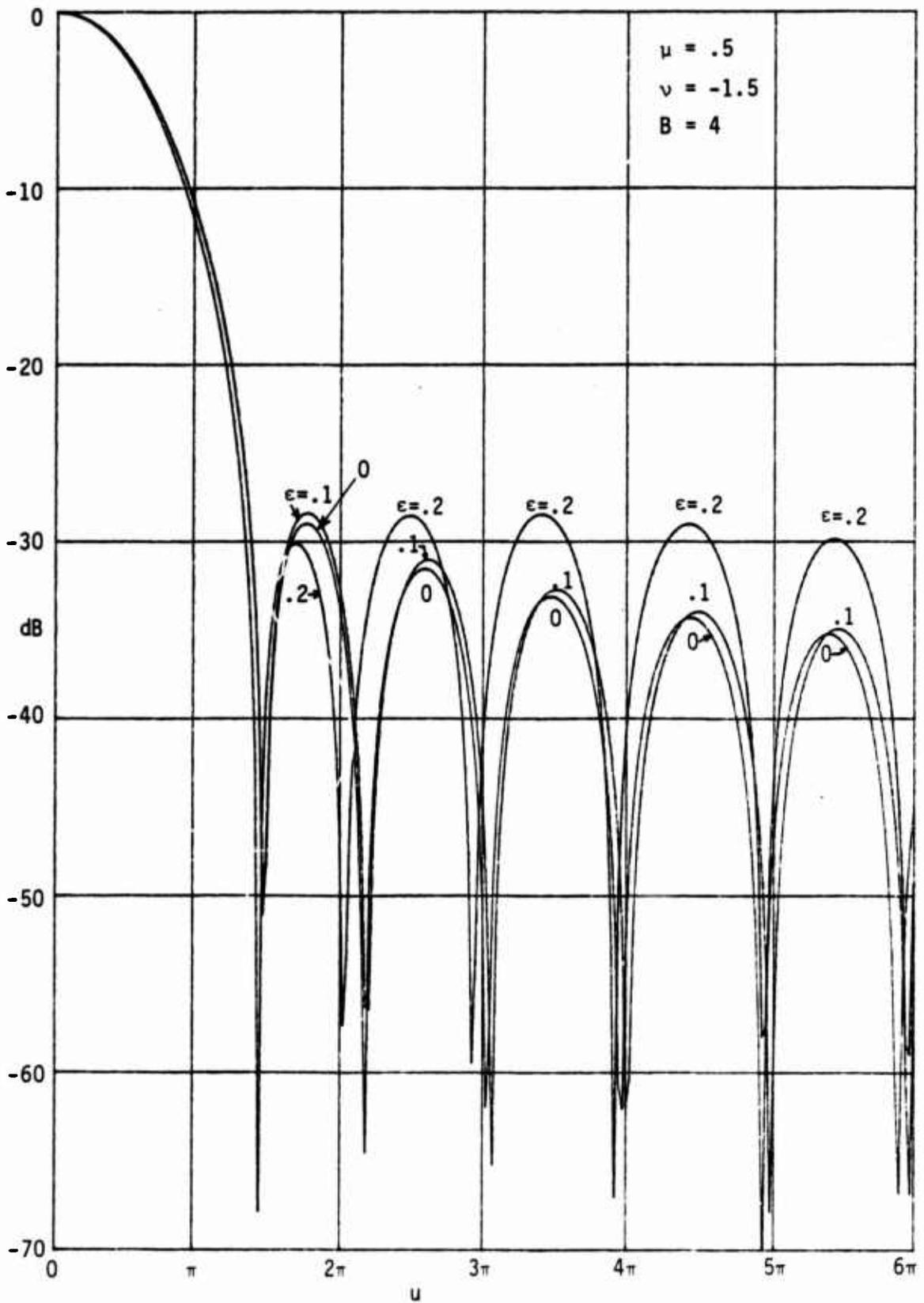
Figure 5. Pattern of Spherical Array for Weighting (36)
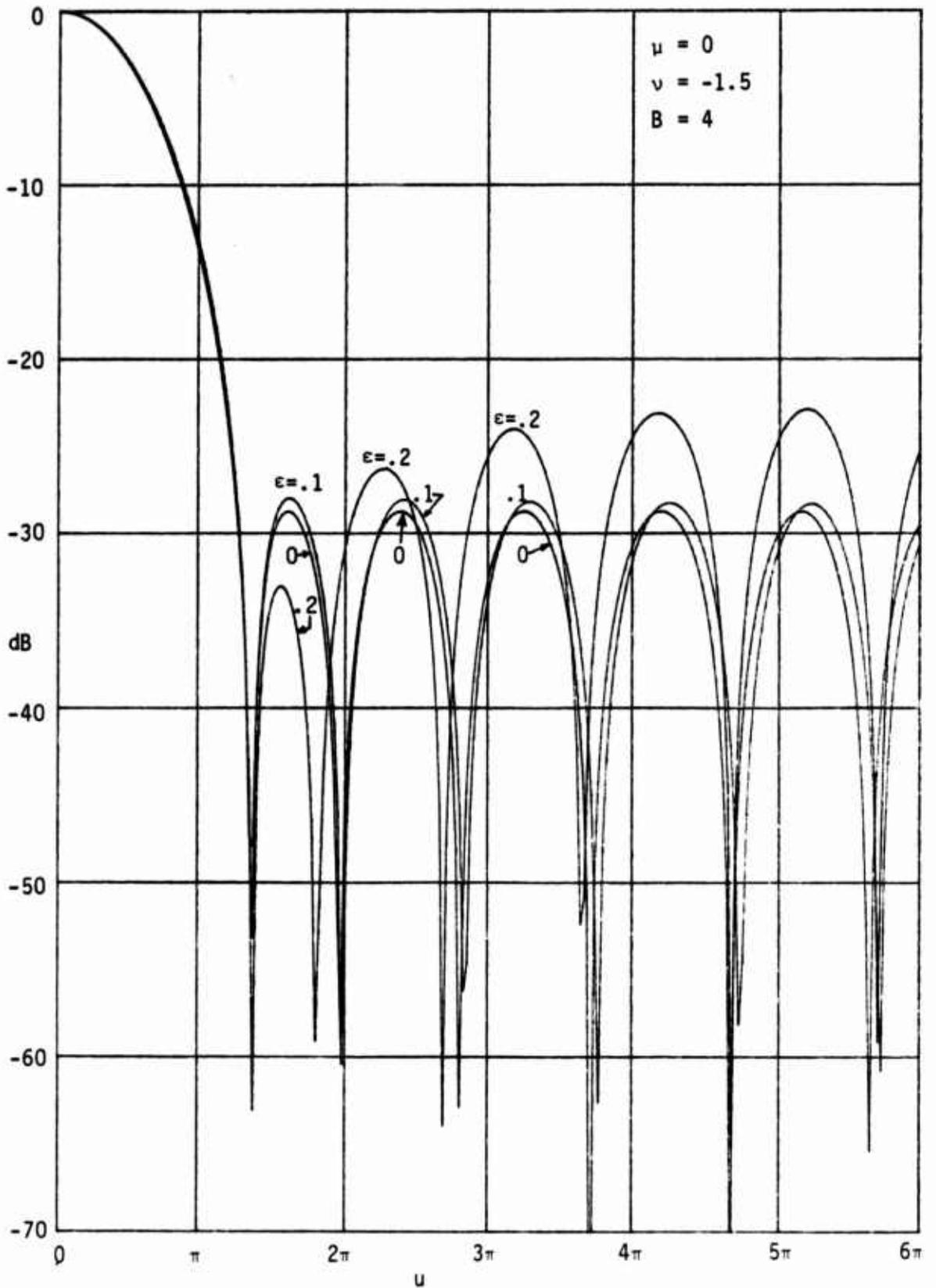
Figure 6.   Pattern of Circular Array for Weighting (36)

USE OF NEUTRALIZERS

An alternative smoother approximation to $f(x)$ and $g(x)$ than afforded by figures 1 and 2 is by the use of neutralizers; see ref. 5, section 3.3. Consider the neutralizer $n_\epsilon(x)$ and its derivative shown in figure 7; the neutralizer is 0 at $x = a$, and 1 at $x = b$. Furthermore, it has derivatives
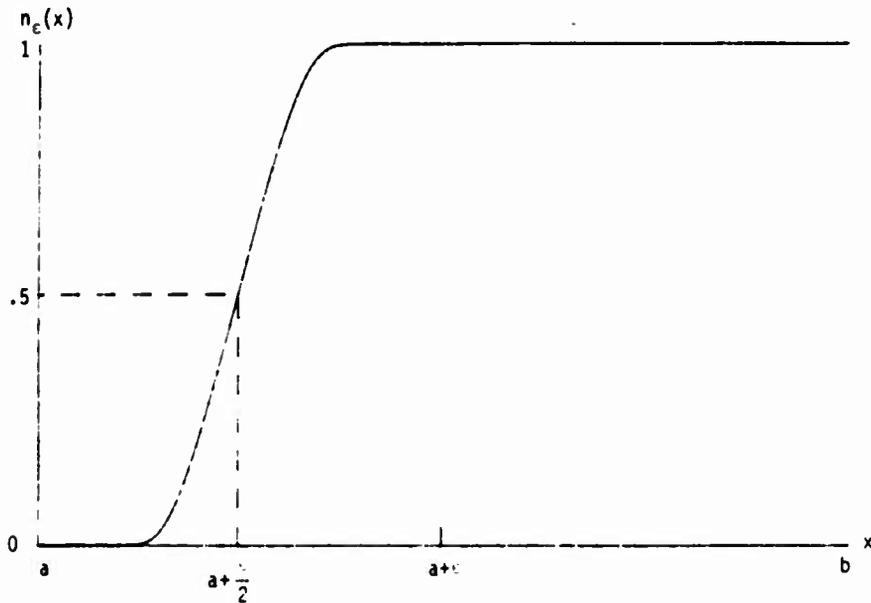
Figure 7A.   Function $n_\epsilon(x)$
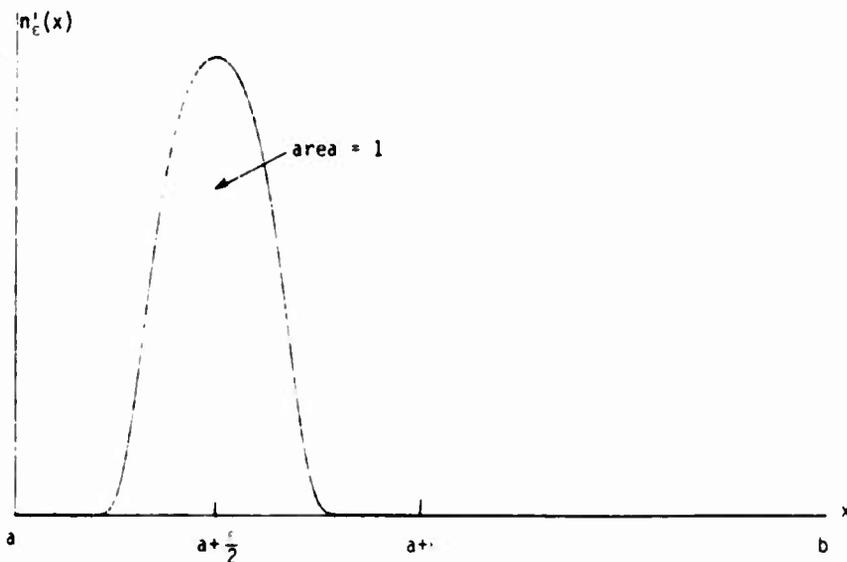
Figure 7B.   Function $n'_\epsilon(x)$

Figure 7.   Neutralizer $n_\epsilon(x)$ and its Derivative

of all orders, all of which are zero at end points a and b. The positive parameter ϵ characterizes the critical point, $x_c = a + \frac{\epsilon}{2}$, where the neutralizer is 1/2; this point x_c will approach a as $\epsilon \to 0+$. The neutralizer has completed its transition to 1 by the value $x = a + \epsilon$. $n_\epsilon(x)$ will approach 1 for all $x > a$, as $\epsilon \to 0$.

The approximation we take to $f(x)$ is the product

$$f_\epsilon(x) \equiv f(x) \, n_\epsilon(x) \; , \tag{43}$$

and the corresponding approximation to generalized function (2) is the smooth function

$$g_\epsilon(x) = f_\epsilon'(x) = f'(x) \, n_\epsilon(x) + f(x) \, n_\epsilon'(x) \; . \tag{44}$$

This approach is similar to the regular sequence of good functions used to define a generalized function in ref. 3, pages 16-17. A representative example is depicted in figure 8. As before, the area under approximation $g_\epsilon(x)$ is independent of ϵ:

$$\int_a^b dx \, g_\epsilon(x) = \left[f_\epsilon(x) \right]_a^b = f_\epsilon(b) = f(b) \; . \tag{45}$$

The result of applying $g_\epsilon(x)$ to function $h(x)$ is now

20
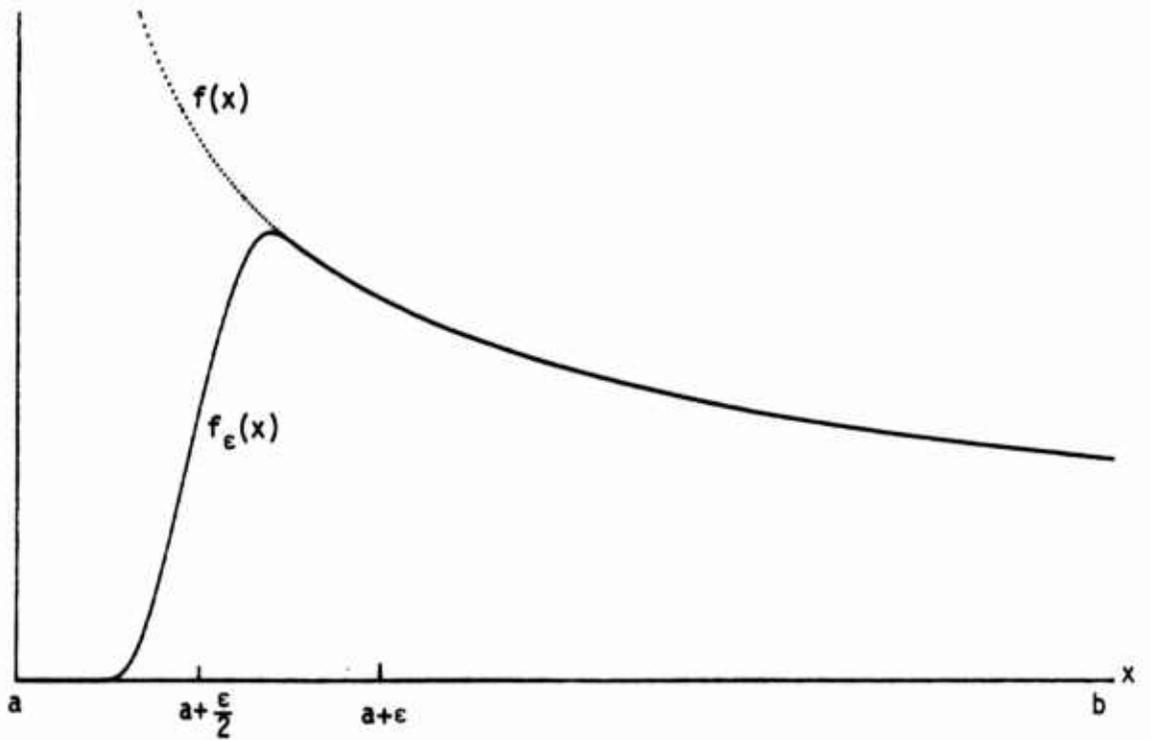
Figure 8A.   Function $f_\epsilon(x)$
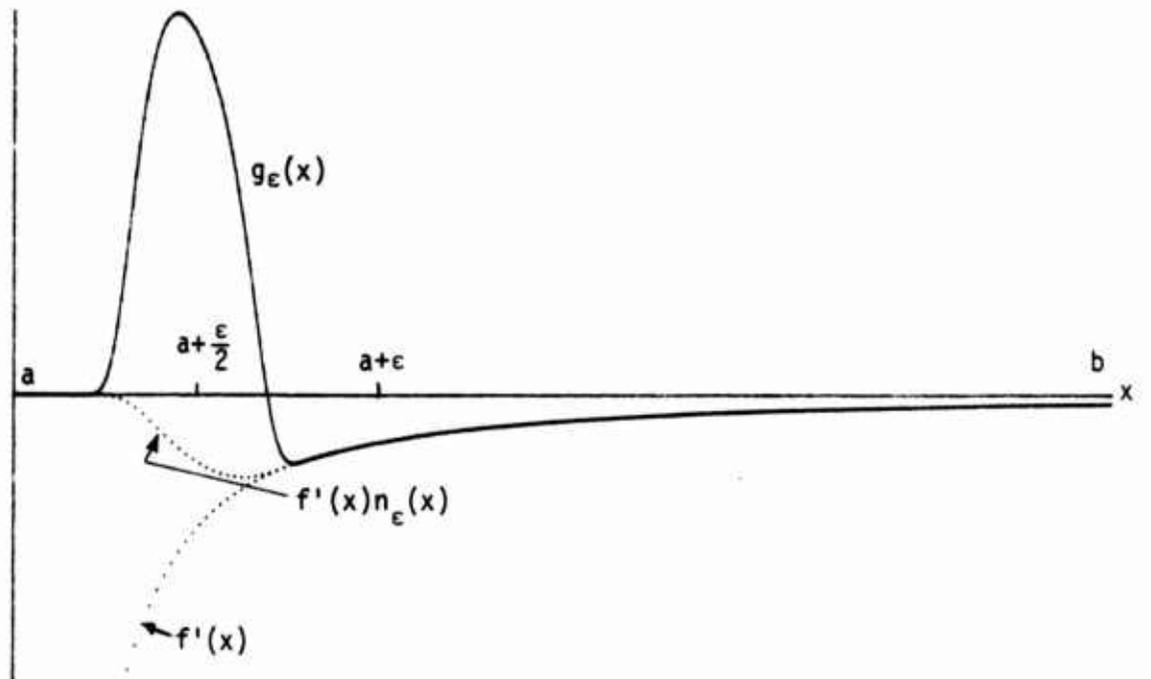
Figure 8B.   Function $g_\epsilon(x)$

Figure 8.   Approximation to Generalized Function via Neutralizer

$$I_\epsilon = \int_a^b dx\, g_\epsilon(x)\, h(x) = \int_a^b dx\, f_\epsilon{}'(x)\, h(x)$$

$$= \left[f_\epsilon(x)\, h(x) \right]_a^b - \int_a^b dx\, f_\epsilon(x)\, h'(x)$$

$$= f(b)\, h(b) - \int_a^b dx\, f(x)\, n_\epsilon(x)\, h'(x) \,, \tag{46}$$

where we have utilized the properties of the neutralizer. Equation (46) now replaces (8). In the limit of $\epsilon \to 0$, (46) yields

$$I = \lim_{\epsilon \to 0} I_\epsilon = f(b)\, h(b) - \int_a^b dx\, f(x)\, h'(x) \,, \tag{47}$$

in agreement with (10).

A limiting case of figure 7 is a step-function at $x = a + \frac{\epsilon}{2}$; i.e.,

$$n_\epsilon(x) = U(x - a - \frac{\epsilon}{2}) \,. \tag{48}$$

Then

$$n_\epsilon{}'(x) = \delta(x - a - \frac{\epsilon}{2}) \tag{49}$$

and (44) yields

$$g_\epsilon(x) = f'(x)\, U(x - a - \frac{\epsilon}{2}) + f(a + \frac{\epsilon}{2})\, \delta(x - a - \frac{\epsilon}{2}) \,. \tag{50}$$

This is almost identical to (6), which has its impulse located at $x = a$ rather than at $x = a + \frac{\epsilon}{2}$.

APPLICATION TO ARRAY

The starting point is (32), again, for the required weighting. The function $f(x)$ is identified as in (34), and we get, from (33) and (44),

$$-x w_\epsilon(x) = g_\epsilon(x) = f'(x) n_\epsilon(x) + f(x) n_\epsilon'(x). \qquad (51)$$

There follows, from (34) and ref. 4, eq. 9.6.28,

$$f'(x) = -x \left(\frac{\sqrt{1-x^2}}{B} \right)^\nu I_\nu \left(B\sqrt{1-x^2} \right) \quad \text{for } 0 < x < 1 . \qquad (52)$$

Substitution of (34) and (52) in (51) yields, for the approximate weighting,

$$w_\epsilon(x) = \left(\frac{\sqrt{1-x^2}}{B} \right)^\nu I_\nu \left(B\sqrt{1-x^2} \right) n_\epsilon(x)$$

$$- \frac{1}{x} \left(\frac{\sqrt{1-x^2}}{B} \right)^{\nu+1} I_{\nu+1} \left(B\sqrt{1-x^2} \right) n_\epsilon'(x) \quad \text{for } 0 < x < 1 . \qquad (53)$$

However, the neutralizer in this case must be chosen to be 1 at $x = 0$, and 0 at $x = 1$; that is, it is a reflected version of figure 7A. For our purposes, it is not necessary for the neutralizer to have derivatives of all orders. Rather, we select $n_\epsilon(x)$ so that $n_\epsilon''(x)$ is continuous for all x, and such that $n_\epsilon'(x)$ and $n_\epsilon''(x)$ are zero at the edges of the transition region $(1-\epsilon, 1)$; see figure 9. Here, letting

$$y = \frac{x - 1 + \frac{\epsilon}{2}}{\epsilon} , \qquad (54)$$

23
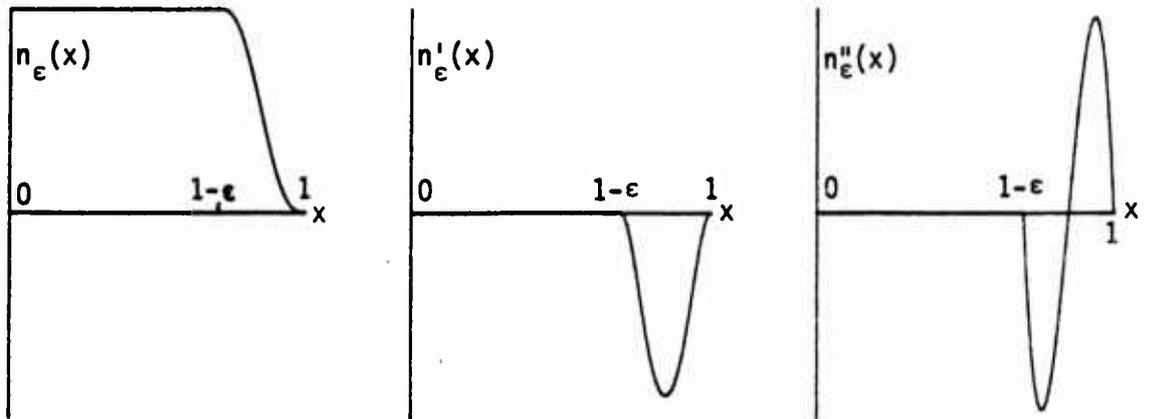
Figure 9. Neutralizer (55) and its Derivatives

we take

$$n_\epsilon(x) = \frac{1}{2} - \frac{1}{8}\left(15y - 40y^3 + 48y^5\right) \quad \text{for } 1-\epsilon \leq x \leq 1, \qquad (55)$$

and 1 for $0 \leq x \leq 1-\epsilon$. Then

$$n'_\epsilon(x) = -\frac{15}{8\epsilon}\left(1-8y^2+16y^4\right) \quad \text{for } 1-\epsilon \leq x \leq 1, \qquad (56)$$

and zero otherwise. Also,

$$n''_\epsilon(x) = \frac{30}{\epsilon^2}\left(y-4y^3\right) \text{ for } 1-\epsilon \leq x \leq 1, \qquad (57)$$

and zero otherwise. From (55)-(57), there follows

$$\left.\begin{array}{l} n_\epsilon(x) \sim \dfrac{10}{\epsilon^3}(1-x)^3 \\[2ex] n_\epsilon'(x) \sim -\dfrac{30}{\epsilon^3}(1-x)^2 \\[2ex] n_\epsilon''(x) \sim \dfrac{60}{\epsilon^3}(1-x) \end{array}\right\} \qquad \text{as } x \to 1- \, . \qquad (58)$$

Combining this with the behavior of the Bessel function near zero argument (ref. 4, eq. 9.6.7), we find that both terms of the approximate weighting $w_\epsilon(x)$ in (53) are proportional to $(1-x)^{3+\nu}$ as x approaches 1. Thus if $\nu \geq -2$, the approximate weighting will approach zero at least linearly at x = 1. An example of $w_\epsilon(x)$ for $\nu = -1.5$ and B = 4 is given in figure 10
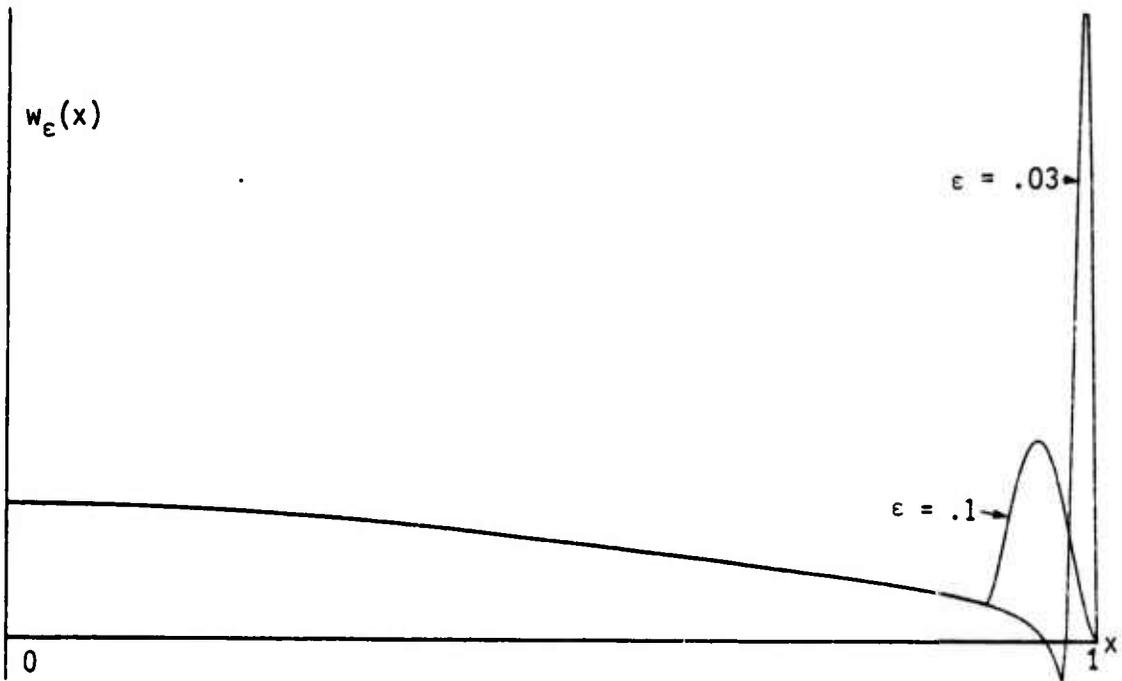
Figure 10. Approximate Weighting $w_\epsilon(x)$ for $\nu = -1.5$, B = 4

for two different selections of ϵ. The large positive spike near $x = 1$ is very pronounced for small ϵ.

The array voltage response corresponding to approximate weighting $w_\epsilon(x)$ in (53) is given by (38) as

$$v_\epsilon(u) = \int_0^1 dx \; x \left(\frac{x}{u}\right)^\mu \; J_\mu(ux) \; w_\epsilon(x)$$

$$= \int_0^1 dx \; x^{2\mu+1} \; \mathcal{J}_\mu(ux) \; w_\epsilon(x) \; , \qquad (59)$$

where we used (40). A program for the evaluation of (59) (along with (53)-(56)) is presented in appendix B. Sample responses are plotted in figure 11 for a spherical array and in figure 12 for a circular array. Comparison with the corresponding results in figures 5 and 6 reveals that the impulsive weighting of (36) yields a better approximation to the ideal pattern ($\epsilon = 0$) than the smoother weighting of (53). That is, ϵ must be chosen smaller in (53) than in (36), in order to realize approximately the same voltage response pattern in the first few sidelobes.
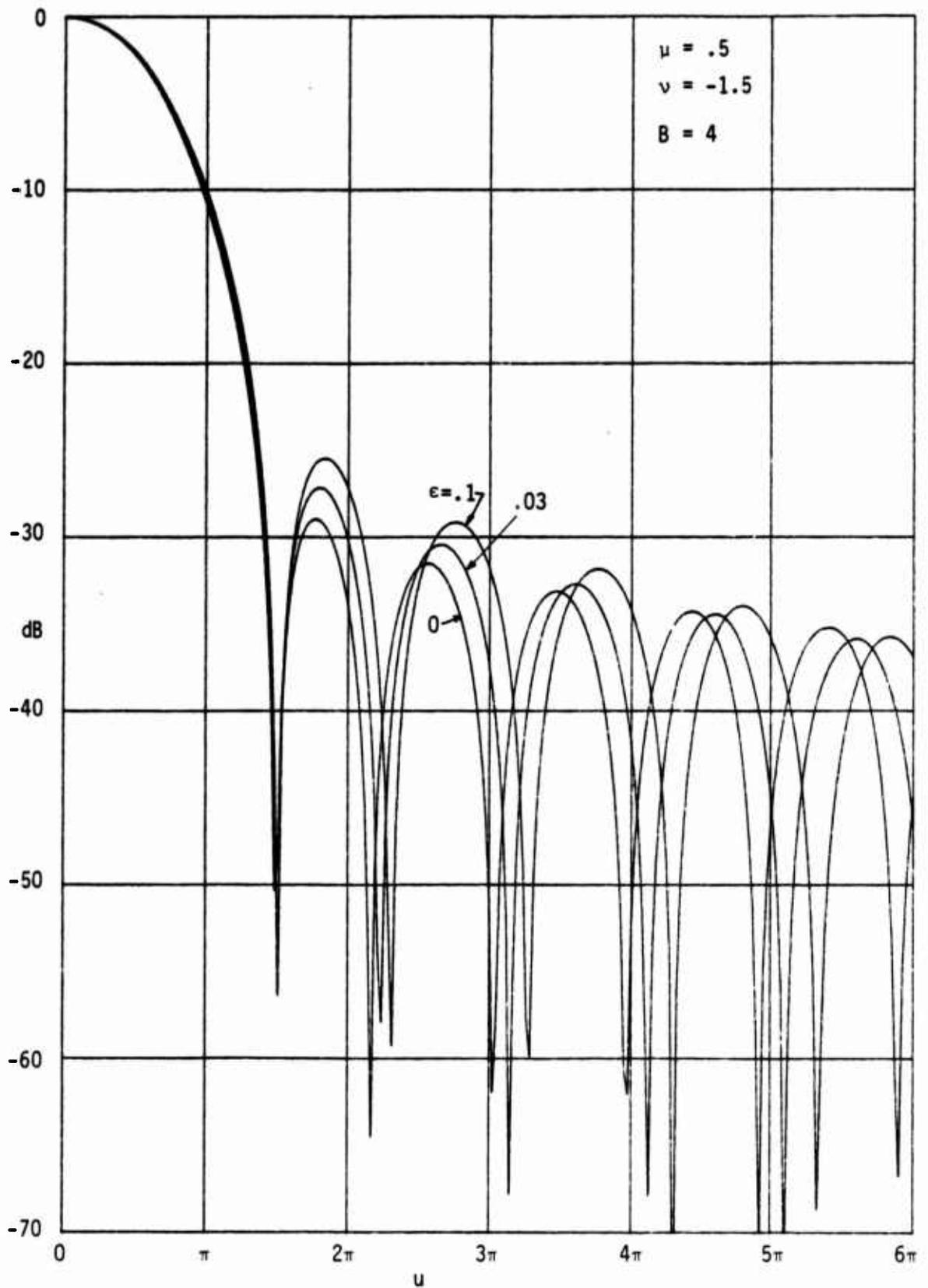
Figure 11.   Pattern of Spherical Array for Weighting (53)
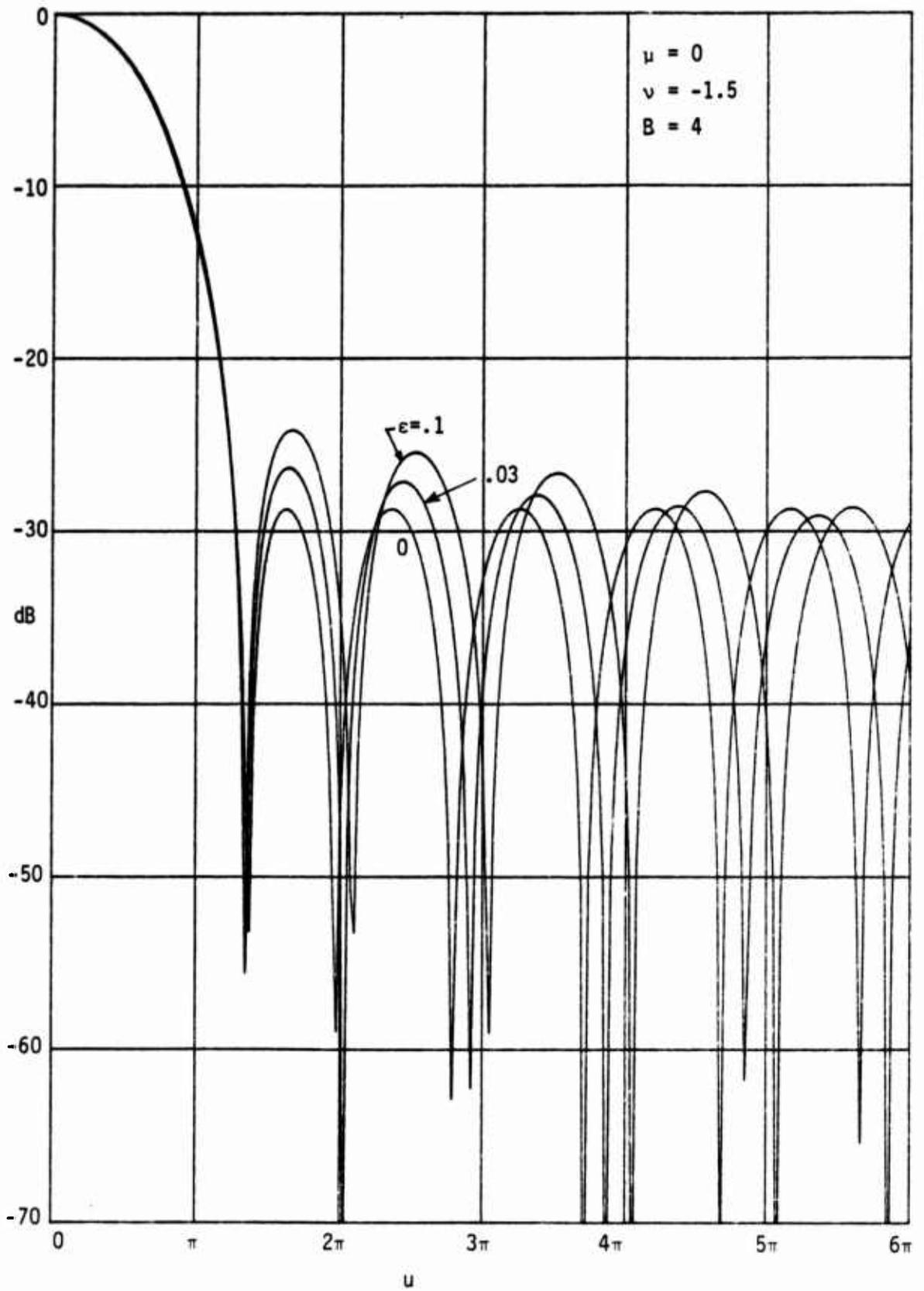
Figure 12.   Pattern of Circular Array for Weighting (53)

SUMMARY

We have indicated how approximations to some generalized functions generated by the process of differentiation may be realized, and then have applied the procedure to the approximation of the weighting required to realize the ideal response patterns of arrays in several dimensions. Two examples, one impulsive and the other smooth, were used for circular and spherical arrays, and the resultant approximate patterns were plotted for different choices of the parameter ϵ governing the transition region near the singularity of the generalized function. How small ϵ must be chosen depends on the form of the approximate weighting and the desired closeness to the specified pattern.

Appendix A
VOLTAGE RESPONSE PATTERN FOR WEIGHTING (36)

With the shorthand notation

$$f_\nu(x) \equiv \left(\frac{\sqrt{1-x^2}}{B}\right)^\nu I_\nu\left(B\sqrt{1-x^2}\right) \quad \text{for } 0 < x < 1, \tag{A-1}$$

approximate weighting (36) takes the form

$$w_\epsilon(x) = f_{\nu+1}(1-\epsilon)\, \delta(x-1) + f_\nu(x)\, U(1-\epsilon-x) \quad \text{for } 0 < x. \tag{A-2}$$

Substitution of (A-2) in (38) yields, upon use of (40), the corresponding voltage response

$$v_\epsilon(u) = f_{\nu+1}(1-\epsilon)\, \mathcal{J}_\mu(u) + \int_0^{1-\epsilon} dx\; x^{2\mu+1} \mathcal{J}_\mu(ux)\, f_\nu(x) \tag{A-3}$$

for any μ; this is the relation programmed in ref. 2, appendix F.

For small ϵ (the case of most interest), the cusp of $f_\nu(x)$ in (A-3) at $x = 1$ causes numerical integration problems. We can alleviate this problem by integrating by parts on (A-3), using

$$U = x^{2\mu} \mathcal{J}_\mu(ux), \quad V = -f_{\nu+1}(x), \tag{A-4}$$

to get

$$v_\epsilon(u) = f_{\nu+1}(1-\epsilon)\left[\mathcal{J}_\mu(u) - (1-\epsilon)^{2\mu} \mathcal{J}_\mu(u(1-\epsilon))\right]$$

$$+ \int_0^{1-\epsilon} dx\; x^{2\mu-1} \mathcal{J}_{\mu-1}(ux)\, f_{\nu+1}(x) \quad \text{for } \mu > 0. \tag{A-5}$$

The function $f_{\nu+1}(x)$ is integrable at $x = 1$ if $\nu > -2$.

The completed integral over $(0,1)$ in (A-5) can be evaluated in closed form (ref. 2), leading to

$$v_\epsilon(u) = \mathcal{J}_{\mu+\nu+1}\left(\sqrt[4]{u^2 - B^2}\right)$$

$$+ f_{\nu+1}(1-\epsilon)\left[\mathcal{J}_\mu(u) - (1-\epsilon)^{2\mu}\mathcal{J}_\mu(u(1-\epsilon))\right]$$

$$- \int_{1-\epsilon}^{1} dx \; x^{2\mu-1} \mathcal{J}_{\mu-1}(ux) \; f_{\nu+1}(x) \; , \tag{A-6}$$

which is advantageous because the interval $(1-\epsilon, 1)$ is small for small ϵ. Finally, the cusp at $x = 1$ can be eliminated for $\nu \geq -3/2$, by changing the variable of integration according to $x = \sin t$, getting

$$v_\epsilon(u) = \mathcal{J}_{\mu+\nu+1}\left(\sqrt[4]{u^2 - B^2}\right)$$

$$+ f_{\nu+1}(1-\epsilon)\left[\mathcal{J}_\mu(u) - (1-\epsilon)^{2\mu}\mathcal{J}_\mu(u(1-\epsilon))\right]$$

$$- \int_{A}^{\pi/2} dt \; (\sin t)^{2\mu-1} (\cos t)^{2\nu+3} \mathcal{J}_{\mu-1}(u \sin t) \, \mathcal{U}_{\nu+1}(B \cos t) \; , \tag{A-7}$$

where $A = \arcsin(1-\epsilon)$, and

$$\mathcal{U}_\alpha(z) \equiv \frac{I_\alpha(z)}{z^\alpha} \; . \tag{A-8}$$

Form (A-7) is good numerically for small ϵ and any value of μ. There is no cusp at $t = \pi/2$ if $\nu \geq -3/2$. A program for (A-7) follows, where ϵ, μ, ν, and B are arbitrary ($\epsilon \geq 0$, $\nu \geq -3/2$, $B \geq 0$).

```
10      Eps=.1          E               ! Pattern Ve(u) via (A-7)
20      Mu=0            μ
30     ·Nu=-1.5         ν
40      Bc=4            B               ! B in (26)
50      OUTPUT 0;"Eps =";Eps;"  Mu = ";Mu;"  Nu = ";Nu;"  Bc =";Bc
60      DIM Ve(0:480)
70      COM U,Bc,M1,N1,M2,N2
80      M1=Mu-1
90      N1=Nu+1
100     M2=2*Mu-1
110     N2=2*Nu+3
120     Alpha=Mu+Nu+1
130     T=2*Eps-Eps^2
140     IF Eps=0 THEN 160
150     F1e=T^N1*FNInuxnu(N1,Bc*SQR(T))      ! for line 2 of (A-7)
160     E1=1-Eps
170     Te=E1^(2*Mu)
180     A=ASN(E1)
190     B=PI/2
200     FOR Iu=0 TO 480
210     U=Iu*PI/40
220     Sq=SQR(ABS(U*U-Bc*Bc))
230     IF U<=Bc THEN T1=FNInuxnu(Alpha,Sq)      ! line 1 of (A-7)
240     IF U>Bc THEN T1=FNJnuxnu(Alpha,Sq)
250     IF Eps>0 THEN 290
260     T3=T1
270     V=0
280     GOTO 480
290     T2=FNJnuxnu(Mu,U)-Te*FNJnuxnu(Mu,U*E1)   ! for line 2 of (A-7)
300     T3=T1+F1e*T2
310     S=(FNS(A)+FNS(B))*.5
320     N=2
330     H=(B-A)*.5
340     F=(B-A)/3
350     Vo=9E99
360     T=0
370     FOR K=1 TO N-1 STEP 2
380     T=T+FNS(A+H*K)
390     NEXT K
400     S=S+T
410     V=(S+T)*F                                ! line 3 of (A-7)
420     IF ABS(V-Vo)<=ABS(V)*1E-5 THEN 480
430     Vo=V
440     N=N*2
450     H=H*.5
460     F=F*.5
470     GOTO 360
480     Ve(Iu)=T3-V                              ! Voltage Response (A-7)
490     PRINT Iu,Ve(Iu)
500     NEXT Iu
510     PLOTTER IS "GRAPHICS"
520     GRAPHICS
530     SCALE 0,480,-70,0
540     GRID 40,10
550     PENUP
560     FOR Iu=0 TO 480
570     PLOT Iu,20*LGT(ABS(Ve(Iu)/Ve(0)))
580     NEXT Iu
590     PENUP
600     END
610     !
```

33

```
620    DEF FNS(T)                    ! integrand of (A-7)
630    COM U,Bc,M1,N1,M2,N2
640    St=SIN(T)
650    Ct=COS(T)
660    T1=FNJnuxnu(M1,U*St)
670    T2=FNInuxnu(N1,Bc*Ct)
680    RETURN St^M2*Ct^N2*T1*T2
690    FNEND
700    !
710    DEF FNGamma(X) !  Gamma(X)  via HART, page 279, #5231
720    N=INT(X)
730    R=X-N
740    IF (N>0) OR (R<>0) THEN 770
750    PRINT "FNGamma(X) IS NOT DEFINED FOR X = ";X
760    STOP
770    IF R>0 THEN 800
780    Gamma2=1
790    GOTO 840
800    P=3.36954359131+R*(1.09850630453+R*(.142928007149+R*3.93013464186E-2)
810    P=43.9410209139+R*(22.9680800836+R*(12.3021698112+R*P))
820    Q=43.9410209191+R*(4.39050474596-R*(7.15075063299-R))
830    Gamma2=P/Q                    ! Gamma(2+R) for 0 < R < 1
840    IF N>2 THEN 880
850    IF N<2 THEN 930
860    Gamma=Gamma2
870    GOTO 980
880    Gamma=Gamma2
890    FOR K=1 TO N-2
900    Gamma=Gamma*(X-K)
910    NEXT K
920    GOTO 980
930    R=1
940    FOR K=0 TO 1-N
950    R=R*(X+K)
960    NEXT K
970    Gamma=Gamma2/R
980    RETURN Gamma
990    FNEND
1000   !
```

```
1010   DEF FNJnuxnu(Nu,X)          !  Jnu(x)/x^nu
1020   IF ABS(X)<1 THEN 1130
1030   A=.797884560803
1040   IF Nu=0 THEN RETURN FNJo(X)
1050   IF Nu=.5 THEN RETURN A*SIN(X)/X
1060   IF Nu=-.5 THEN RETURN A*COS(X)
1070   IF Nu=1 THEN RETURN FNJ1(X)/X
1080   IF Nu=-1 THEN RETURN -FNJ1(X)*X
1090   IF Nu=1.5 THEN RETURN A*(SIN(X)-X*COS(X))/X^3
1100   IF Nu=-1.5 THEN RETURN -A*(X*SIN(X)+COS(X))
1110   IF Nu=2 THEN RETURN (2*FNJ1(X)-X*FNJo(X))/X^3
1120   IF Nu=-2 THEN RETURN (2*FNJ1(X)-X*FNJo(X))*X
1130   A=Nu
1140   IF (INT(A)<>A) OR (A>=0) THEN 1160
1150   K=A=-Nu
1160   S=T=1/(2^A*FNGamma(A+1))
1170   R=-.25*X*X
1180   Big=ABS(S)
1190   FOR N=1 TO 100
1200   T=T*R/(N*(N+A))
1210   S=S+T
1220   Big=MAX(Big,ABS(S))
1230   IF ABS(T)<=1E-11*ABS(S) THEN 1270
1240   NEXT N
1250   PRINT "100 TERMS IN FNJnuxnu(Nu,X) AT ";Nu;X
1260   PAUSE
1270   D=12-LGT(ABS(Big/S))   ! NO. OF SIGNIF. DIGITS
1280   IF K>0 THEN S=S*(4*R)^K
1290   RETURN S
1300   FNEND
1310   !
1320   DEF FNJo(X)     ! Jo(X) via 9.4.1 & 9.4.3
1330   Y=ABS(X)
1340   IF Y>3 THEN 1390
1350   T=Y*Y/9
1360   Jo=.0444479-T*(.0039444-T*.00021)
1370   Jo=1-T*(2.2499997-T*(1.2656208-T*(.3163866-T*Jo)))
1380   GOTO 1450
1390   T=3/Y
1400   Jo=9.512E-5-T*(.00137237-T*(.00072805-T*.0014475))
1410   Jo=.79788456-T*(7.7E-7+T*(.00552740+T*Jo))
1420   S=.00262573-T*(.00054125+T*(.00029333-T*.00013558))
1430   T=Y-.78539816-T*(.04166397+T*(3.954E-5-T*S))
1440   Jo=Jo*COS(T)/SQR(Y)
1450   RETURN Jo
1460   FNEND
1470   !
1480   DEF FNJ1(X)     ! J1(X) via 9.4.4 & 9.4.6
1490   Y=ABS(X)
1500   IF Y>3 THEN 1550
1510   T=Y*Y/9
1520   J1=.00443319-T*(.00031761-T*.00001109)
1530   J1=X*(.5-T*(.56249985-T*(.21093573-T*.03954289-T*J1)))
1540   GOTO 1610
1550   T=3/Y
1560   J1=.00017105-T*(.00249511-T*(.00113653-T*.00020033))
1570   J1=.79788456+T*(1.56E-6+T*(.01659667+T*J1))
1580   S=.00637879-T*(.00074348+T*(.00079824-T*.00029166))
1590   T=Y-2.35619449+T*(.12499612+T*(5.65E-5-T*S))
1600   J1=SGN(X)*J1*COS(T)/SQR(Y)
1610   RETURN J1
1620   FNEND
1630   !
```

35

```
1640    DEF FNInuxnu(Nu,X)              !   Inu(x)/x^nu
1650    IF ABS(X)<1 THEN 1770
1660    A=.398942280401
1670    E=EXP(X)
1680    IF Nu=0 THEN RETURN FNIo(X)
1690    IF Nu=.5 THEN RETURN A*(E-1/E)/X
1700    IF Nu=-.5 THEN RETURN A*(E+1/E)
1710    IF Nu=1 THEN RETURN FNI1(X)/X
1720    IF Nu=-1 THEN RETURN FNI1(X)*X
1730    IF Nu=1.5 THEN RETURN A*((X-1)*E+(X+1)/E)/X^3
1740    IF Nu=-1.5 THEN RETURN A*((X-1)*E-(X+1)/E)
1750    IF Nu=2 THEN RETURN (X*FNIo(X)-2*FNI1(X))/X^3
1760    IF Nu=-2 THEN RETURN (X*FNIo(X)-2*FNI1(X))*X
1770    A=Nu
1780    IF (INT(A)<>A) OR (A>=0) THEN 1800
1790    K=A=-Nu
1800    S=T=1/(2^A*FNGamma(A+1))
1810    R=.25*X*X
1820    Big=ABS(S)
1830    FOR N=1 TO 100
1840    T=T*R/(N*(N+A))
1850    S=S+T
1860    Big=MAX(Big,ABS(S))
1870    IF ABS(T)<=1E-11*ABS(S) THEN 1910
1880    NEXT N
1890    PRINT "100 TERMS IN FNInuxnu(Nu,X) AT ";Nu;X
1900    PAUSE
1910    D=12-LGT(ABS(Big/S))  ! NO. OF SIGNIF. DIGITS
1920    IF K>0 THEN S=S*(4*R)^K
1930    RETURN S
1940    FNEND
1950    !
1960    DEF FNIo(X)      ! Io(X) via 9.8.1 & 9.8.2
1970    Y=ABS(X)
1980    IF Y>3.75 THEN 2030
1990    T=Y*Y/14.0625
2000    Io=.2659732+T*(.0360768+T*.0045813)
2010    Io=1+T*(3.5156229+T*(3.0899424+T*(1.2067492+T*Io
2020    GOTO 2070
2030    T=3.75/Y
2040    Io=.00916281-T*(.02057706-T*(.0263553T-T+.0164T+ 2-T*.00392377))
2050    Io=.39894228+T*(.01328592+T*(.00225319-T*.00157565-T+Io)
2060    Io=Io*EXP(Y)/SQR(Y)
2070    RETURN Io
2080    FNEND
2090    !
2100    DEF FNI1(X)      ! I1(X) via 9.8.3 & 9.8.4
2110    Y=ABS(Y)
2120    IF Y>3.75 THEN 2170
2130    T=Y*Y/14.0625
2140    I1=.02658733+T*(.00301532+T*.00032411)
2150    I1=X*(.5+T*(.87890594+T*(.51498869+T*(.15084934+T*I1
2160    GOTO 2210
2170    T=3.75/Y
2180    I1=.01031555-T*(.02282967-T*(.02895312-T*.01787654-T-.00420059))
2190    I1=.39894228-T*(.03988024+T*(.00362018-T*.00163801-T*I1))
2200    I1=SGN(X)*I1*EXP(Y)/SQR(Y)
2210    RETURN I1
2220    FNEND
```

Appendix B
PROGRAM FOR THE EVALUATION OF (59)

The integral of interest is, from (59),

$$v_\epsilon(u) = \int_0^1 dx \, \mathcal{J}_\mu(ux) \, x^{2\mu+1} \, w_\epsilon(x) \, . \tag{B-1}$$

We approximate this integral by sampling at increment $\Delta = \dfrac{1}{N_x}$ and using Simpson's rule. There follows

$$v_\epsilon(u) \cong \frac{\Delta}{3} \sum_{n=0}^{N_x} \bar{w}_n \, \mathcal{J}_\mu(nu\Delta) \, (n\Delta)^{2\mu+1} \, w_\epsilon(n\Delta) \, , \tag{B-2}$$

where

$$\left\{ \bar{w}_n \right\} = 1, \, 4, \, 2, \, 4, \, \ldots, \, 4, \, 2, \, 4, \, 1 \, . \tag{B-3}$$

The program below evaluates (B-2). Values of $x^{2\mu+1} \, w_\epsilon(x)$ are stored in array Xmuwe.

```
10      Eps=.1          ! TRANSITION PARAMETER ε
20      Mu=0            ! DIMENSIONALITY PARAMETER μ
30      Nu=-1.5         ! WEIGHTING PARAMETER ν
40      Bc=4            ! WEIGHTING PARAMETER B
50      Nx=2^8          ! NUMBER OF INCREMENTS IN X
60      Niu=240         ! NUMBER OF INCREMENTS IN U
70      DIM Xmuwe(0:1024),Ve(0:240)
80      REDIM Xmuwe(0:Nx),Ve(0:Niu)
90      Mu21=Mu*2+1
100     Nu1=Nu+1
110     Del=1/Nx
120     Xmuwe(Nx)=0
130     FOR I=0 TO Nx-1
140     X=I*Del
150     R2=1-X*X
160     Rs=SQR(R2)
170     Br=Bc*Rs
180     Iv=R2^Nu*FNInuxnu(Nu,Br)
190     IF X>1-Eps THEN 220
200     Xmuwe(I)=X^Mu21*Iv
210     GOTO 260
220     Iv1=R2^Nu1*FNInuxnu(Nu1,Br)
230     Ne=FNNeut(Eps,X)
240     Nep=FNNeutp(Eps,X)
250     Xmuwe(I)=X^Mu21*(Iv*Ne-Iv1*Nep/X)
260     NEXT I
270     FOR Iu=0 TO Niu
280     U=Iu/Niu*6*PI
290     Ud=U*Del
300     T=Xmuwe(0)+FNJnuxnu(Mu,U)*Xmuwe(Nx)
310     So=Se=0
320     FOR Ns=1 TO Nx-1 STEP 2
330     So=So+FNJnuxnu(Mu,Ns*Ud)*Xmuwe(Ns)
340     NEXT Ns
350     FOR Ns=2 TO Nx-2 STEP 2
360     Se=Se+FNJnuxnu(Mu,Ns*Ud)*Xmuwe(Ns)
370     NEXT Ns
380     Ve(Iu)=T+4*So+2*Se
390     PRINT Iu,Ve(Iu)
400     NEXT Iu
410     PLOTTER IS "GRAPHICS"
420     GRAPHICS
430     SCALE 0,Niu,-70,0
440     GRID Niu/6,10
450     PENUP
460     T=20*LGT(Ve(0))
470     FOR Iu=0 TO Niu
480     Y=20*LGT(ABS(Ve(Iu)))-T
490     PLOT Iu,Y
500     NEXT Iu
510     PENUP
520     END
530     !
```

```
540    DEF FNNeut(E,X)
550    IF X<=1-E THEN RETURN 1
560    Y=(X-1+.5*E)/E
570    T=Y*Y
580    T=15-T*(40-T*48)
590    RETURN .5-.125*Y*T
600    FNEND
610    !
620    DEF FNNeutp(E,X)
630    IF X<=1-E THEN RETURN 0
640    Y=(X-1+.5*E)/E
650    T=Y*Y
660    T=1-T*(3-16*T)
670    RETURN -1.875*T/E
680    FNEND
690    !
```

ALL OTHER FUNCTIONS ARE LISTED IN APPENDIX A.

Performance of Three Averaging Methods, for Various Distributions

A. H. Nuttall

ABSTRACT

The performance of three averaging methods, namely the sample median, the sample arithmetic mean, and the sample geometric mean, are analyzed in terms of their bias, variance, and mean square error. The bias and variance are numerically evaluated for various parent distributions and plotted versus the number, N, of data points employed in the sample statistics. Also, the limiting behaviors, as N increases without limit, are derived. It is found that the best averaging method is very dependent upon the distribution of the data, with the sample median being favored for data with occasional large out-liers.

PERFORMANCE OF THREE AVERAGING METHODS,
FOR VARIOUS DISTRIBUTIONS

by

Albert H. Nuttall
Naval Underwater Systems Center
New London, CT 06320 USA

ABSTRACT

The performance of three averaging methods, namely the sample median, the sample arithmetic mean, and the sample geometric mean, are analyzed in terms of their bias, variance, and mean square error. The bias and variance are numerically evaluated for various parent distributions and plotted versus the number, N, of data points employed in the sample statistics. Also, the limiting behaviors, as N increases without limit, are derived. It is found that the best averaging method is very dependent upon the distribution of the data, with the sample median being favored for data with occasional large out-liers.

INTRODUCTION

Estimation of average properties, such as the average power in a particular angular sector and/or frequency bin is often accomplished by taking N independent measurements of such data and calculating a simple arithmetic average. However, when the desired process is subject to random fade-outs or occasional large out-liers, this sample arithmetic mean (SAM) is severely perturbed, and alternative averaging methods should be considered. Two possible candidates are the sample median (SMD) and the sample geometric mean (SGM); these nonlinear processors of the available data have the potential of suppressing the deleterious effects mentioned above. Here we investigate* the performance of all three of these averaging methods in terms of the number, N, of independent data points employed in the pertinent average, and the parent distribution of the data. A wide variety of distributions are considered, some with parameters which allow for significantly different character and shapes of the governing probability functions.

---

*The basic analysis, derivations, and programs are given in Ref. 1.

1

DEFINITIONS

We have available N statistically-independent identically-distributed samples (random variables) x_1, x_2, ..., x_N from some parent population with cumulative distribution function $P(u) = \text{Prob}\{x<u\}$ and probability density function, PDF, $p(u) = P'(u)$. The SAM of the available measurements is

$$a(N) = \frac{1}{N}\left(x_1 + x_2 + \ldots + x_N\right);\tag{1}$$

the SMD is

$$q(N) = \text{middle value of } \left\{x_1, x_2, \ldots, x_N\right\}, \text{ for N odd};\tag{2}$$

and the SGM is (for non-negative random variables)

$$g(N) = \left(x_1 \, x_2 \ldots x_N\right)^{1/N} = \exp\left(\frac{\ln(x_1) + \ldots + \ln(x_N)}{N}\right)$$

$$= B^{\left(\frac{1}{A} \frac{A \log_B(x_1) + \ldots + A \log_B(x_N)}{N}\right)} \qquad \begin{array}{l}\text{for any base } B > 0 \\ \text{and scaling A.}\end{array}\tag{3}$$

The last form in (3) for base $B = 10$ goes under the name of dB averaging.

As N tends to infinity, the sample quantities above tend to definite (non-random) limits. In particular, as $N \rightarrow \infty$,

$$a(N) \rightarrow \text{arithmetic mean} = \int du \, u \, p(u);$$

$$q(N) \rightarrow \text{median} = u_{1/2}, \text{ where } P\left(u_{1/2}\right) = \frac{1}{2};$$

$$g(N) \rightarrow \text{geometric mean} = \exp\left(\int du \, \ln(u) \, p(u)\right);\tag{4}$$

where we drop the prefix 'sample' for these deterministic quantities. The last result in (4) follows from the exponential form of the SGM in (3). If the 1/2 in the median definition is replaced by r, we have for q(N) the sample quantile of order r (Ref. 2, page 181).

The three limiting quantities in (4) will generally not be equal. For example, for an exponential parent PDF

$$p(u) = \frac{1}{m} \exp\left(-\frac{u}{m}\right) \text{ for } u > 0,\tag{5}$$

2

we have

arithmetic mean = m;

median = m ln(2) = m .693;   (6)

geometric mean = me$^{-\gamma}$ = m .562.

Thus we define the bias of each of the sample statistics (1)-(3) as the difference between their mean value and their asymptotic value:

$$\text{bias} \{SAM\} = \overline{a(N)} - \text{mean};$$

$$\text{bias} \{SMD\} = \overline{q(N)} - \text{median};$$

$$\text{bias} \{SGM\} = \overline{g(N)} - \text{geometric mean}. \qquad (7)$$

By virtue of this definition, all three biases will tend to zero as $N \to \infty$; that is, all three estimators, (1)-(3), are asymptotically unbiased, <u>each with respect to its</u> desired value as given by (4), respectively.

It is then convenient to define a normalized bias, NB, for each sample statistic as

$$NB(N) = N \frac{\text{bias}}{\sigma}, \qquad (8)$$

where σ is the standard deviation of parent PDF p(u). The scale factor of N leads to a non-zero value of the normalized bias for large N, while the scale factor of σ is convenient in that it eliminates the dependence of the normalized bias on the absolute scale of the input data. For large N, (8) yields

$$\text{bias} \sim \sigma \frac{NB(\infty)}{N} \text{ as } N \to \infty; \qquad (9)$$

thus $NB(\infty)$ is an important measure of quality of the particular sample statistic under consideration.

The variances of sample statistics (1)-(3) are defined as

$$\text{var} \{SAM\} = \overline{a^2(N)} - \overline{a(N)}^2;$$

$$\text{var} \{SMD\} = \overline{q^2(N)} - \overline{q(N)}^2;$$

$$\text{var} \{SGM\} = \overline{g^2(N)} - \overline{g(N)}^2. \qquad (10)$$

3

Again, since these quantities tend to zero for large N, it is more convenient to define a normalized variance, NV, as

$$NV(N) = N \frac{variance}{\sigma^2} \quad . \tag{11}$$

Then we can state that

$$variance \sim \sigma^2 \frac{NV(\infty)}{N} \text{ as } N \rightarrow \infty; \tag{12}$$

thus $NV(\infty)$ is also an important measure of the quality of a particular sample statistic.

We present results here for NB(N) and NV(N), along with their asymptotic values at $N = \infty$, for a variety of parent distributions P(u). Additional results for the sample quantile with $r = .75$ and .9, and for the PDF, cumulative distribution function, characteristic function, cumulants, and moments of the various sample statistics are available in Ref. 1.

RESULTS

The first case we consider is the Gaussian PDF with mean m and variance σ^2. Since this random variable can go negative, the SGM is undefined. The SAM and SMD are unbiased for all N; thus NB(N) = 0 for all N, for this example. Results for the normalized variance are presented in Fig. 1 for the number of samples, N, between 1 and 51, for both the SAM and the SMD. The normalized variance for the SMD is computed only at odd values of N, indicated by an X, and straight lines drawn between these points for ease of association of values. It is seen that the variance of the SMD is always greater than that for the SAM, the limiting value, $NV(\infty)$, being $\pi/2$ for the SMD; see also Ref. 2, page 369. Observe that the parameters m and σ of this PDF have dropped out of this normalized plot.

It is worth pointing out here, and for similar results to follow, that although the curve for the SMD increases with N, that does not mean that the variance increases with N; rather, the normalizing factor of N in definition (11) causes this behavior. The actual (unnormalized) variance decreases monotonically with N, eventually being of order 1/N.

For a Rayleigh random variable, the results for the normalized bias are given in Fig. 2. The SAM is unbiased for all N, whereas the SGM and SMD are, of course, only asymptotically unbiased. The limiting values, NB(∞), for both of these latter sample statistics are given by analytically complicated expressions and are not repeated here, for the sake of brevity; they are indicated numerically by horizontal lines at the right edge of the figure. The corresponding results for the normalized variance are given in Fig. 3. They indicate that whereas the SGM has about the same stability as the SAM, the variance for the SMD is about 65% greater.

For an exponential PDF (as given by (5)), the normalized bias and variance results are presented in Figs. 4 and 5 respectively. The biases of the SGM and SMD are comparable, but we observe that the variance for the SGM is twice as small as that for the SAM and the SMD.

4

For a log-normal PDF,

$$p(u) = \frac{1}{\sigma_y\, u\, \sqrt{2\pi}} \exp\left[-\frac{\left(\ln(u) - m_y\right)^2}{2\sigma_y^2}\right] \quad \text{for } u > 0 , \tag{13}$$

the NB(N) and NV(N) results are independent of location parameter m_y, but they do depend on spread factor σ_y. This may be anticipated by plotting the PDF (13) for various values of σ_y and observing that the <u>shape</u> changes as σ_y does. Since NB(N) and NV(N) depend upon the shape of the PDF (rather than upon absolute location and scale), results will depend on the particular value of σ_y selected. An example of NB(N) and NV(N) for $\sigma_y = 1$ is presented in Figs. 6 and 7. Now we observe that the variance of the SMD is 3 times better, and that of the SGM 4.6 times better, than for the SAM, at least for larger values of N. However, as $\sigma_y \to 0$, the log-normal PDF in (13) approaches a Gaussian PDF about the point $u = \exp(m)$, and the behaviors would revert back to Fig. 1 then.

The next example is the Rice PDF; physically, this corresponds to the squared-envelope of the sum* of a sine wave and a centered narrowband Gaussian noise process. That is, the PDF is

$$p(u) = \frac{1}{2\sigma_i^2} \exp\left(-\frac{u + A^2}{2\sigma_i^2}\right) I_0\left(\frac{A\sqrt{u}}{\sigma_i^2}\right) \text{for } u > 0 , \tag{14}$$

where A is the sine wave amplitude and σ_i is the noise standard deviation. Once again, the shape of the PDF depends on a parameter, namely A/σ_i. Results for $A/\sigma_i = 1$ are given in Figs. 8 and 9. The SMD has 14% greater variance than the SAM, but the SGM has about 60% of the SAM variance. As $A/\sigma_i \to 0$, the exponential PDF results are obtained, whereas as $A/\sigma_i \to \infty$, the Gaussian case is realized. Thus (14) represents a transition case between these extremes.

The last example we consider here is an exponential PDF with out-liers. That is, each sample or measurement $\{x_k\}$ in (1)-(3) is given by

$$x_k = x_a + x_b, \tag{15}$$

where x_a has an exponential PDF,

$$p_a(u) = \frac{1}{m_a} \exp\left(-\frac{u}{m_a}\right) \quad \text{for } u > 0, \tag{16}$$

---

*If each of the observed random variables x_1, x_2, ..., x_N in (1)-(3) is obtained by first summing up the envelope-squared outputs of M narrowband filters, as for example in diversity reception, the PDF in (14) is replaced by the Q_M distribution. Results for this case are available in Ref. 1, but are not given here, for sake of brevity.

and disturbance x_b is a random variable which is zero most of the time, but occasionally takes on a large value (out-lier) L. That is, its PDF is

$$p_b(u) = (1 - Q) \, \delta(u) + Q \, \delta(u - L), \qquad\qquad (17)$$

where Q is the probability of an out-lier. Then the parent PDF of observation (sample) x_k is the convolution of (16) and (17):

$$p(u) = \frac{1 - Q}{m_a} \exp\left(- \frac{u}{m_a}\right) U(u) + \frac{Q}{m_a} \exp\left(- \frac{u - L}{m_a}\right) U(u - L), \qquad (18)$$

where unit step

$$U(t) = \begin{cases} 1 \text{ for } t > 0 \\ 0 \text{ for } t < 0 \end{cases}. \qquad\qquad (19)$$

The important parameter now is L/m_a, which obviously affects the shape of PDF (18).

Now however, before we get into the detailed bias and variance results, another consideration is of paramount importance. Our sample statistics, (1)-(3), will no longer extract (estimate) the arithmetic mean, median, and geometric mean, respectively, of the (disturbance-free) exponential PDF (16), but perforce, the corresponding statistics of the measurement PDF (18). If, however, we are really interested in the parameters of (16), then we must inquire into the quantitative disturbance caused by the out-liers described in (17). Here we merely cite the results for one numerical case; additional results are given in Ref. 1.

For probability Q = .05, and out-lier value L/m_a = 6, we find that the ratio of means, for (18) with respect to (16), is 1.3. The corresponding ratio of medians is only 1.08, whereas the ratio of geometric means is 1.13. Thus the SMD and SGM are more resistant to the presence of infrequent out-liers, insofar as their effects on the particular parameters of median and geometric mean.

The results for the normalized bias and variance are given in Figs. 10 and 11. The variances of the SMD and SGM are again smaller than that for the SAM. The bias of the SMD and SGM are comparable.

If we define a mean-square error as the average value of the squared difference between a sample statistic s and a desired parameter d_a of the disturbance-free PDF, we can develop it as follows:

$$MSE = \overline{(s - d_a)^2} = \overline{(s - \bar{s} + \bar{s} - d_a)^2}$$

$$= \overline{(s - \bar{s})^2} + (\bar{s} - d_a)^2 . \qquad\qquad (20)$$

But the first term is the variance of the sample statistic, and the second term can be expressed as

$$\bar{s} - d_a = (\bar{s} - d) + (d - d_a) = \text{bias} + \text{deflection in desired parameter}, \qquad (21)$$

where d is the modified value of the desired parameter d_a, due to the disturbance. Thus

$$\text{MSE} = \text{variance} + (\text{bias} + \text{deflection})^2 . \qquad (22)$$

Now the bias and variance are $O(N^{-1})$ for large N, whereas the deflection of the desired parameter does <u>not decay</u> with N at all; in fact, it is independent of N. Thus the considerations above, whether for the ratio of means or medians or geometric means, are very important, since they dominate the magnitude of the mean-square error for very many samples available.

DISCUSSION

The ability of the SMD and SGM to suppress deleterious effects due to occasional large interferences is very pronounced for some probability density functions. Not only is the deflection of the desired parameter (mean or median or geometric mean) decreased, but the bias and variance of the estimate can be markedly reduced in some cases. The exact amounts depend on the magnitude and frequency of the interference.

Another possible approach to alleviate the effects of additive large out-liers is to subject the available samples $x_1, x_2, \ldots, x_N$ to a nonlinear transformation such as saturation, in order to suppress the large contributions, prior to evaluating the SAM or SGM or SMD. Knowledge of relative levels (such as L/m_a for the above example) would be required for optimal adjustment of the saturation level, but performance could be markedly improved. The nonlinear transformation would reduce the deflection, while the averaging of N samples would reduce the bias and variance. This possibility has not yet been pursued.

REFERENCES

1. A. H. Nuttall, "Statistics of Sample Median, Quantile, Geometric Mean, and Arithmetic Mean, for Various Distributions," Technical Report 6689, Naval Underwater Systems Center, New London, CT USA.

2. H. Cramer, <u>Mathematical Methods of Statistics</u>, Princeton University Press, Princeton, N.J., 1961.
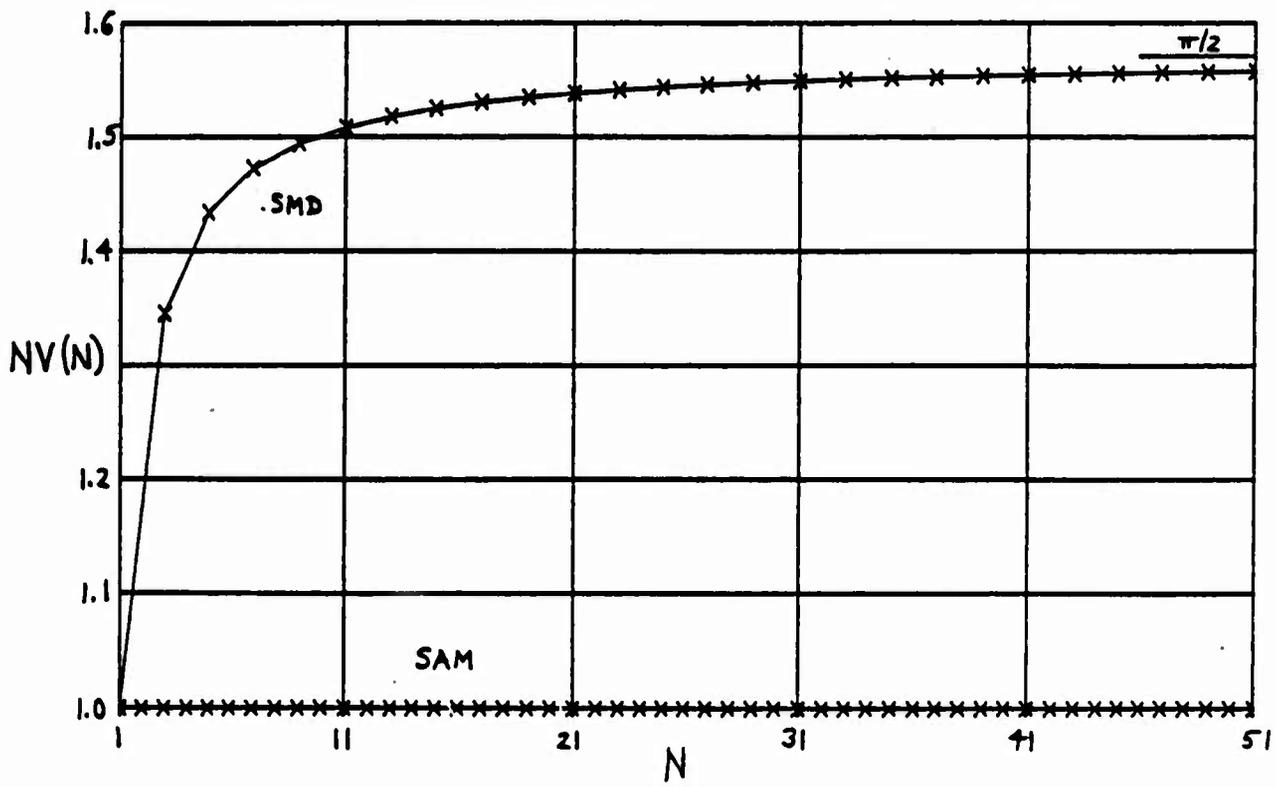
Figure 1. Normalized Variance; Gaussian PDF
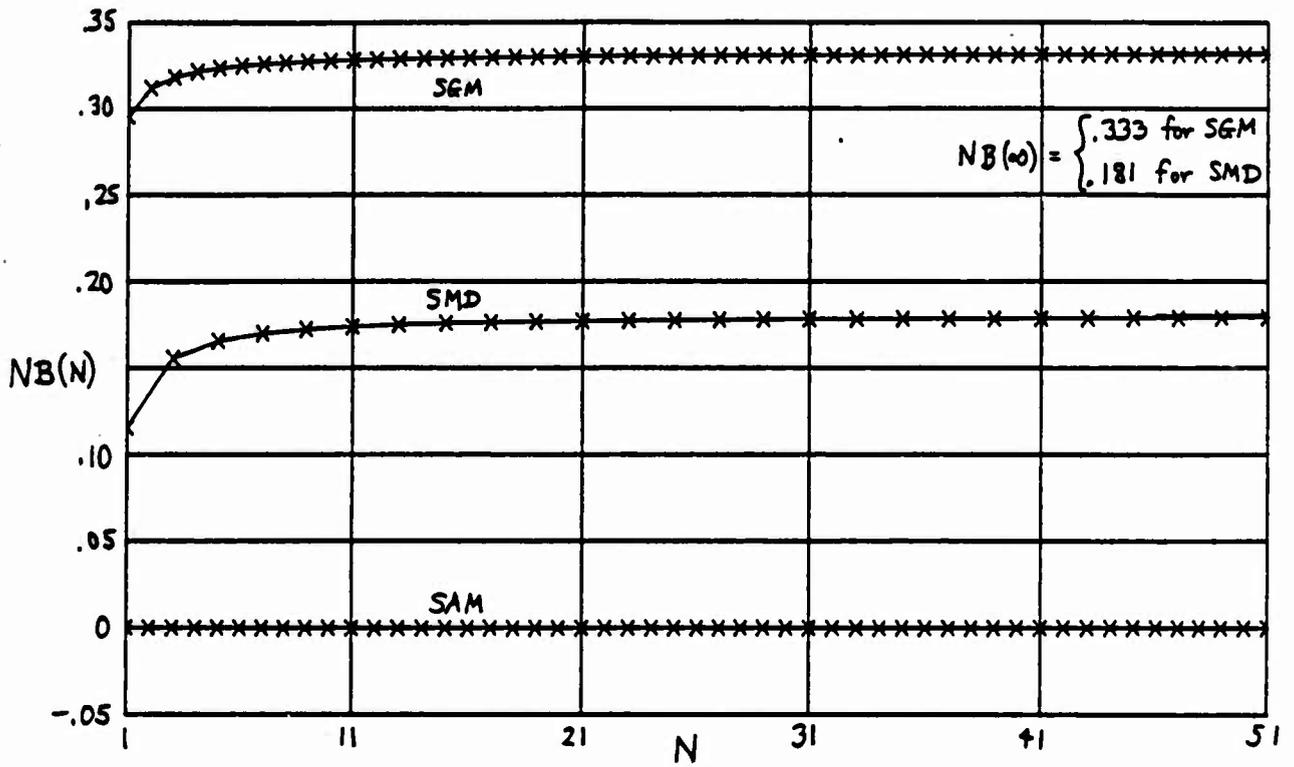
Figure 2. Normalized Bias; Rayleigh PDF

$$NB(\infty) = \begin{cases} .333 \text{ for SGM} \\ .181 \text{ for SMD} \end{cases}$$
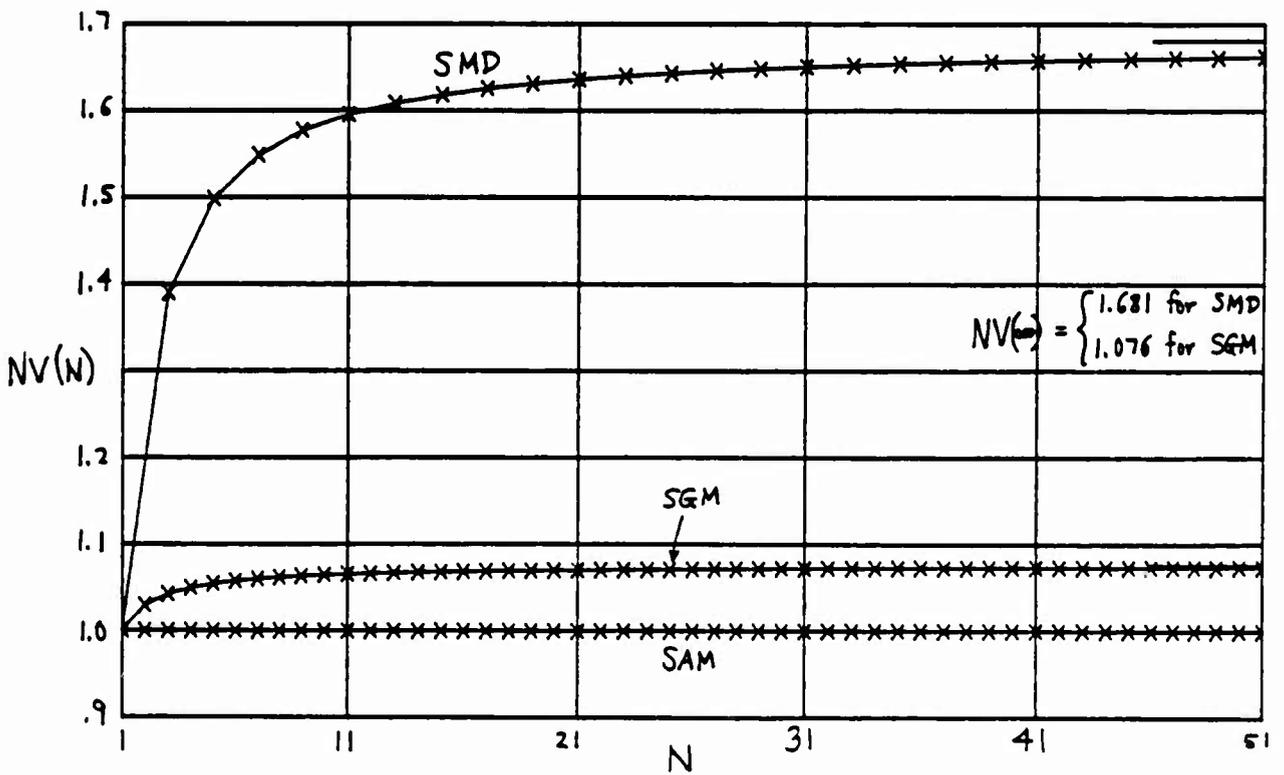
Figure 3. Normalized Variance; Rayleigh PDF

$$NV(\infty) = \begin{cases} 1.681 \text{ for SMD} \\ 1.076 \text{ for SGM} \end{cases}$$
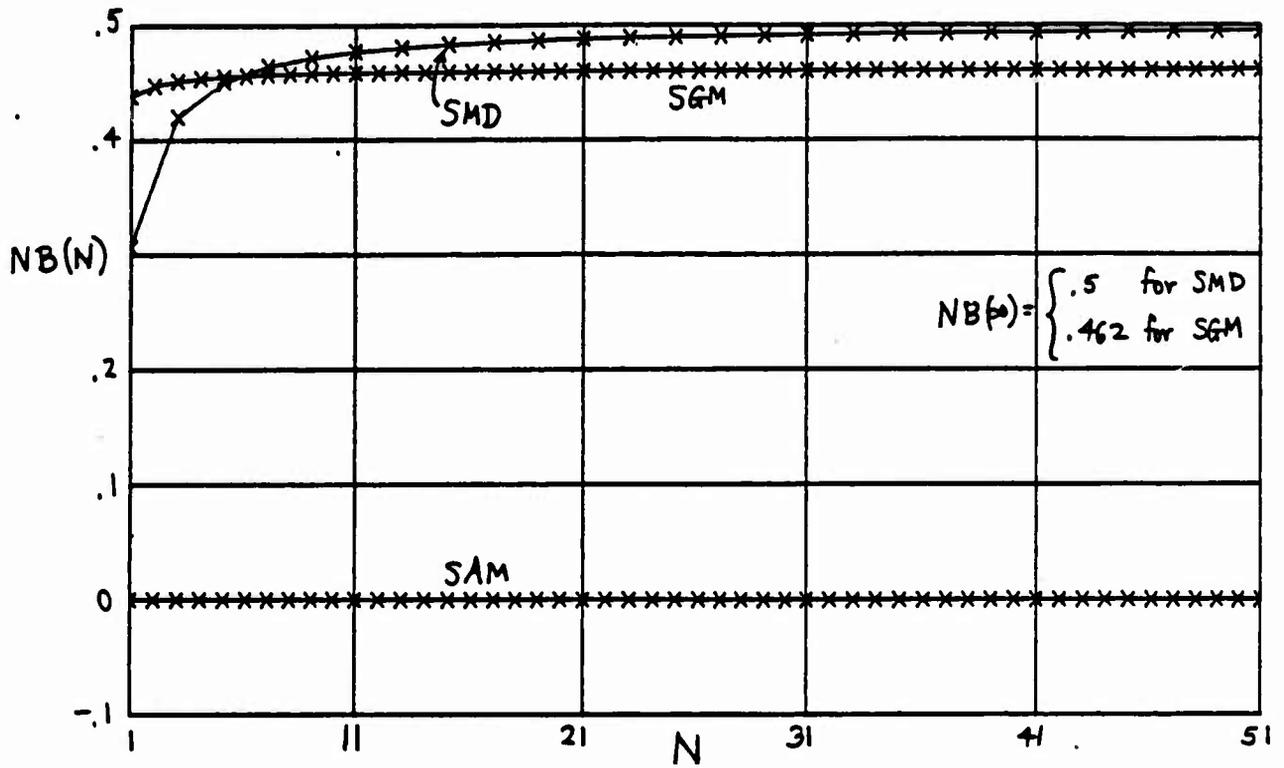
Figure 4. Normalized Bias; Exponential PDF
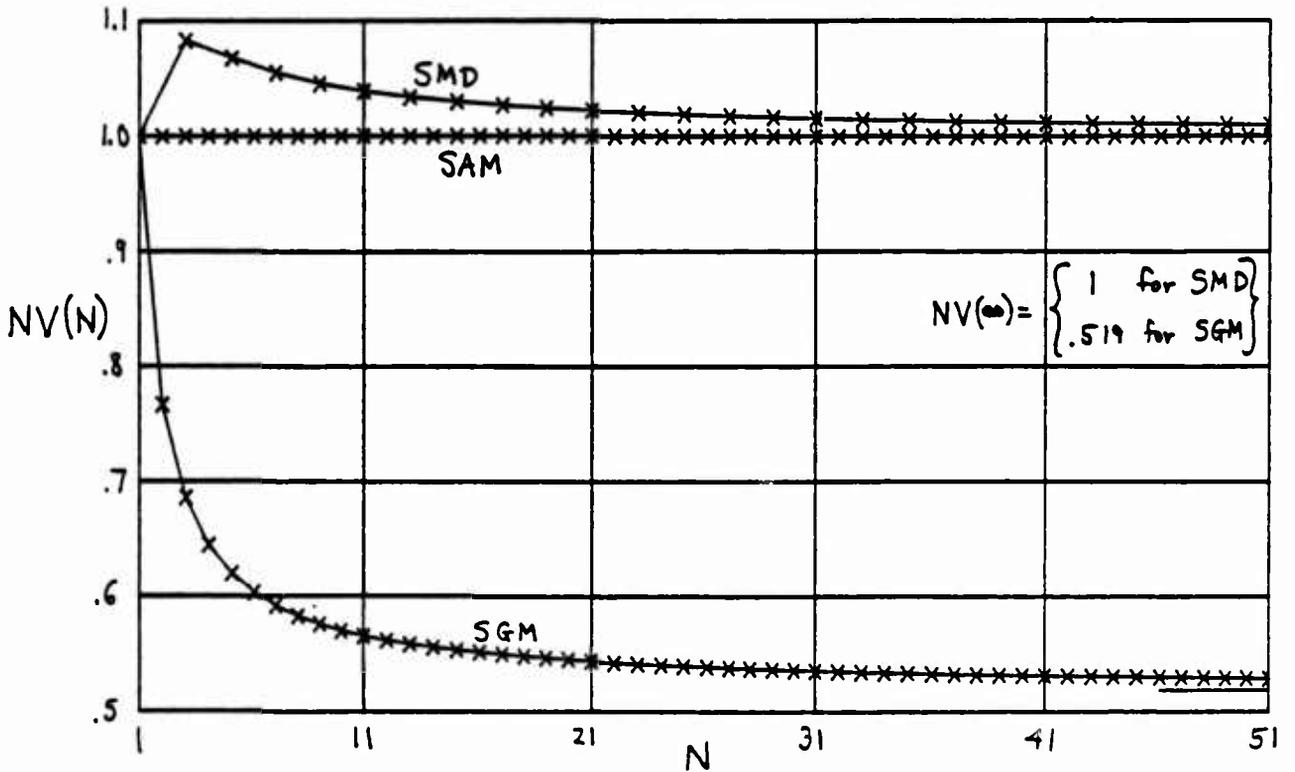
Figure 5. Normalized Variance; Exponential PDF

10
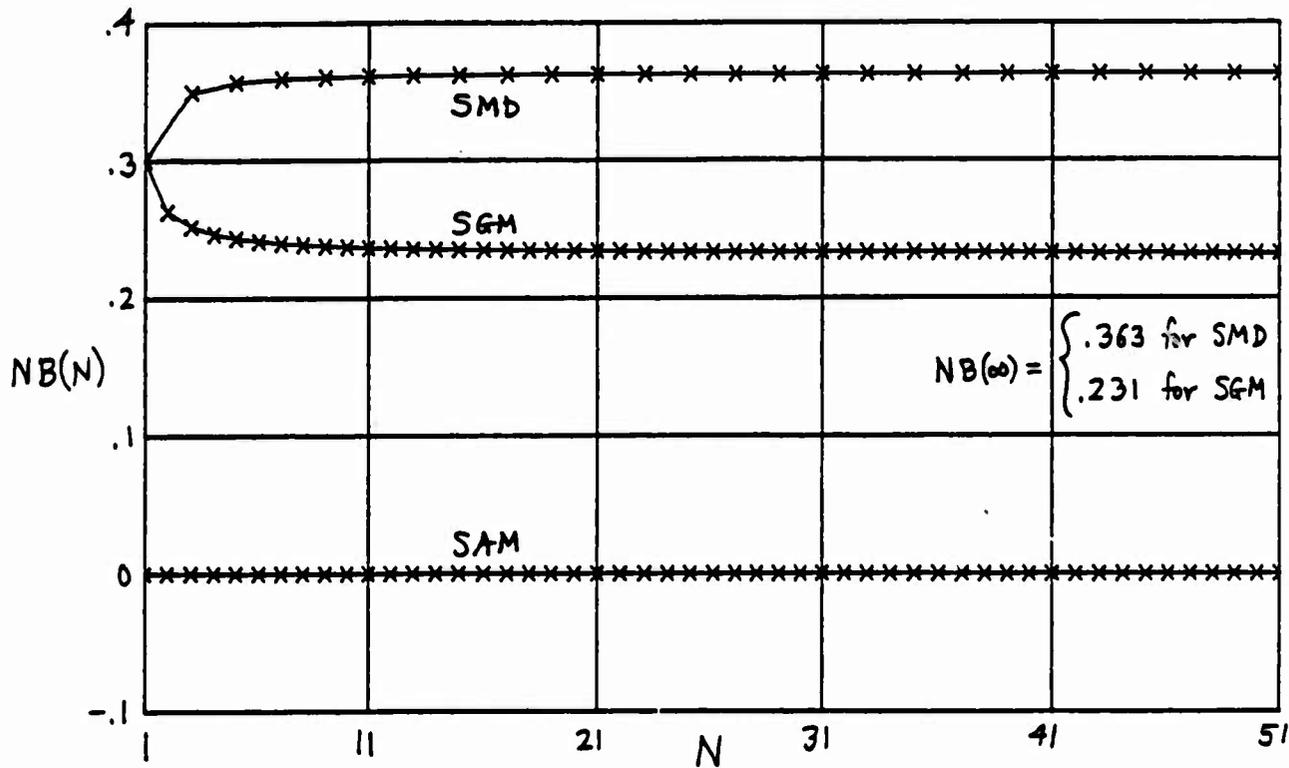
Figure 6. Normalized Bias; Log-Normal PDF ($\sigma_y = 1$)

$$NB(\infty) = \begin{cases} .363 \text{ for SMD} \\ .231 \text{ for SGM} \end{cases}$$
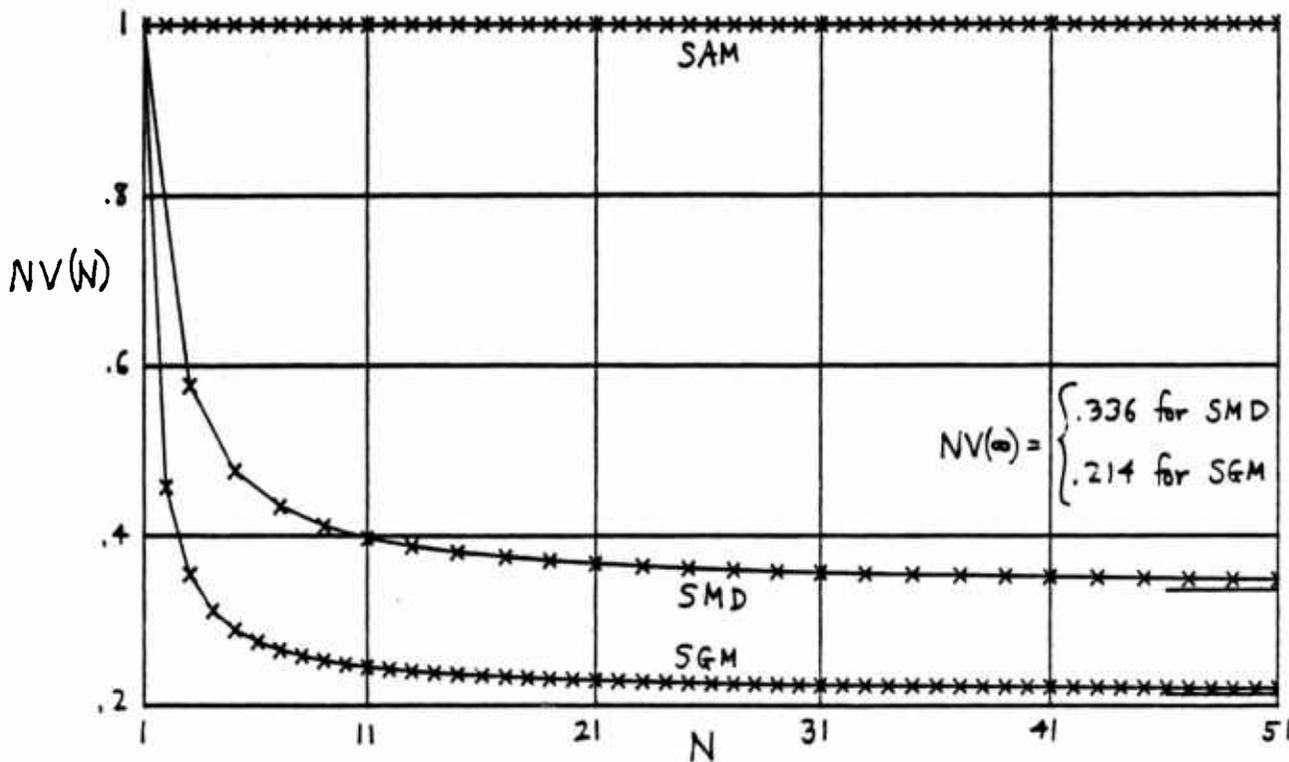
Figure 7. Normalized Variance; Log-Normal PDF ($\sigma_y = 1$)

$$NV(\infty) = \begin{cases} .336 \text{ for SMD} \\ .214 \text{ for SGM} \end{cases}$$
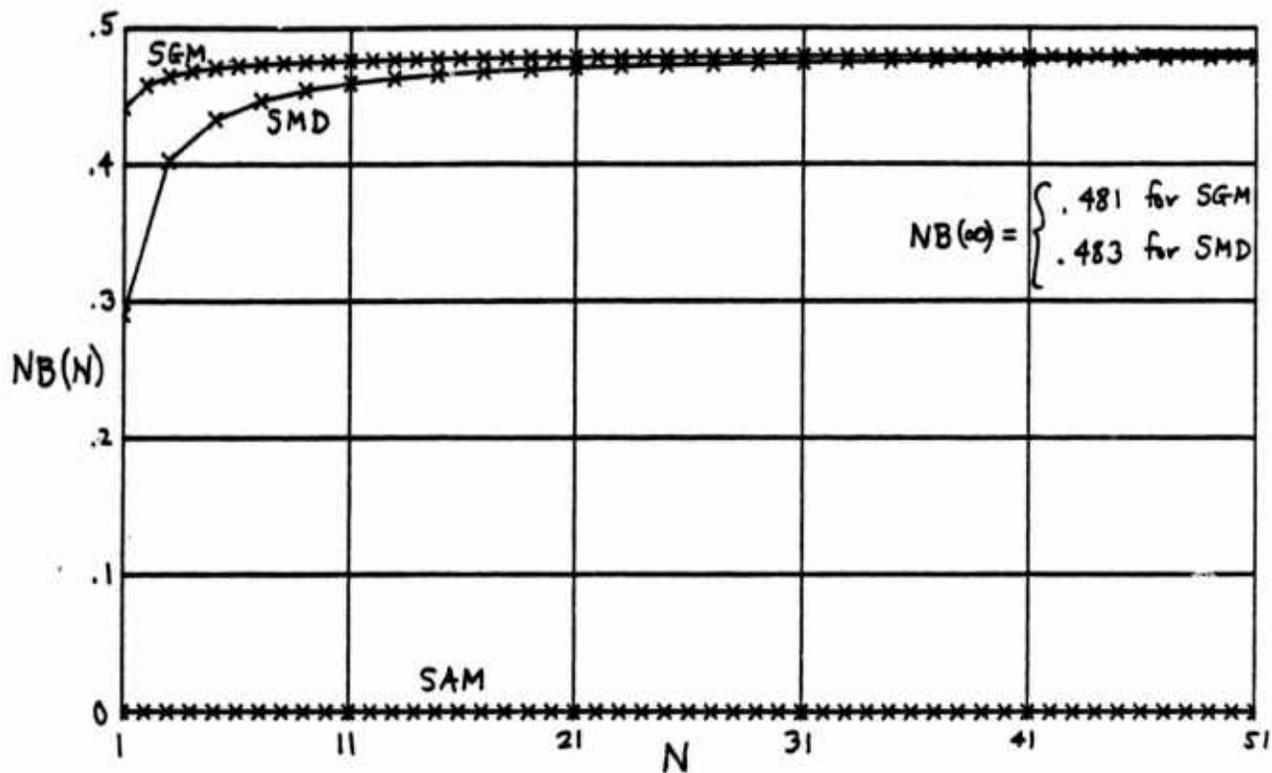
Figure 8. Normalized Bias; Rice PDF $(A/\sigma_1 = 1)$

$$NB(\infty) = \begin{cases} .481 \text{ for SGM} \\ .483 \text{ for SMD} \end{cases}$$
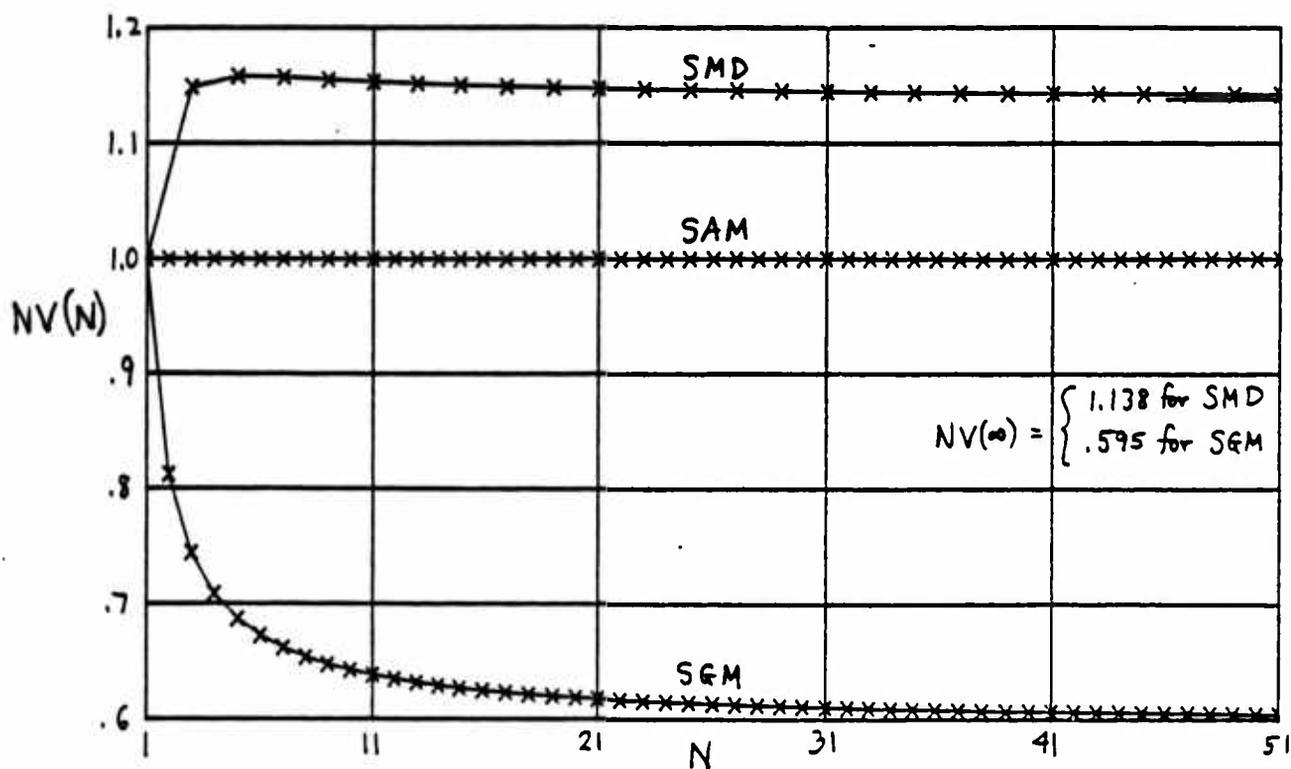
Figure 9. Normalized Variance; Rice PDF $(A/\sigma_1 = 1)$

$$NV(\infty) = \begin{cases} 1.138 \text{ for SMD} \\ .595 \text{ for SGM} \end{cases}$$
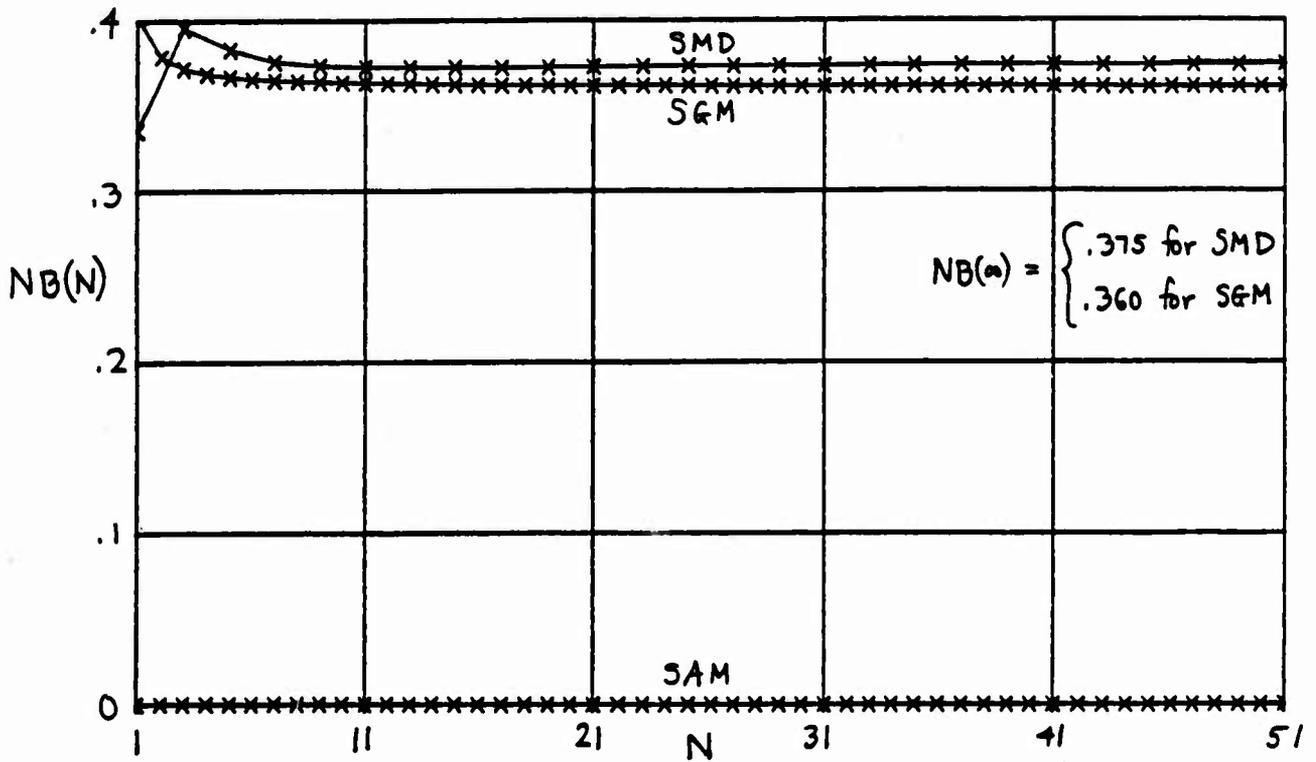
Figure 10. Normalized Bias; Exponential PDF With Out-liers $\left(\dfrac{L}{m_a} = 6,\ Q = .05\right)$
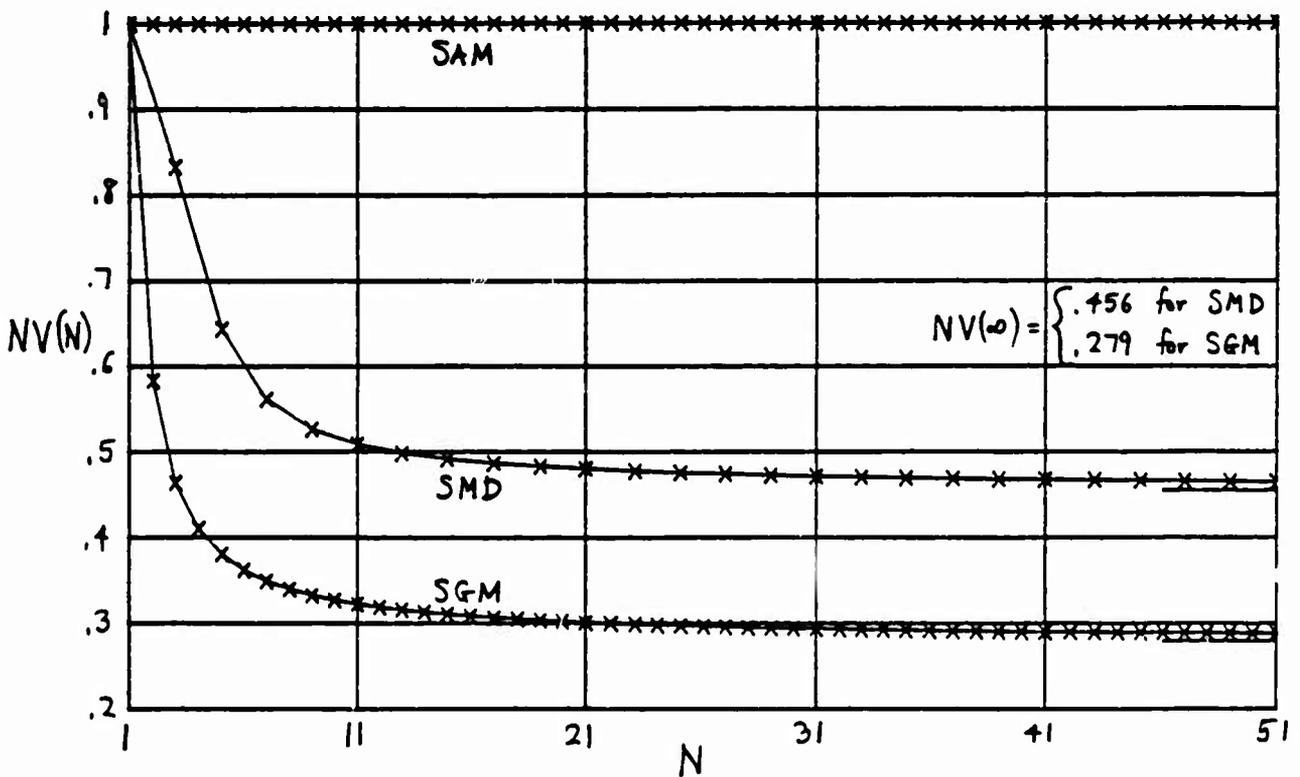
Figure 11. Normalized Variance; Exponential PDF With Out-liers $\left(\dfrac{L}{m_a} = 6,\ Q = .05\right)$

Detection Performance Characteristics For a System With Quantizers, Or-ing, And Accumulator

A. H. Nuttall

ABSTRACT

The false alarm and detection probabilities of an N-channel system subject to $L+1$ level quantization, or-ing, and accumulation of M time samples, are evaluated exactly, with no Gaussian assumptions, for arbitrary values of N, L, M, and for a quantizer with arbitrary breakpoint locations. The channel noises are independent but can have arbitrary statistics. The signal occupies one unknown (possibly changing) channel, if present. Two FFTs suffice to sweep out a complete detection probability vs. threshold curve. The optimum placement of quantizer breakpoints, for a fixed total number of levels, $L+1$, is a subtle one and is shown to depend on N, M, and the desired level of performance; a simple rule-of-thumb is presented which yields near-optimum capability over the useful range of the operating characteristics. A brief comparison with the optimum processor is made.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF SYMBOLS

N	number of input channels
m	sample time
$x_n(m)$	m-th time sample of n-th channel input waveform
M	number of time samples accumulated
L	number of quantizer abscissa breakpoints; L+1 is the number of quantizer output levels
$q\{\ \}$	quantizer nonlinear characteristic
b_ℓ	ℓ-th quantizer breakpoint location
$y(m)$	input to accumulator
z	output of accumulator; decision variable
$w(m)$	output of or-ing device in figure 5
$p_x^{(0)}$	cumulative distribution function of input x for signal absent
$p_x^{(1)}$	cumulative distribution function of input x for signal present
H_0, H_1	hypotheses 0 and 1: signal absent and present, respectively
P_w	cumulative distribution function of random variable w
α_ℓ	area of impulse at ℓ in the probability density function of random variable y
β_n	area of impulse at n in the probability density function of random variable z
f_z, f_y	characteristic functions of z and y, respectively
N_f	size of discrete Fourier transform; $N_f \geq ML+1$
J	integer-valued threshold
w_0, w_1	characteristic points of probability density function of w; figure 8
m_0, m_1	means of input x under H_0, H_1, respectively
σ	standard deviation of x
Φ	Gaussian cumulative distribution function; (17)
d_i	input deflection statistic; (18)
γ_ℓ	normalized quantizer breakpoint; (20)
v	normalized variable; (23)
P_D, P_F	detection, false alarm probabilities, respectively

DETECTION PERFORMANCE CHARACTERISTICS FOR A
SYSTEM WITH QUANTIZERS, OR-ING, AND ACCUMULATOR

INTRODUCTION

Practical realizations of desired systems often incorporate approximations to the ideal processor or devices, for the sake of reduced expense and equipment complexity. In particular, quantization is frequently employed, since it facilitates data handling and processing in terms of storage and execution time. In addition, tne large number of alternatives for signal presence and location that must be considered often dictates that data reduction procedures, sucn as or-ing, be adopted. Both of these suboptimum approaches, quantization and or-ing, degrade system performance, and it is important to know tne extent of the degradation. Alternatively, it is desirable to know how much the received signal strength must be increased in order to maintain a specified level of performance.

The effects of or-ing by itself were analyzed in ref. 1, where the required input signal-to-noise ratio for specified false alarm and detection probabilities was evaluated as a function of tne number of input channels and tne observation time. In ref. 2, the additional degradation caused by the inclusion of quantization was derived, where tne number of levels and breakpoint locations of tne quantizer were completely arbitrary. Both of these analyses, however, were limited to second-order moments of the decision variable, and were tnerefore most appropriate to the situation of large observation time and moderate false alarm probabilities. That is, the Gaussian presumption played a prevalent role in tne analysis. A cursory study of good quantizer breakpoint locations was also given in ref. 2.

Numerical evaluation of tne or-ing losses, based upon tne derivations in refs. 1 and 2 , were given in refs. 3 and 4, and a more extensive investigation of quantizer characteristics was conducted in ref. 5, which corroborated the results given in ref. 2. But again, all results were based on a second-order moment approacn.

Here we will derive exact results for the false alarm and detection probabilities of a system with quantization, or-ing, and accumulation-in-time, for arbitrary input signal and noise statistics. Surprisingly, it turns out that inclusion of the additional nonlinearity (quantization) actually simplifies tne analysis, and the judicious use of FFTs (fast Fourier transforms) makes tne numerical evaluation of the probabilities an efficient and accurate procedure.

SYSTEM DESCRIPTION AND ASSUMPTIONS

The system of interest is depicted in figure 1. There are N input channels, each of which is subject to noise statistically independent of the other channel noises. A signal is either present on one (and only one) unknown input channel or it is not present at all; we wish to make a decision on presence versus absence with good quality. The input signal (if present) need not be additive to the input noise; all that is required is knowledge of the probability density function of the input channel random variable $x_n(m)$ (at sample time m) for signal present. By setting the signal strength to zero, we obtain the probability density function for noise-only, of course.

It is presumed that the signal, if present, remains so for all M discrete time samples accumulated at the or-ing output. However, the signal could remain in one channel, or it could wander over any number of channels in a deterministic or random manner, from time sample to time sample.

The probability density functions of inputs $\{x_n(m)\}$ are arbitrary except that they must be identical for all the noise-only channels. Extension to non-identically distributed channel noises appears possible but has not been pursued. No Gaussian assumptions are made at any point of the system of figure 1, for the general analysis to follow. However, the numerical example that we eventually pursue is for Gaussian inputs, although this could easily be replaced by a different case of interest; in fact, some candidate examples are listed later in an appendix.

Each channel input $x_n(m)$ is subject to memoryless quantization, where the quantizer output $q\{x_n(m)\}$ has a total of L+1 levels, and the L quantizer abscissa breakpoints are arbitrary; see figure 2. A general analysis for this most-general quantizer is possible and will be outlined later. We confine our attention here to the less-general quantizer depicted in figure 3, where the ordinate (output) levels are limited to the equi-spaced values 0, 1, 2, ..., L, but the abscissa (input) breakpoints $\{b_\ell\}$ are arbitrary, except that $b_\ell \leq b_{\ell+1}$ for $1 \leq \ell \leq L-1$, without loss of generality.
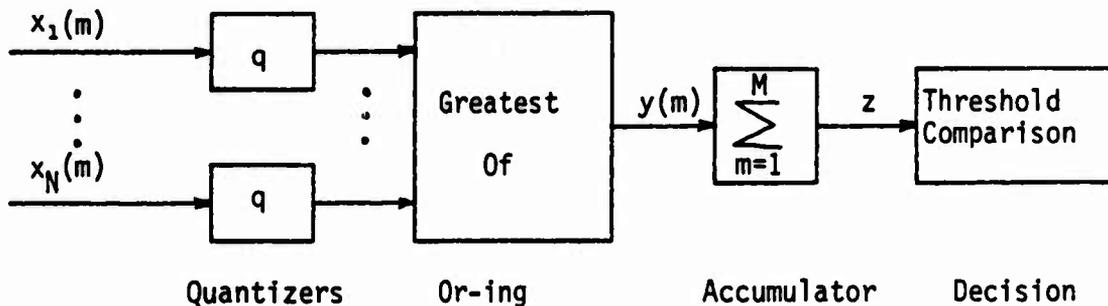
Figure 1. Nonlinear System
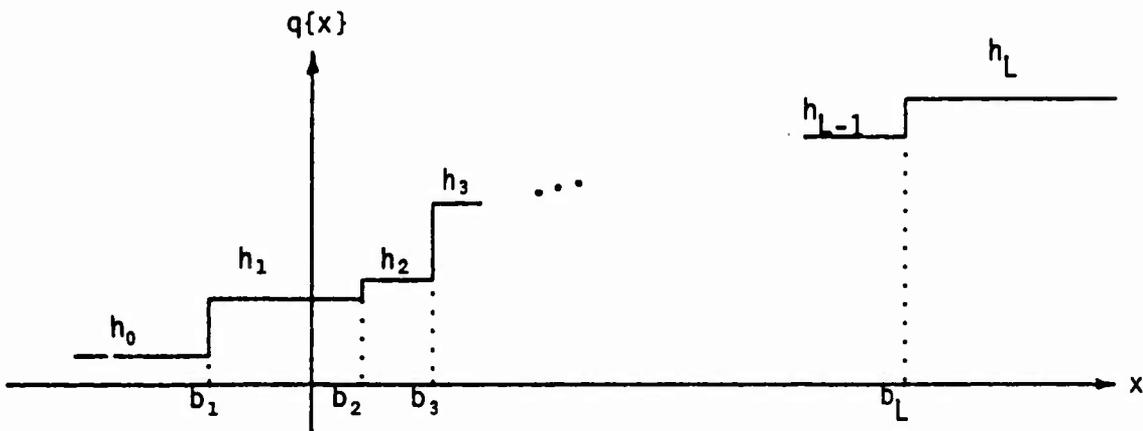
Figure 2. General Quantizer Characteristic
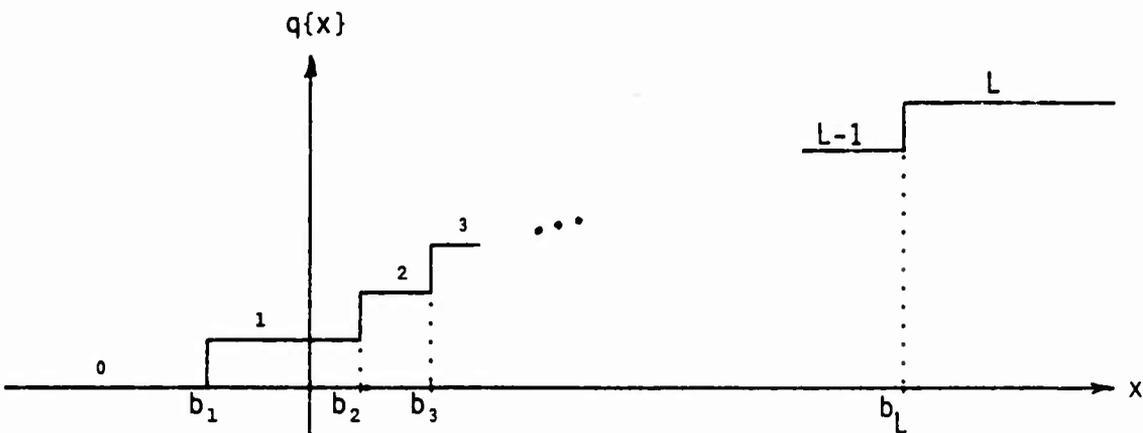
Figure 3. Specific Quantizer Characteristic

This limitation on the possible quantizer output levels is not as restrictive as first appears. Reference to figure 1 quickly reveals that if the quantizer output levels were at equi-spaced values $h_0 + \ell \Delta h$, for $0 \leq \ell \leq L$, the operating characteristics, namely detection probability versus false alarm probability, would be completely <u>independent</u> of h_0 and Δn (> 0). That is, h_0 and Δn affect the absolute scale of $y(m)$ and z, but they can be absorbed in a modified threshold at the system output. In fact, if one wants to approximate a given nonlinearity (such as $\tanh(x)$, for example) by a quantizer, L can be selected large enough, h_0 and Δh chosen without restriction, and the $\{b_\ell\}_0^L$ selected for a good fit. Then the analysis, as contained here for the quantizer of figure 3, applies directly, where h_0 and Δh (determined from tne nonlinear fit) are discarded. h_0 and Δh are temporarily used for the fitting procedure, but are not fundamental to the system operating cnaracteristics.

An example of a fit to $\tanh(x)$ is given in figure 4 for L = 7, where the
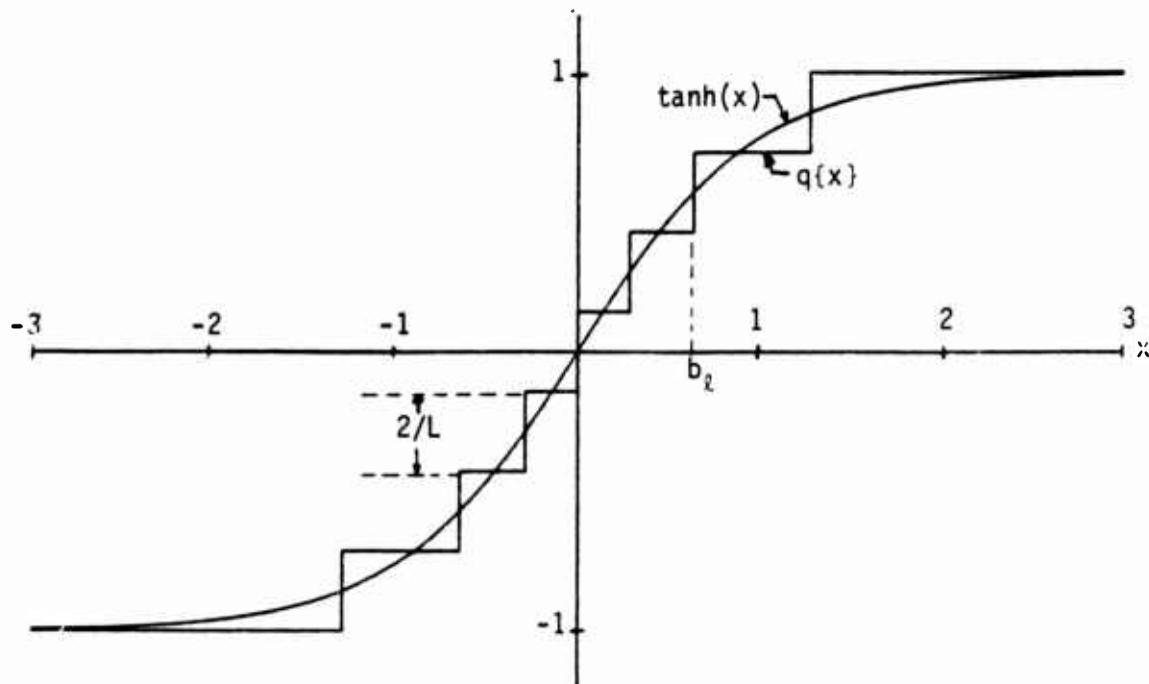
Figure 4. A Quantizer Fit to tanh(x) for L = 7

5

maximum ordinate error has been minimized and equalized at all the steps, and both functions approach ± 1 as x tends to $\pm \infty$. The best locations of the breakpoints are at

$$h_\ell = \frac{1}{2} \ln \left(\frac{\ell - \frac{1}{2}}{L + \frac{1}{2} - \ell} \right) \quad \text{for } 1 \leq \ell \leq L \quad , \tag{1}$$

for this example. We also have $h_0 = -1$, $\Delta h = 2/L$.

The or-ing device in figure 1 is subject to N inputs at each discrete time sample m. It selects the largest of these N random variables at each m and emits it to the accumulator as variable $y(m)$.

The accumulator adds up M input time samples to yield the decision variable

$$z = \sum_{m=1}^{M} y(m) \quad . \tag{2}$$

This output z is compared with a threshold. If the threshold is exceeded, a signal is declared present at the input to the system; otherwise, no signal is declared. For present analysis purposes, it is presumed that the discrete time samples (which can actually take place on the input channels) are sufficiently separated in time that the M random variables entering the accumulator are statistically independent.

When quantizer q in figure 1 is monotonically nondecreasing, the non-linear system in figure 1 is equivalent (for all inputs) to that shown in figure 5, where the quantization and or-ing operations have been interchanged.* The random variable $y(m)$ in figure 5 is identical to that in figure 1. We have added another random variable, $w(m)$, in figure 5 that has no counterpart in figure 1, for purposes of analysis. This interchange of operations is valid whether the quantizer of figures 2 or 3 is used.

Extensions of the above assumptions to more general situations, such as statistically dependent inputs or dependent accumulator samples, are discussed in ref. 2. However, the analysis there is limited to second-order moments, not threshold-crossing probabilities as considered here.

---

*See refs. 2 and 4 for two different proofs.
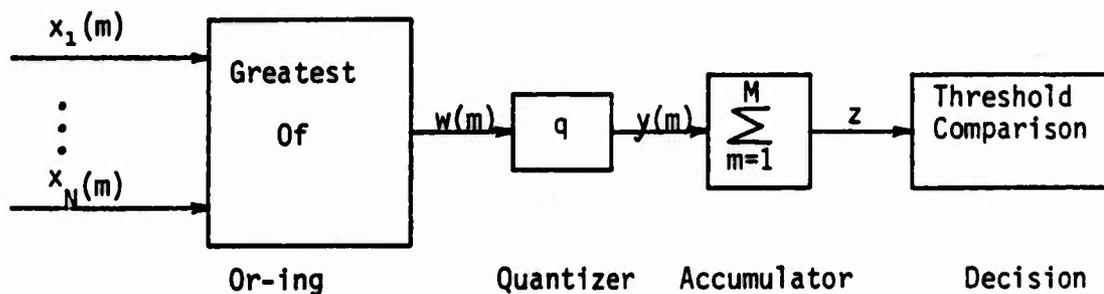
Figure 5. Equivalent Nonlinear System
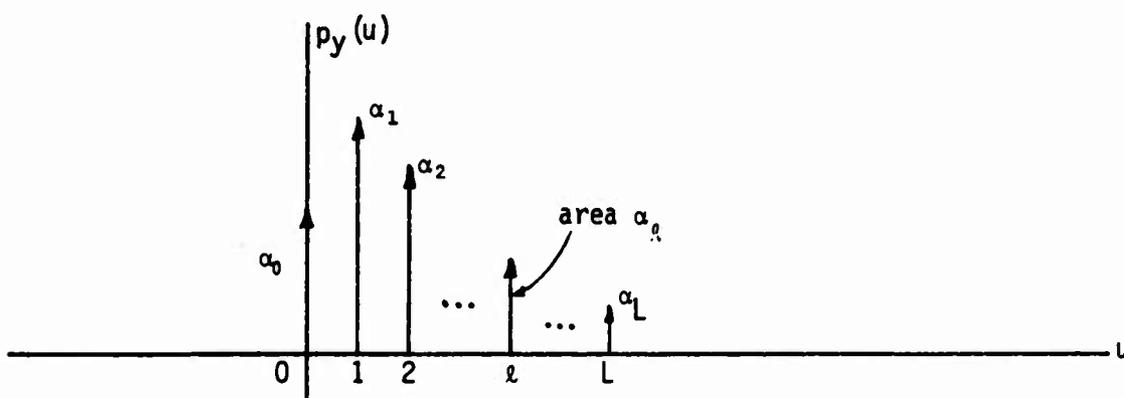
Figure 6. Probability Density Function of y(m)
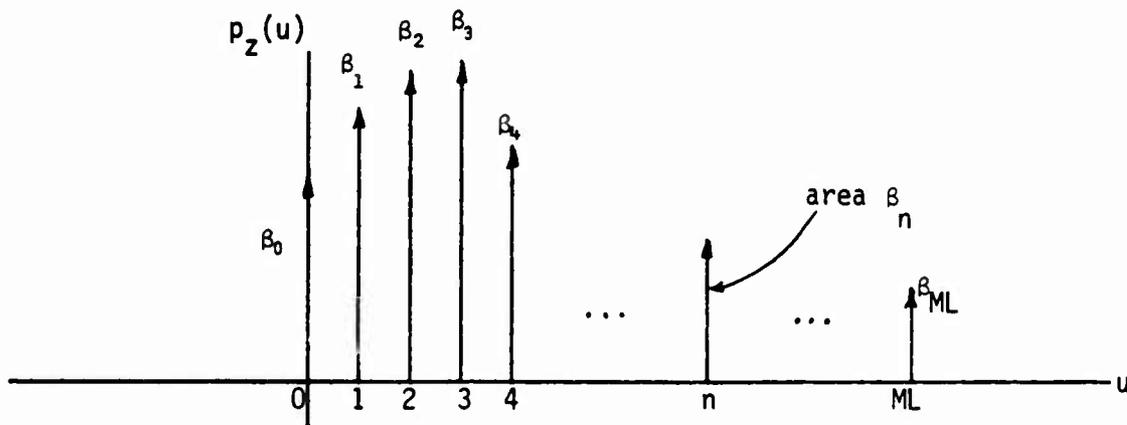
Figure 7. Probability Density Function of z

ANALYSIS OF PERFORMANCE

In this section, we evaluate the false alarm and detection probabilities of the system in figure 5. The cumulative distribution function of an individual channel input, for signal absent, is $P_x^{(0)}$, where

$$P_x^{(0)}(u) = \text{Probability } (x < u | \text{signal absent}) \quad . \quad (3)$$

The cumulative distribution function of the channel input with signal-present is $P_x^{(1)}$. Both of these cumulative distribution functions are presumed known. The superscript will denote either hypothesis H_0 (no signal present) or hypothesis H_1 (signal present on that channel). Notice there is no restriction on the forms of $P_x^{(0)}$ or $P_x^{(1)}$; thus arbitrary input statistics (including nonadditive signals) are allowed. However, the noise-only channels are identically distributed.

Since the or-ing output in figure 5 is

$$w(m) = \max\left\{x_1(m), \ldots, x_N(m)\right\} \quad \text{for } 1 \leq m \leq M \quad , \quad (4)$$

the cumulative distribution function of $w(m)$ is, for any time instant m,

$$P_w(u) = \begin{cases} \left[P_x^{(0)}(u)\right]^N & \text{for } H_0 \\ \left[P_x^{(0)}(u)\right]^{N-1} P_x^{(1)}(u) & \text{for } H_1 \end{cases} , \quad (5)$$

since the signal (if present) is in only one (arbitrary) input channel. In fact, the occupied channel can change from sample m to m+1, in a random or deterministic fashion, and (5) is still applicable.

The output $y(m)$ of the quantizer in figure 5 is limited to the L+1 values $\ell = 0, 1, 2, \ldots, L$; see figure 3. Therefore the probability density function of random variable $y(m)$ is impulsive, as shown in figure 6, where the area α_ℓ of the impulse at ℓ is given by

$$\alpha_\ell = \begin{cases} P_w(b_1) & \text{for } \ell = 0 \\ P_w(b_{\ell+1}) - P_w(b_\ell) & \text{for } 1 \le \ell \le L-1 \\ 1 - P_w(b_L) & \text{for } \ell = L \end{cases} \quad . \tag{6}$$

Here $\{b_\ell\}_1^L$ are the L abscissa breakpoints of the quantizer, and P_w is the cumulative distribution function of w, given by (5).

The output of the accumulator in figure 5 is called the decision variable and is given by the sum of M statistically independent random variables according to

$$z = \sum_{m=1}^{M} y(m) \quad . \tag{7}$$

Since $y(m)$ can only take on the values 0, 1, 2, ..., L, the decision variable z can only take on the values 0, 1, 2, ..., ML. Thus the probability density function of z is also impulsive, as depicted in figure 7; the area of the impulse at n is denoted by β_n.

We now need to relate the $\{\beta_n\}$ of figure 7 to the $\{\alpha_\ell\}$ of figure 6. To accomplish this, we resort to the characteristic functions of $y(m)$ and z. Using the statistical independence of the $\{y(m)\}$, the characteristic function of z is

$$f_z(\xi) = \overline{\exp(i\xi z)} = \left[\overline{\exp(i\xi y)}\right]^M = \left[f_y(\xi)\right]^M \quad , \tag{8}$$

where f_y is the characteristic function of random variable $y(m)$. But by use of figures 6 and 7, we have

$$f_y(\xi) = \int_{-\infty}^{\infty} du\, p_y(u)\, \exp(i\xi u) = \sum_{\ell=0}^{L} \alpha_\ell \exp(i\xi\ell) \quad , \tag{9}$$

and

$$f_z(\xi) = \int_{-\infty}^{\infty} du\, p_z(u)\, \exp(i\xi u) = \sum_{n=0}^{ML} \beta_n \exp(i\xi n) \quad . \tag{10}$$

Substitution of (9) and (10) in (8) yields

$$f_z(\xi) = \sum_{n=0}^{ML} \beta_n \exp(i\xi n) = \left[\sum_{\ell=0}^{L} \alpha_\ell \exp(i\xi\ell) \right]^M . \qquad (11)$$

Now the results of appendix A indicate that the set of numbers $\left\{\beta_n\right\}_0^{ML}$ can be recovered directly by an N_f-point DFT (discrete Fourier transform) of the set of characteristic function samples

$$f_z(2\pi n/N_f) \quad \text{for } 0 \leq n \leq N_f-1 \quad , \qquad (12)$$

provided that

$$N_f \geq ML+1 \quad . \qquad (13)$$

And from (11), we have the required characteristic function samples as

$$f_z(2\pi n/N_f) = \left[\sum_{\ell=0}^{L} \alpha_\ell \exp(i2\pi n\ell/N_f) \right]^M . \qquad (14)$$

The sum on ℓ here is the conjugate of an N_f-point DFT of the L+1 real nonzero numbers $\left\{\alpha_\ell\right\}_0^L$ augmented by zeros. When this DFT is raised to the M-th power, it constitutes the required N_f samples of the characteristic function f_z that are needed for the DFT that leads to $\left\{\beta_n\right\}$.

To summarize, (5) yields the cumulative distribution function of w, and (6) gives the impulse areas $\left\{\alpha_\ell\right\}_0^L$. An N_f-point DFT of this sequence (augmented with zeros) is taken, then conjugated and raised to the M-th power. Another N_f-point DFT is then taken and the results divided by N_f. The end result is impulse areas $\left\{\beta_n\right\}$. The values returned (by the second DFT) for $\beta_{ML+1}, \ldots, \beta_{N_f-1}$ should all be zero, as should all the imaginary parts.

Although any value for N_f that satisfies (13) is permissible, the smallest power of 2 is most reasonable since then we can employ two FFTs above; this time-saving feature is employed here.

The probability that decision variable z equals or exceeds integer value J is, from figure 7,

$$\text{Probability} \left\{ z \geq J \right\} = \sum_{n=J}^{ML} B_n \qquad \text{for } 0 \leq J \leq ML \quad .$$ (15)

The above analysis for the exceedance probability in (15) is exact. When signal is present, (15) is the detection probability, whereas when signal is absent, (15) is the false alarm probability. The fundamental input statistics, $P_x^{(0)}$ and $P_x^{(1)}$, required in (5), are arbitrary.

QUANTIZER BREAKPOINT LOCATIONS

For a given number, L+1, of quantizer output levels, the selection of the L breakpoints $\{b_\ell\}$ should be done so as to optimize the performance of the system, that is, maximum detection probability for a specified false alarm probability. The equivalence of the two processors in figures 1 and 5 is the guide to good selection of the breakpoints; namely $\{b_\ell\}$ should lead to a maximum difference in outputs $y(m)$ for hypothesis H_1 vs. H_0. However, as figure 5 shows, this selection of $\{b_\ell\}$ is governed by the probability density functions of quantizer input w, _not_ x; that is, the quantizer should take account of the or-ing nonlinearity and the number of input channels, N.

A problem arises here, however, in that the probability density functions of w under H_1 and H_0 depend on the particular value of input signal-to-noise ratio that obtains. Thus, the quantizer design should take the input signal-to-noise ratio into account. This situation is frequently encountered in likelihood ratio processing, in which the optimum processor often requires knowledge of absolute levels of input signal and noise. Since this knowledge is almost always lacking, a design that is good for representative values of input signal-to-noise ratio, that is, which correspond to adequate levels of performance, should be adopted. If the input signal-to-noise ratio is larger than these representative values, improved performance will result; if smaller, inadequate performance is expected anyway, regardless of quantizer breakpoint placements.

To get at the quantizer design, consider the probability density functions in figure 8 for random variable w under H_1 and H_0. w_0 is a point beyond which there is a small chance of w ever reaching under H_0; w_1 is a point below which w hardly ever reaches under H_1. Generally $w_1 < w_0$ for cases of practical importance; otherwise near-perfect performance is possible at this signal-to-noise ratio. Under H_0, we would like to locate the first breakpoint b_1 greater than w_0; then the false alarm probability would be substantially zero, regardless of the remaining $\{b_\ell\}$. On the other hand, under H_1, we would like to locate the last breakpoint b_L less than w_1; then the detection probability would be essentially 1.
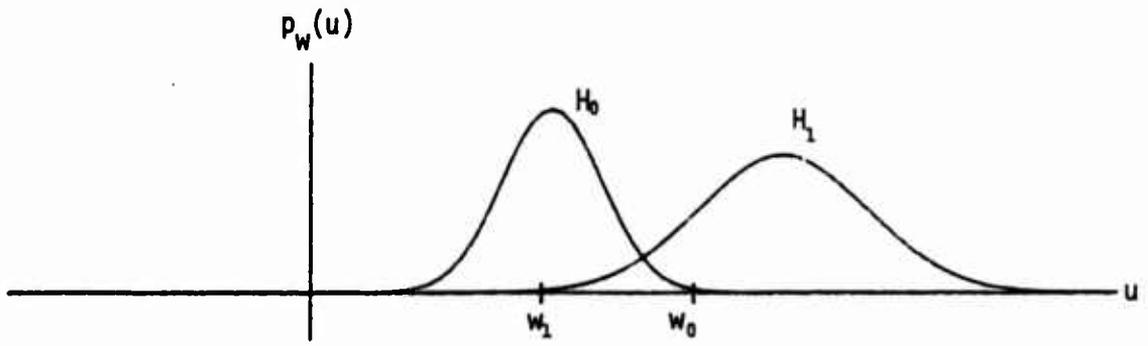
Figure 8. Probability Density Functions of w in figure 5
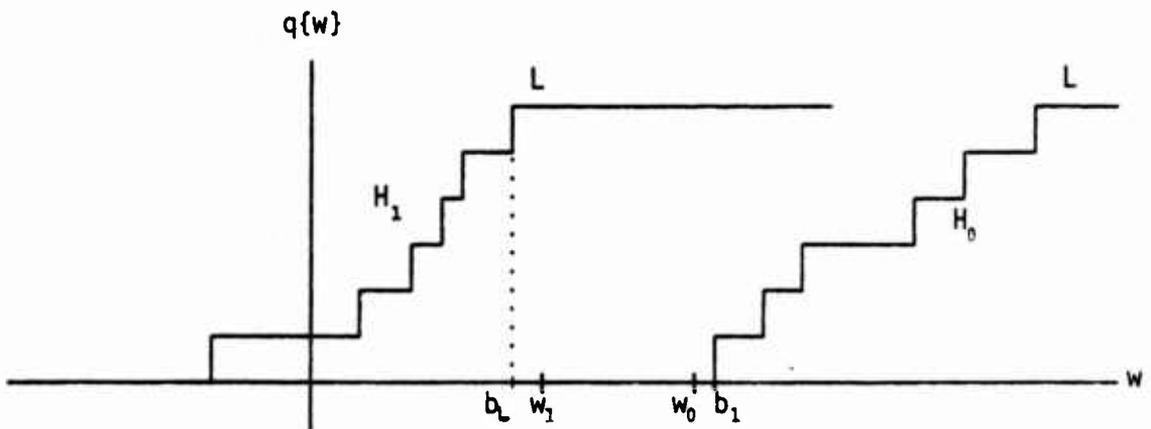
Figure 9. Desirable Quantizer Characteristics
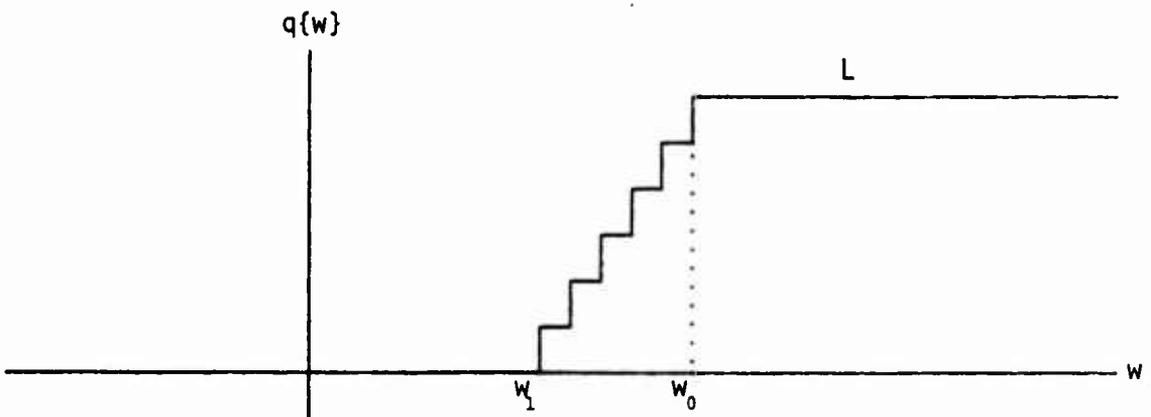
Figure 10. Compromise Quantizer Characteristic

These desirable quantizer characteristics under H_0 and H_1 are depicted in figure 9. It is immediately seen that the desired features of both cases can be realized for $w < w_1$ and $w > w_0$; that is, choose $q=0$ for $w < w_1$, and choose $q=L$ for $w > w_0$. However there is an inherent conflict in the intermediate region $w_1 < w < w_0$. The only way to strike a reasonable compromise is to make q small near w_1 and make q large near w_0. That is, locate the breakpoints in w_1, w_0, as indicated in figure 10. We shall make them equally spaced on the input w, that is, $b_{\ell+1} - b_\ell$ independent of ℓ; nonuniform abscissa spacings, such as in figure 4, are possible and could give slightly better performance. However, sample computer runs have demonstrated that for $L > 4$, essentially optimum performance is attained via uniform breakpoint spacing. Also, for the larger values of input signal-to-noise ratio, it will be shown that the optimum processor of the available inputs takes precisely the form of figure 5, where the quantizer is replaced by a linear gain.

The best quantizer placements in figure 10 obviously depend on the number, N, of input channels and the input signal-to-noise ratio, since the probability density functions of w displayed in figure 8 depend on these quantities. But there is an additional less-obvious dependence on M, the number of samples accumulated at the quantizer output. For larger M, the separation of the probability density functions of z under H_0 and H_1 will become better, if the input signal-to-noise ratio is held fixed. But often, the larger values of M are employed so that lower input signal-to-noise ratios can be tolerated and yet realize adequate performance levels; thus, the probability density functions of w in figure 8 generally overlap more for the larger values of M. This means that w_1 will be smaller and therefore the breakpoints should be relocated. Further discussion of quantizer breakpoints is deferred until the numerical investigation is undertaken.

INPUT STATISTICS

Up to now, the statistics of inputs $\{x_n\}$ in figures 1 and 5 have been arbitrary. We now specialize to the case of Gaussian noises; some other candidate statistics are given in appendix B. In particular, the cumulative distribution functions under H_0 and H_1 are

$$P_x^{(0)}(u) = \Phi\left(\frac{u-m_0}{\sigma}\right), \quad P_x^{(1)}(u) = \Phi\left(\frac{u-m_1}{\sigma}\right) \quad , \tag{16}$$

respectively, where Φ is the Gaussian cumulative distribution function

$$\Phi(x) = \int_{-\infty}^{x} dt \, (2\pi)^{-1/2} \exp(-t^2/2) \equiv \int_{-\infty}^{x} dt \, \phi(t) \quad . \tag{17}$$

The means of x_n are m_0 and m_1 under H_0 and H_1, respectively, while the standard deviation is the common value σ in both cases. For later use, we define the deflection statistic of the inputs as

$$d_i = \frac{m_1 - m_0}{\sigma} \quad . \tag{18}$$

Reference to (5) and (6) indicates that we need the quantities

$$\left.\begin{array}{l} P_x^{(0)}\left(b_\ell\right) = \Phi\left(\frac{b_\ell - m_0}{\sigma}\right) = \Phi\left(v_\ell\right) \\[2em] P_x^{(1)}\left(b_\ell\right) = \Phi\left(\frac{b_\ell - m_1}{\sigma}\right) = \Phi\left(v_\ell - d_i\right) \end{array}\right\} \text{for } 1 \le \ell \le L, \tag{19}$$

where we have used (16), (17), and defined the normalized breakpoints of the quantizer as

$$v_\ell = \frac{b_\ell - m_0}{\sigma} \quad \text{for } 1 \le \ell \le L \quad . \tag{20}$$

Then (5) and (19) yield

$$P_w(b_\ell) = \begin{cases} \Phi^N(v_\ell) & \text{for } H_0 \\ \Phi^{N-1}(v_\ell)\Phi(v_\ell - d_i) & \text{for } H_1 \end{cases} \quad \text{for } 1 \leq \ell \leq L \quad . \quad (21)$$

These are the quantities needed in (6) for the areas $\{a_\ell\}$ in figure 6. (If the noise standard deviations were σ_0 and σ_1 under H_0 and H_1, instead of the common σ in (16), the only changes would be to replace $v_\ell - d_i$ in the second line of (19) and (21) by $\frac{\sigma_0}{\sigma_1}\left(v_\ell - d_i\right)$). The results in the next section are based upon this Gaussian noise example.

RESULTS

By combining the general result for the cumulative distribution function of w in (5) with the Gaussian example in (16), we obtain

$$P_w(u) = \Phi^{N-1}\left(\frac{u-m_0}{\sigma}\right)\Phi\left(\frac{u-m_1}{\sigma}\right)$$

$$= \Phi^{N-1}(v)\,\Phi(v-d_i) \quad , \tag{22}$$

where we define normalized variable

$$v = \frac{u-m_0}{\sigma} \tag{23}$$

and have used (18). The probability density function of w under H_1 is

$$p_w(u) = \frac{d}{du}P_w(u) = \frac{1}{\sigma}\frac{d}{dv}\left\{\Phi^{N-1}(v)\,\Phi(v-d_i)\right\}$$

$$= \frac{1}{\sigma}\left[(N-1)\,\Phi(v-d_i)\,\phi(v) + \Phi(v)\phi(v-d_i)\right]\Phi^{N-2}(v) \quad , \tag{24}$$

where we used (17). This probability density function is plotted in figure 11
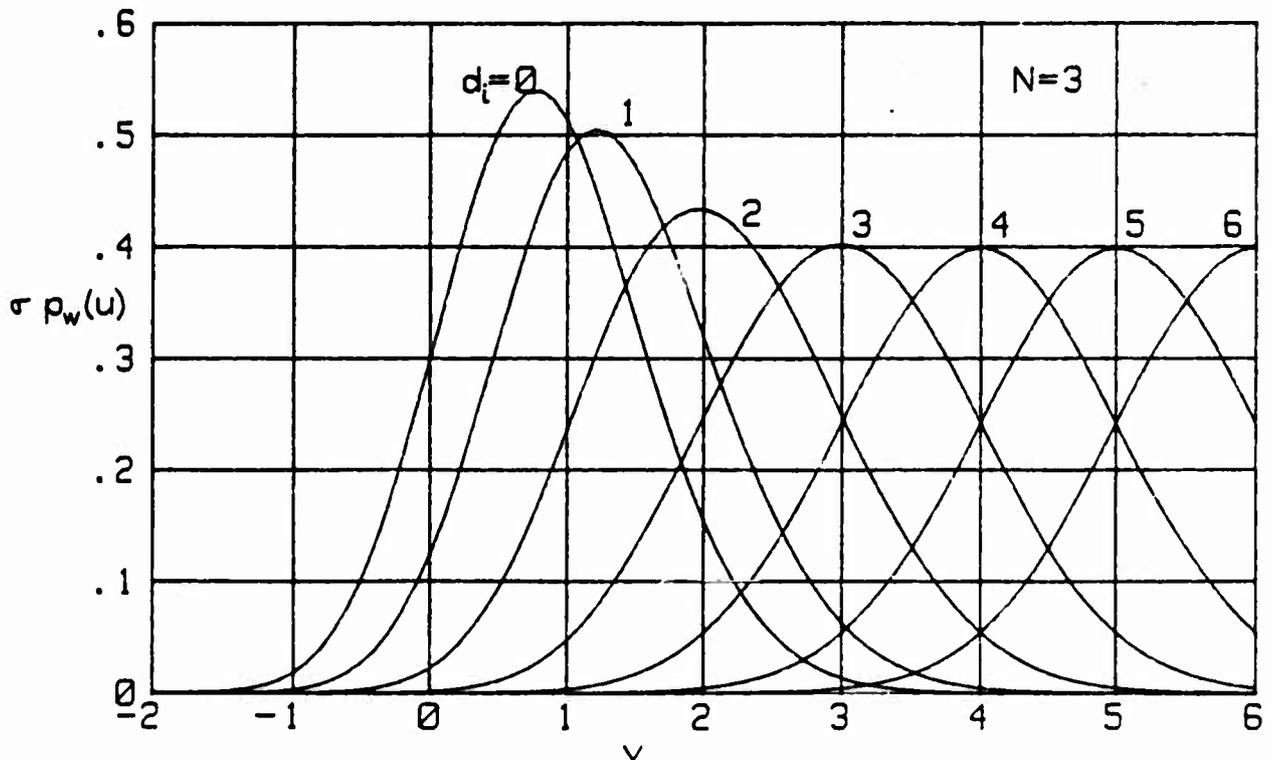
Figure 11.   Probability Density Function of w for N = 3

for $N=3$ and several values of d_i. If we select, for example, $d_i=5$, the points w_1 and w_0 in figure 8 correspond approximately to $v = 1.4$ and 4.1 in figure 11. Tne curve for $d_i=0$ is tne probability density function of w under H_0.

Tne false alarm probability, as given by (15) et seq., is plotted in figure 12 for an example with

$$N=3 \text{ input channels,}$$
$$M=5 \text{ time samples accumulated,}$$
$$L=7 \text{ quantizer breakpoints,}$$
$$\{v_\ell\} = 1.4(.45)4.1, \text{ normalized breakpoints.*} \qquad (25)$$

(Since $L+1=8$ here, this is called a 3-bit quantizer.) Since threshold J can only take on integer values, the false alarm probability is only defined at those values, and is so indicated by crosses in figure 12. The detection probability P_D is plotted vs. tne false alarm probability P_F , in figure 13,

by eliminating tne parametric dependence of botn on J. Again, both P_D and P_F are only defined at discrete points, as indicated by crosses; straight lines have been drawn between these points for ease of association of values. (In tne curves to follow, tnese crosses are suppressed.) Tne program for the generation of figures 12 and 13 is given in appendix C.

The cnoice of breakpoints in (25) has realized near-optimum performance for $L=7$ and $d_i \approx 5$, in the upper-left corner of tne plot. The curves for the smaller values of input signal-to-noise ratio, tnat is small d_i, are more crowded together; this reflects the usual small-signal suppression that is cnaracteristic of nonlinear processors.

To demonstrate tne effects of a bad choice of quantizer breakpoints, the previous example in (25) is rerun, with normalized breakpoints

$$\{v_\ell\} = 1,1,1,1,1,1,2, \qquad (26)$$

instead of the uniform spacing. Tne results in figure 14 illustrate much poorer performance tnan figure 13, in addition to a very erratic appearance.

---

* Tne notation $a(b)c$ denotes tne sequence a, $a+b$, $a+2b$, ..., $c-b$, c.
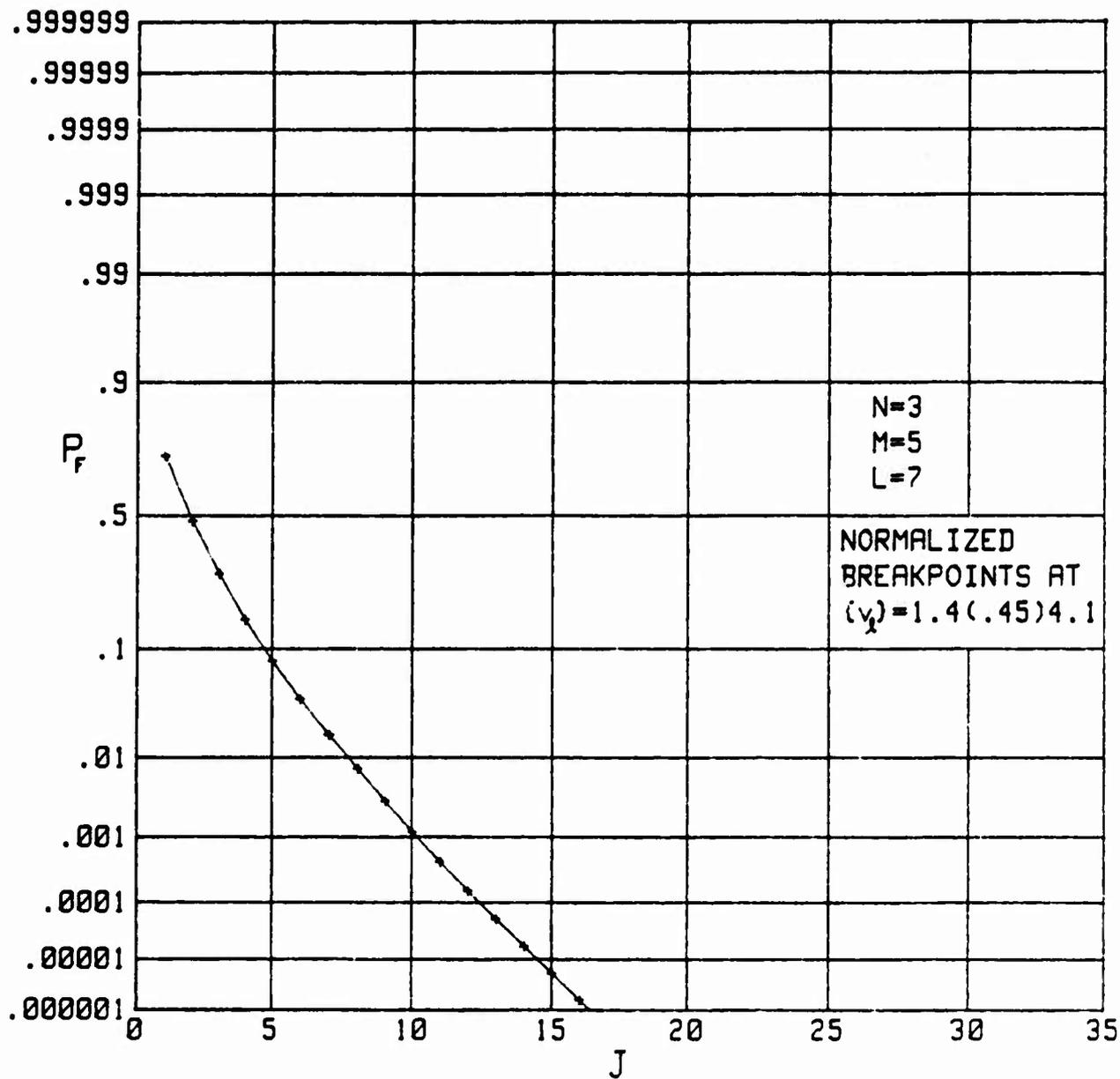
Figure 12.   Probability of False Alarm vs. Threshold J
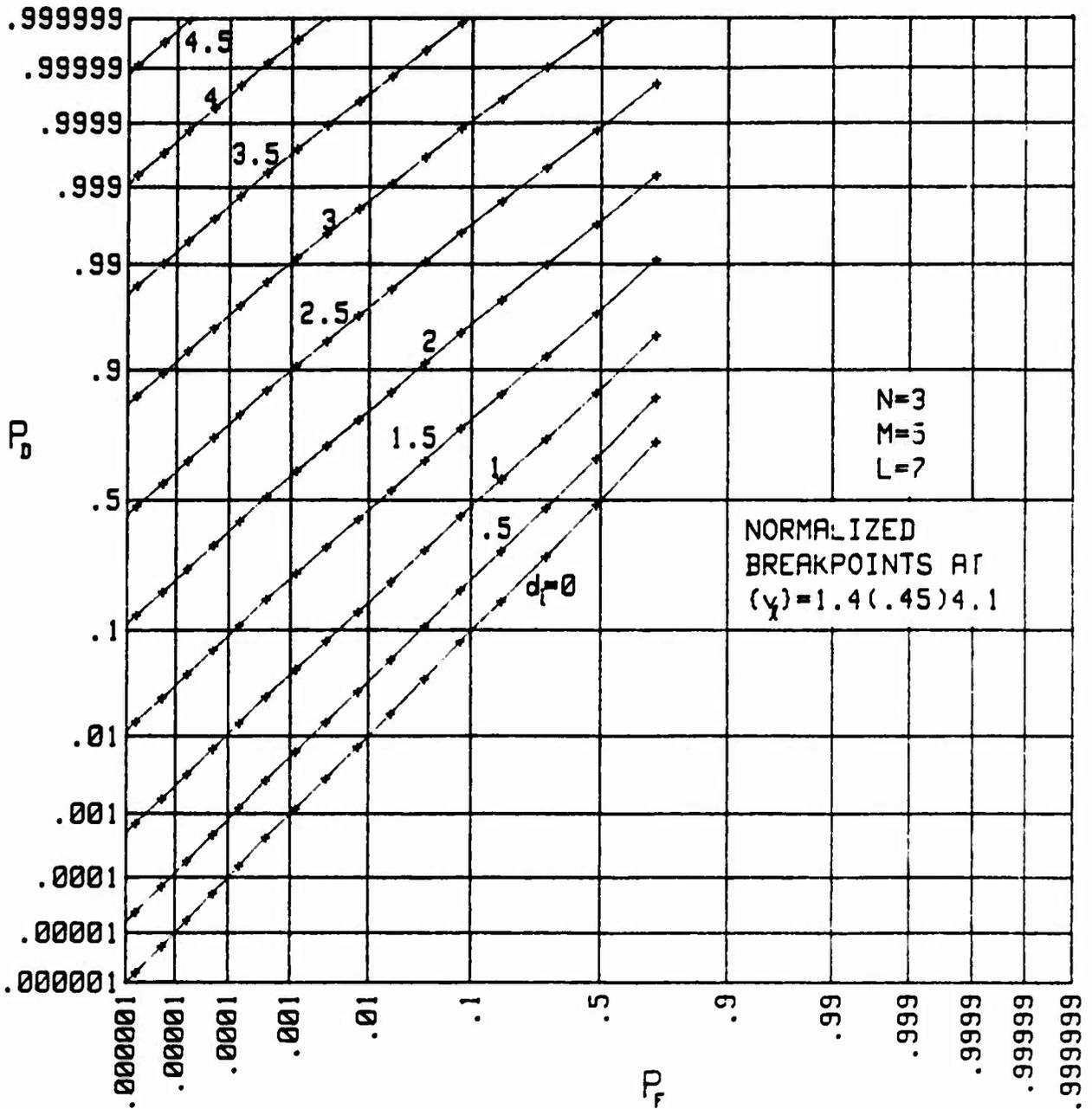
Figure 13.   Operating Characteristics for $\{v_\ell\}$ = 1.4(.45)4.1
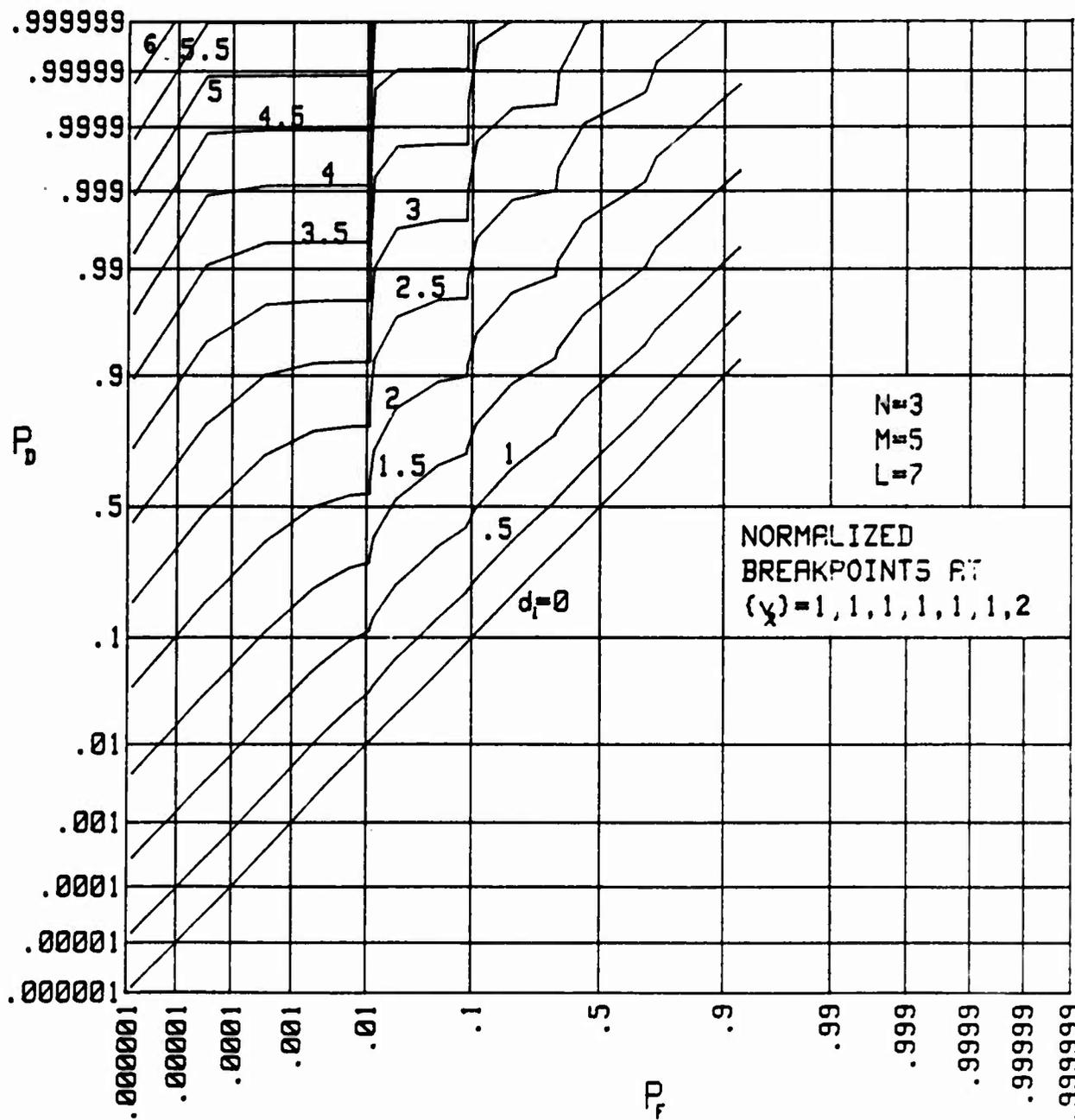
Figure 14.   Operating Characteristics for $\{v_\ell\}$ = 1,1,1,1,1,1,2

The curves for constant d_i must have a nonnegative slope, and never cross, but can have a wide variety of shapes. A similar bad placement for a 2-bit (L=3) quantizer is given in figure 15. This jagged behavior of the operating characteristic is typical when the breakpoints are bunched together instead of being uniformly spaced.

The next example is for

$$N=3, \; M=16, \; L=7, \; \{v_\ell\} = 0(1/3) \; 2 \quad , \qquad (27)$$

and is displayed in figure 16. This can be compared directly with the results of ref. 2, figure 5. Both require $d_i=1.22$ for $P_F=10^{-3}$, $P_D=.5$, and both require about $d_i=1.75$ for $P_F=10^{-6}$, $P_D=.5$. Thus the simplified analysis in ref. 2 is very reliable for large M, where it is reasonable to expect Gaussian statistics to hold.

In order to see if better performance is attainable by modifying the breakpoints, we return to the probability density function in figure 11 (for N=3) and observe that if we want to optimize for $d_i=3$ (i.e., top left corner of figure 16), we should choose $w_1 \approx 0$, $w_2 \approx 4$. However, the narrowing effect on the decision-variable probability density function, due to the averaging caused by large M, indicated (by trial and error) that the best normalized breakpoints were

$$\{v_\ell\} = .5(.5)3.5 \quad . \qquad (28)$$

The corresponding operating characteristic is displayed in figure 17; it is slightly better than figure 16. Thus the significant modification in normalized breakpoints from figure 16 to figure 17 did not yield significantly better performance, for this example with L=7.

The next example is run for comparison with ref. 2, figure 8. Namely we have

$$N=3, \; M=32, \; L=7 \quad , \qquad (29)$$

and three different quantizer breakpoint sets:

24
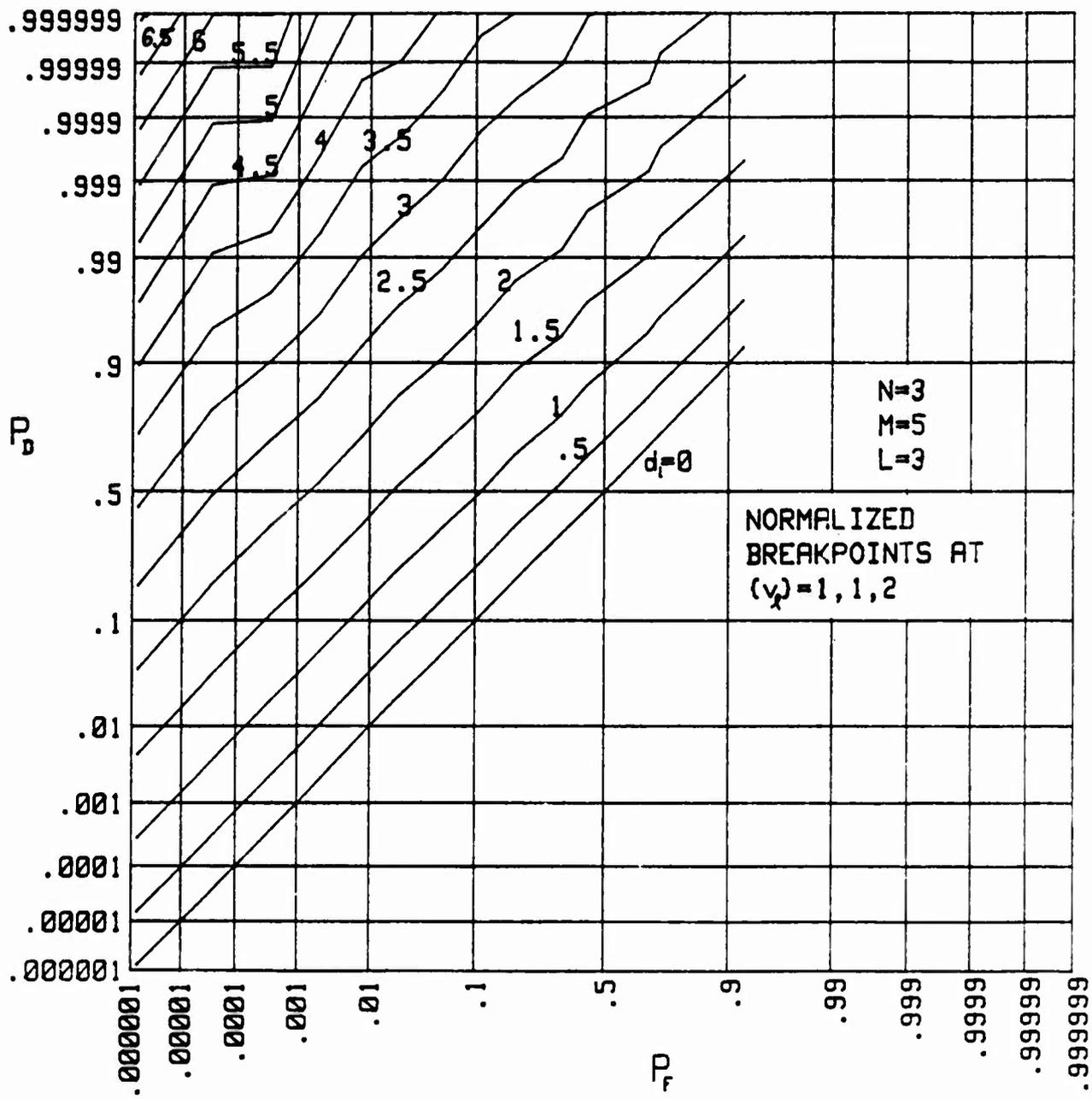
Figure 15. Operating Characteristics for $\{v_\ell\}$ = 1,1,2
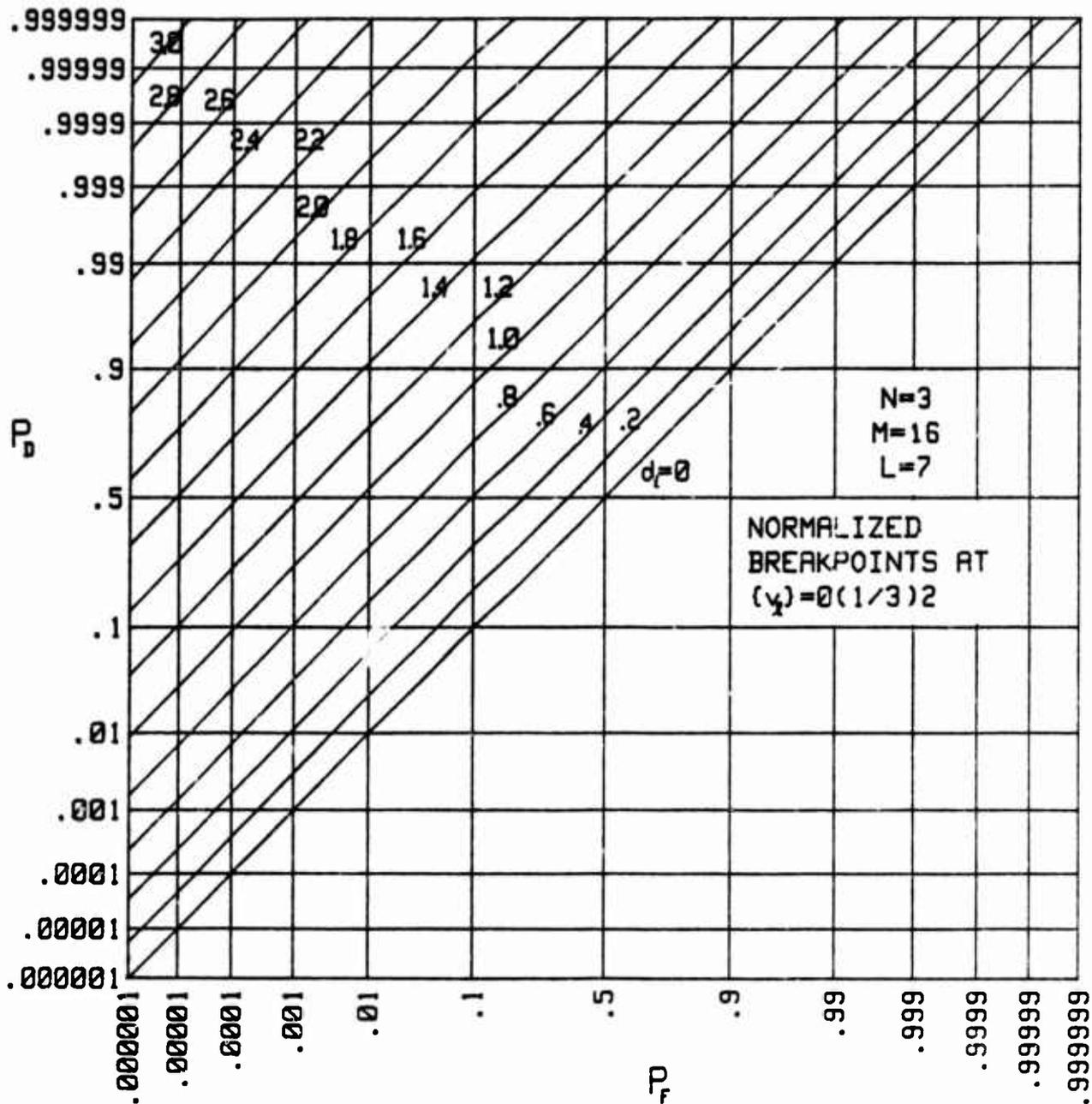
Figure 16. Operating Characteristics for M = 16, $\{v_\ell\}$ = 0(1/3)2
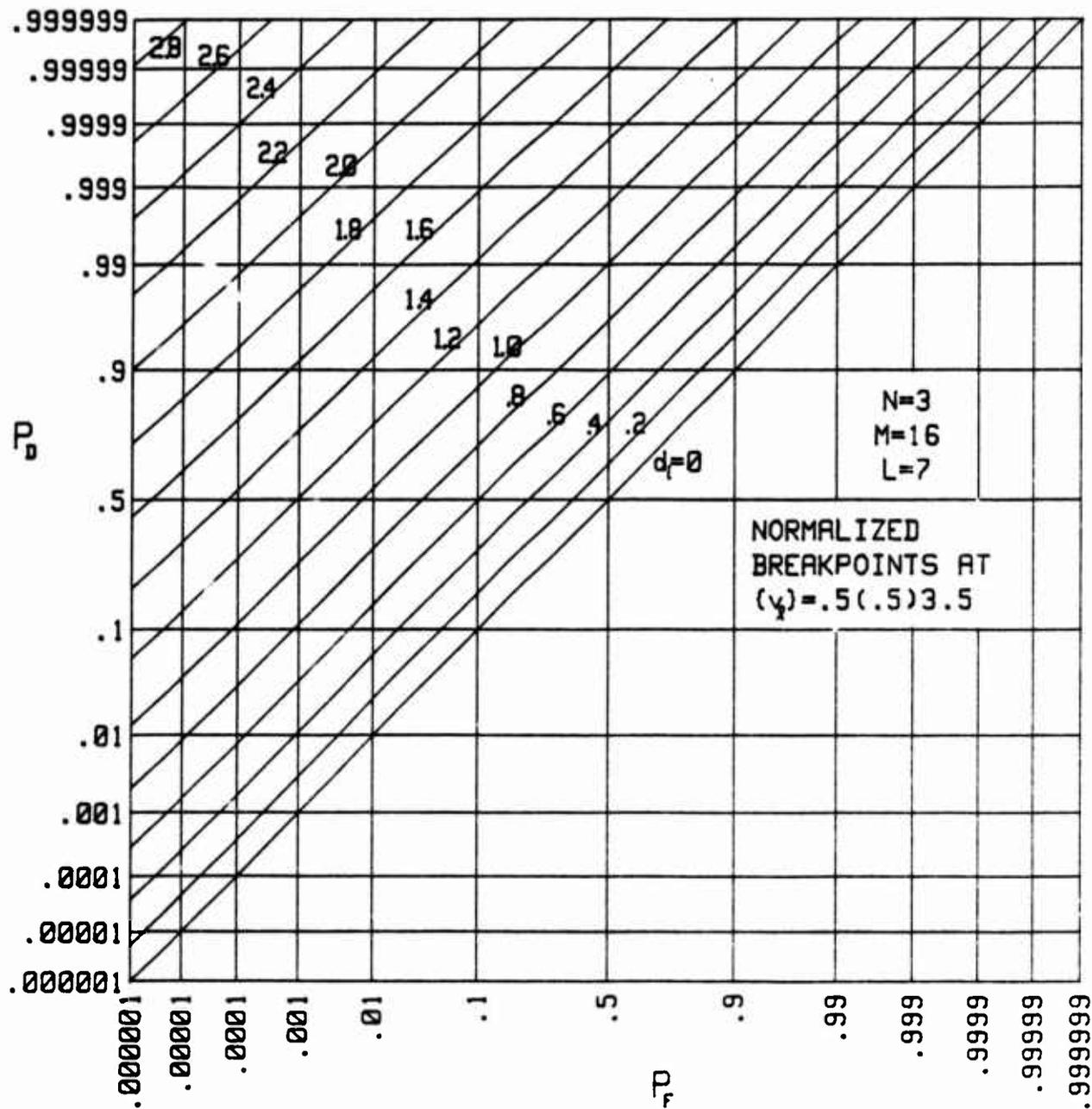
Figure 17.   Operating Characteristics for M = 16, $\{v_\ell\}$ = .5(.5)3.5

$$\{v_\ell\} = \begin{cases} 0(1/3)2 & \text{in figure } 18 \\ 0(.5)3 & \text{in figure } 19 \\ 0(.7)4.2 & \text{in figure } 20 \end{cases} . \tag{30}$$

The results in figures 18 through 20 are in almost perfect agreement with the simplified analysis in ref. 2. The normalized breakpoints in figure 19 are best in the intermediate range of detection probabilies; figure 18 is just slightly better for the high performance region of $d_i = 2.2$.

We now present a series of comparisons where some of the parameters are held constant, while the remainder are varied in order to determine the effect upon the operating characteristics. The first comparison is for M=5, N=10, and the quantizer varied as follows:

$$\begin{array}{lll} L=7, & \{v_\ell\} = 1.4(.45)4.1 & \text{in figure } 21 \\ L=3, & \{v_\ell\} = 2(.75)3.5 & \text{in figure } 22 \\ L=1, & v_\ell = 2.75 & \text{in figure } 23 \end{array} . \tag{31}$$

These correspond to 3-bit, 2-bit, and 1-bit quantizers respectively, where the normalized breakpoints have been chosen in each case so as to optimize the performance for $d_i = 4$. Increasing L beyond 7, and changing the breakpoints, failed to improve the operating characteristic noticeably above that of figure 21. The increase in the input deflection, d_i, needed to maintain the same performance at $P_F = 1 - P_0 \approx 10^{-3.5}$ is approximately 4.5/4 for L=1 vs L=7. If inputs $\{x_n\}$ are interpreted as voltages, this corresponds approximately to a 1 dB degradation for the hard clipper, L=1.

The next series is for M=5 and the quantizer fixed at the 3-bit characteristic in the first line of (31). N is varied over the values 1, 5, 10, 20, 40 in figures 24-28, respectively. The slight improvement in performance, that might ensue from modifying the breakpoints at each N, was not investigated in this comparison. The effect of increasing N is to degrade the performance, since the or-ing must select one of the input channels for accumulation, and it will not always pick the signal-bearing channel. The increase in d_i required to maintain $P_F = 1 - P_0 \approx 10^{-5}$ is approximately 5.25/4 = 2.35 dB for N=40 vs N=1.
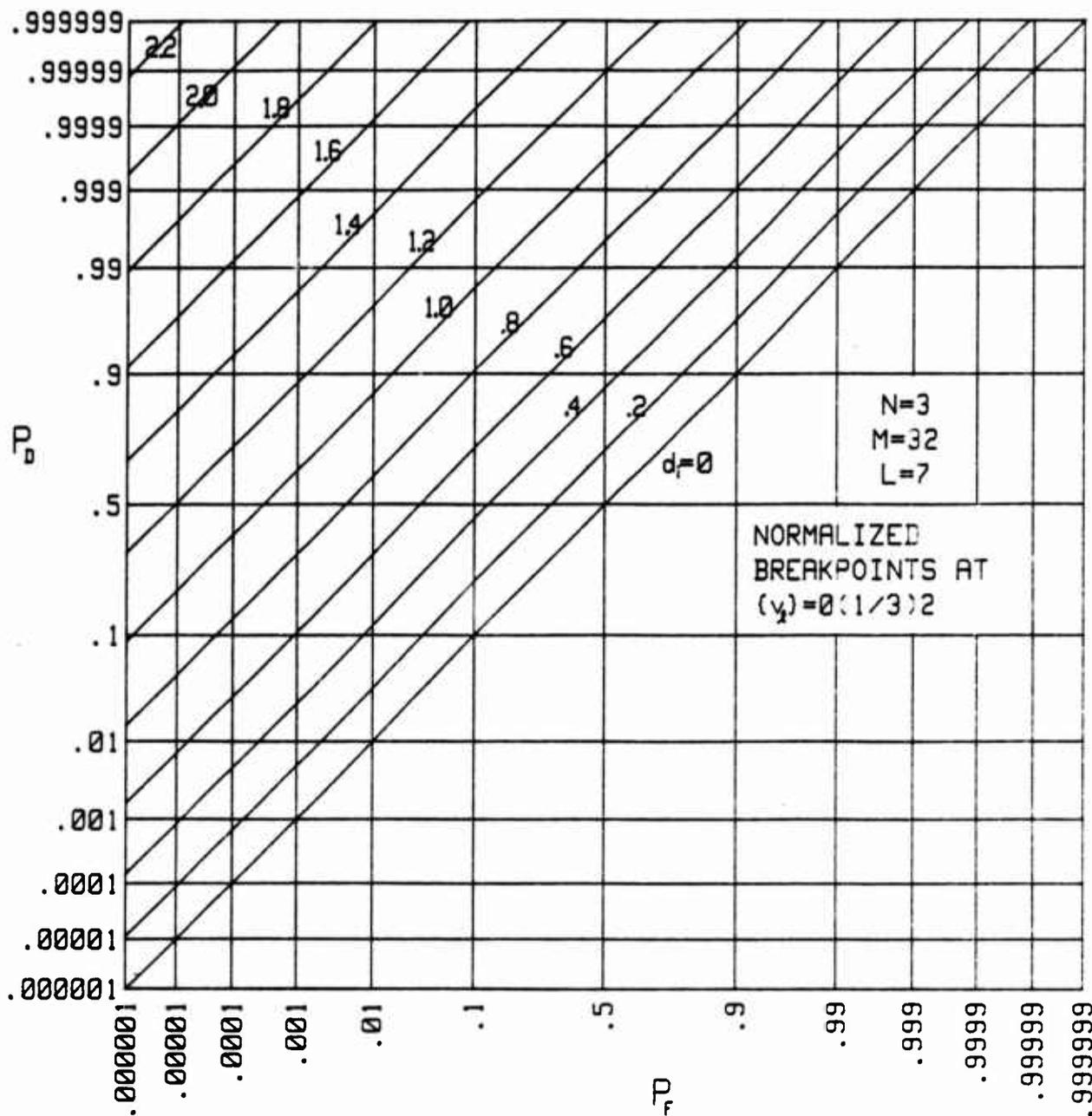
Figure 18.   Operating Characteristics for M = 32, $\{v_\ell\}$ = 0(1/3)2
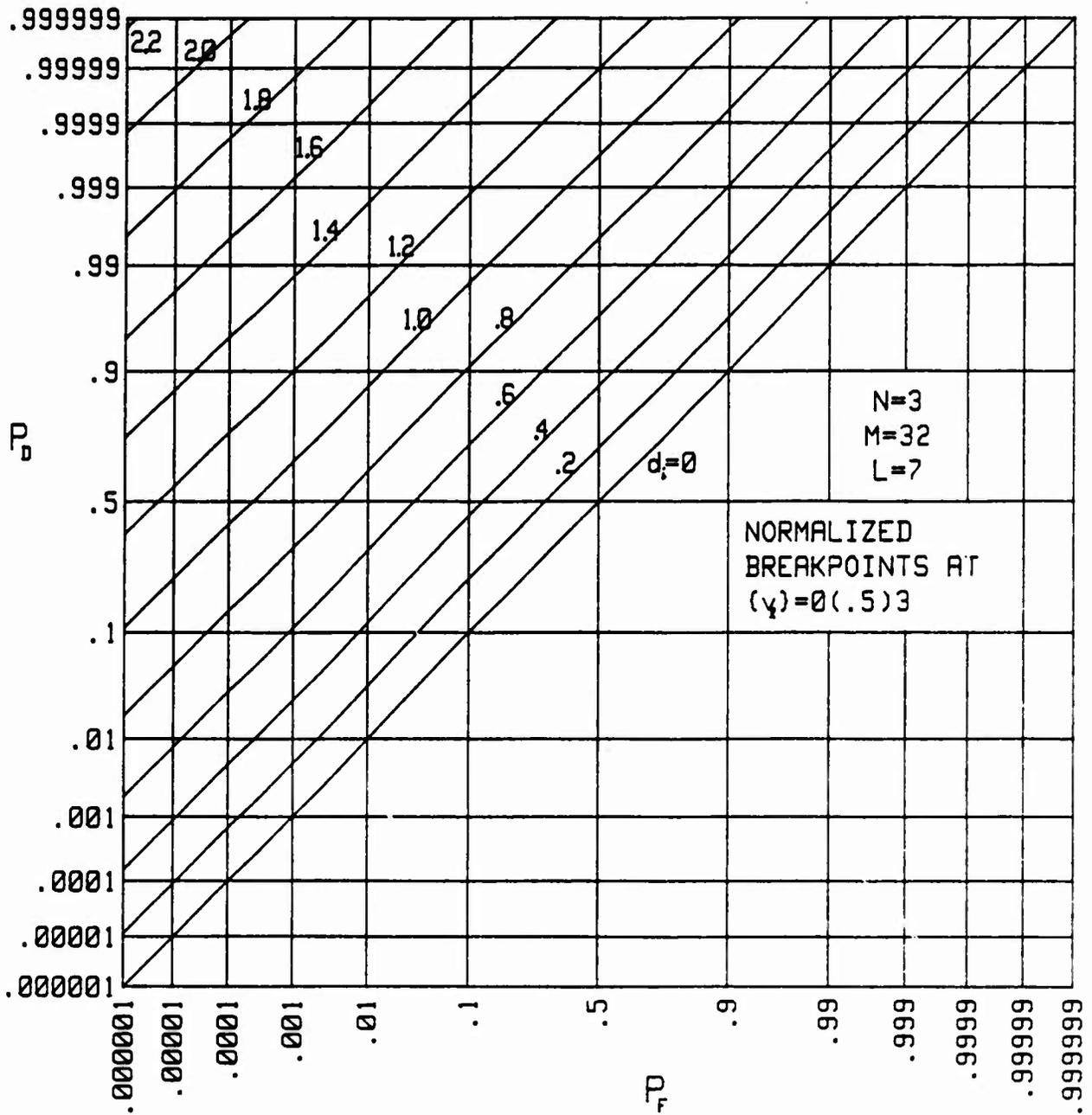
Figure 19. Operating Characteristics for M = 32, $\{v_\ell\}$ = 0(.5)3
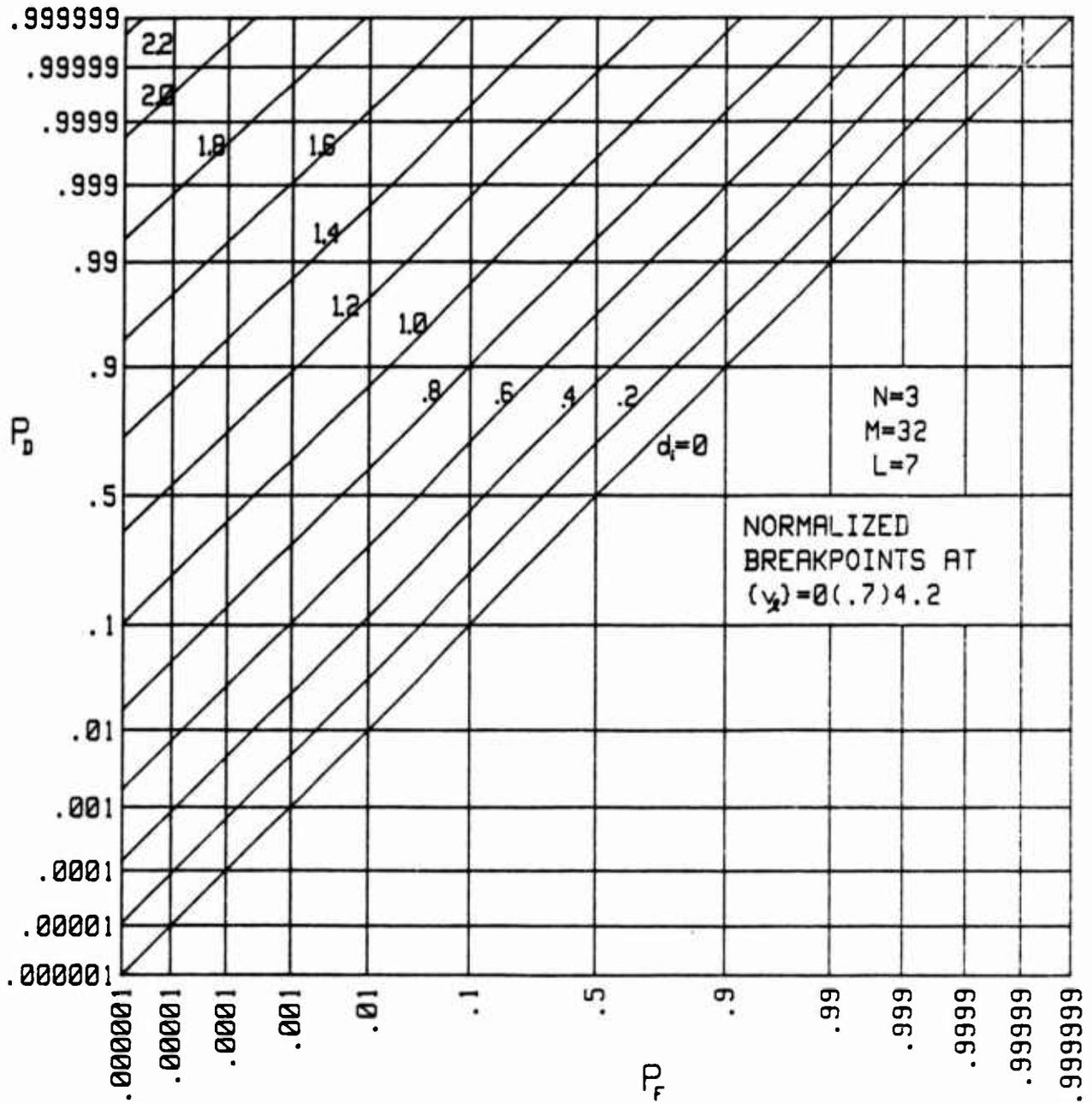
Figure 20. Operating Characteristics for M = 32, $\{v_\ell\}$ = 0(.7)4.2
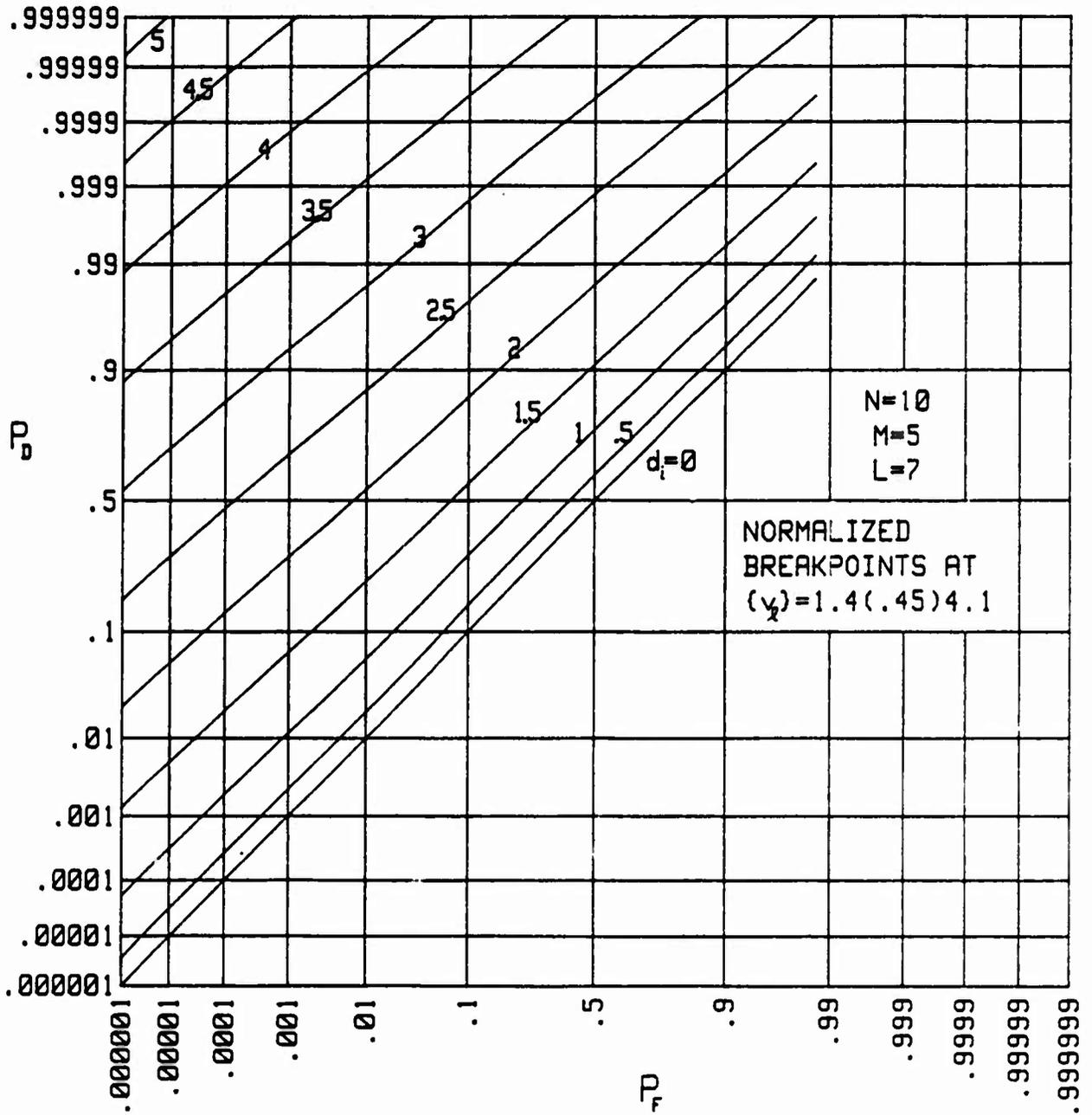
Figure 21. Operating Characteristics for L = 7
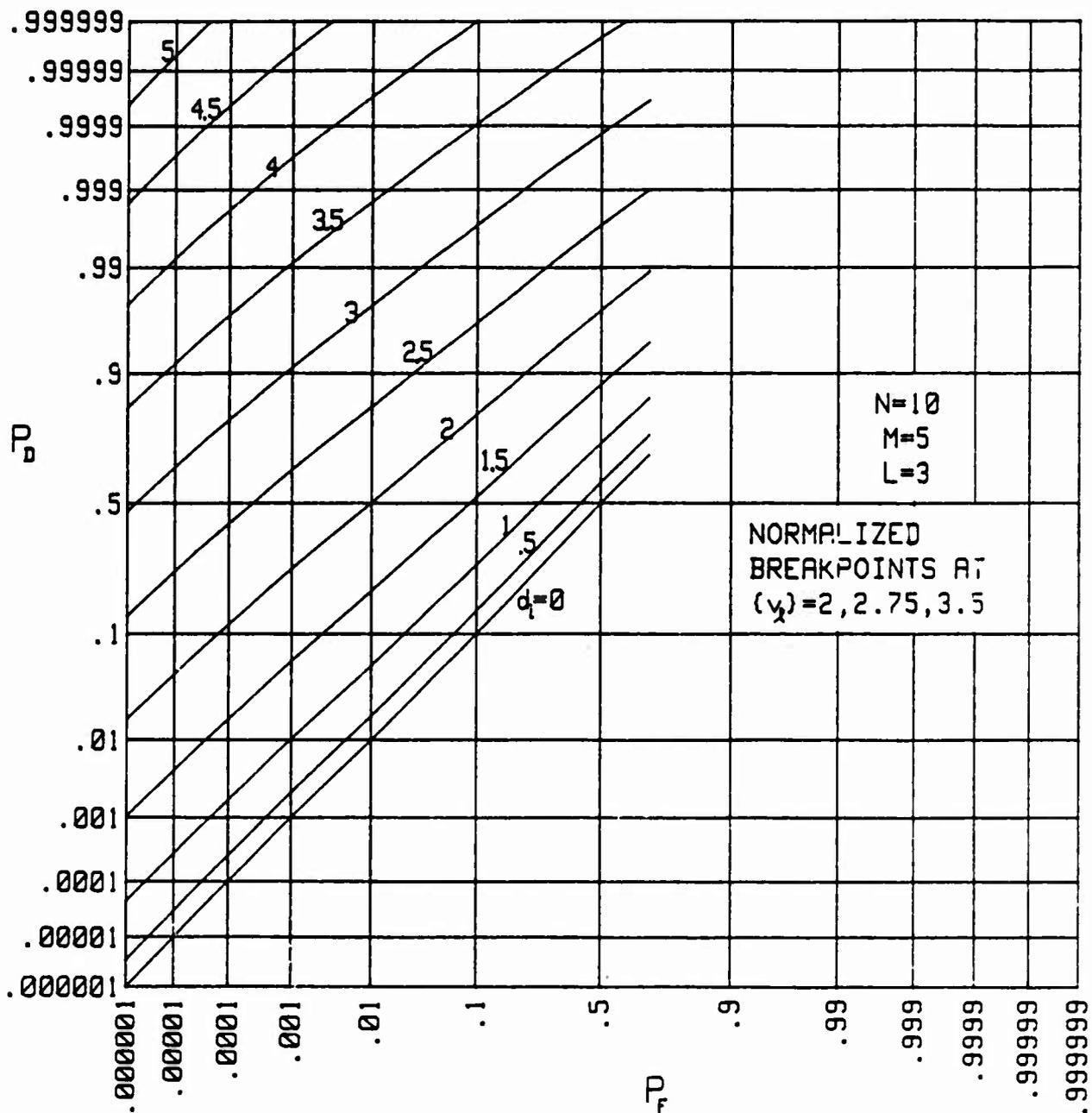
Figure 22. Operating Characteristics for L = 3
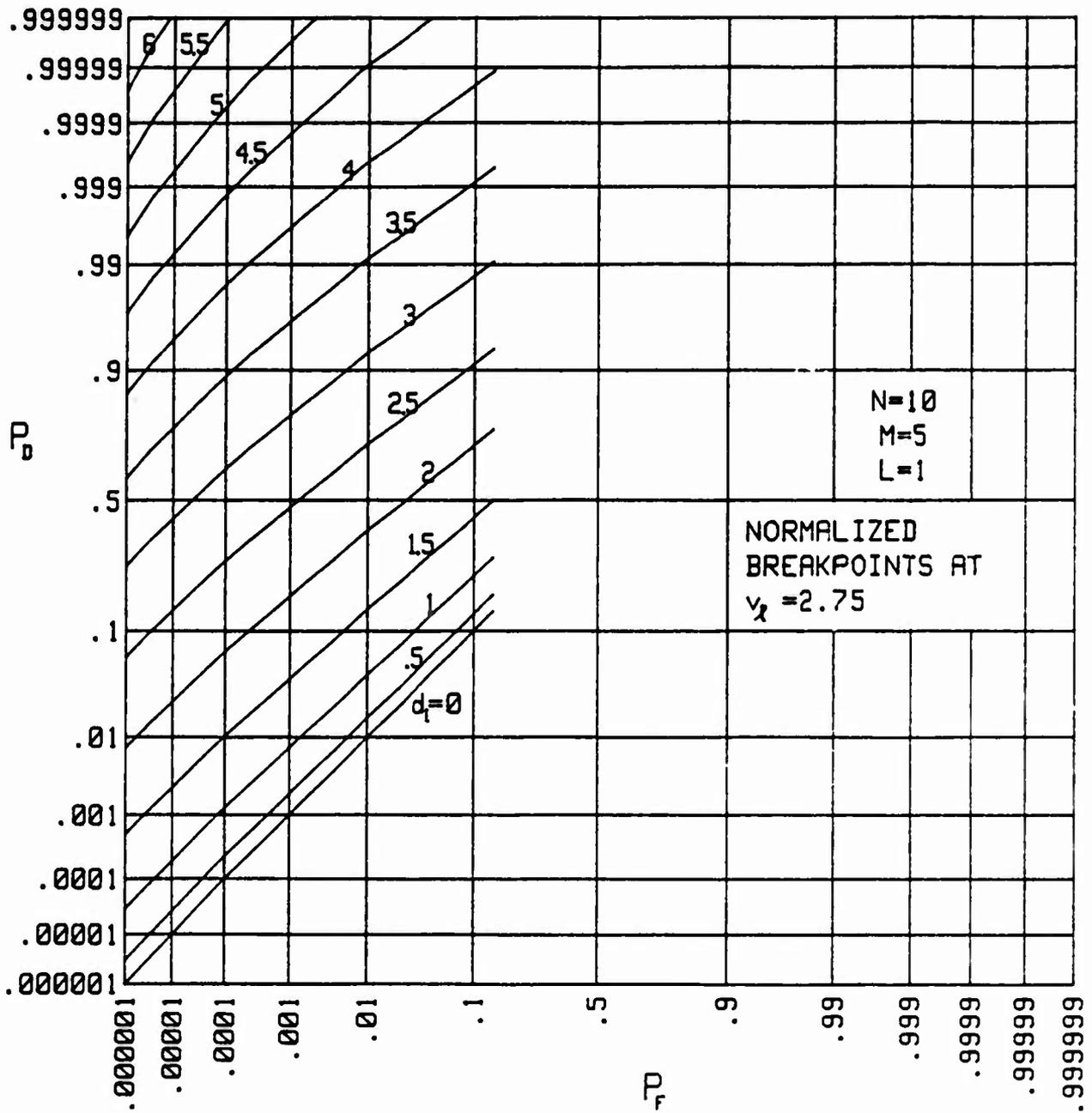
Figure 23.   Operating Characteristics for L = 1
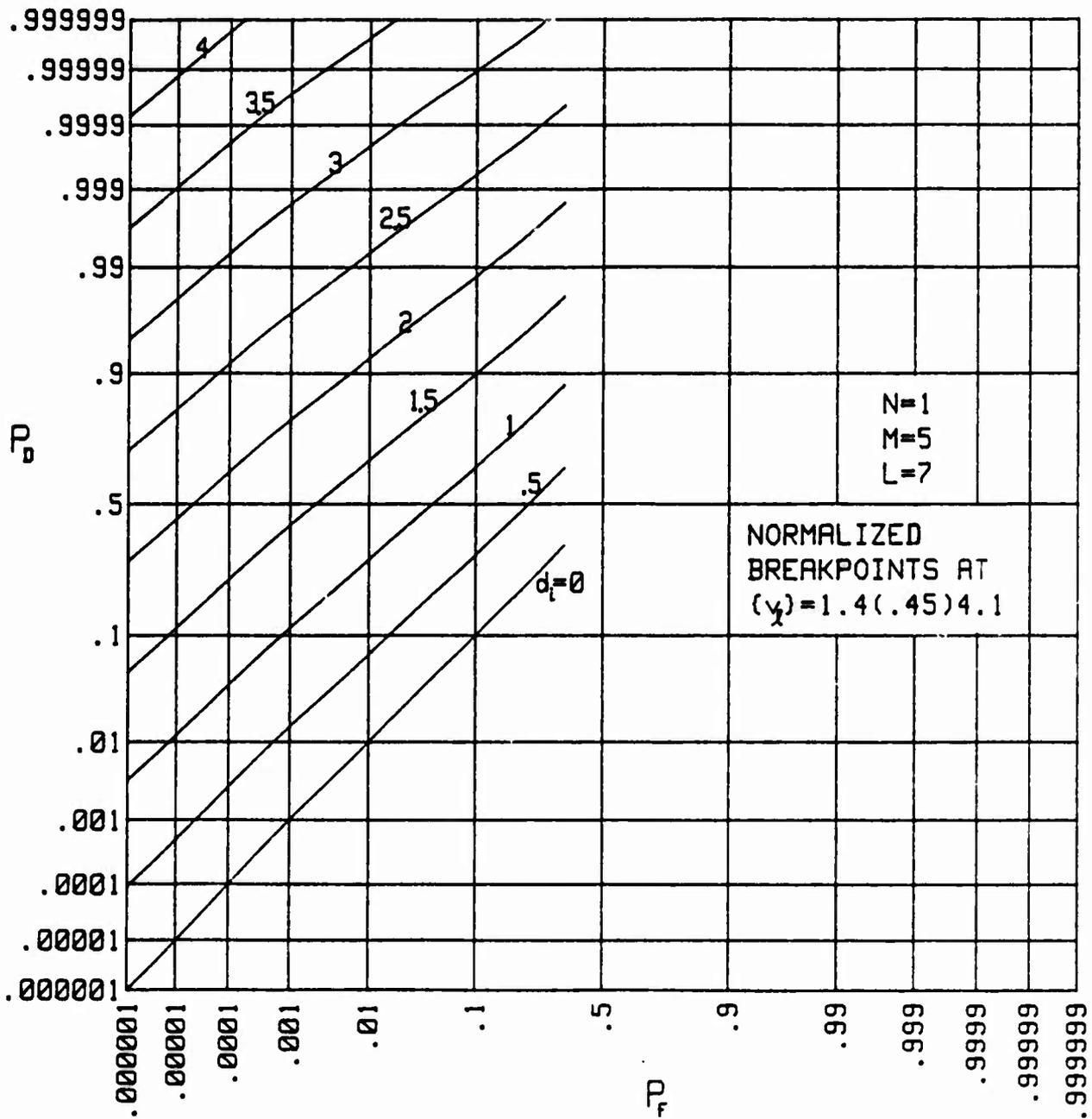
Figure 24.   Operating Characteristics for N = 1

Figure 25.   Operating Characteristics for $N = 5$
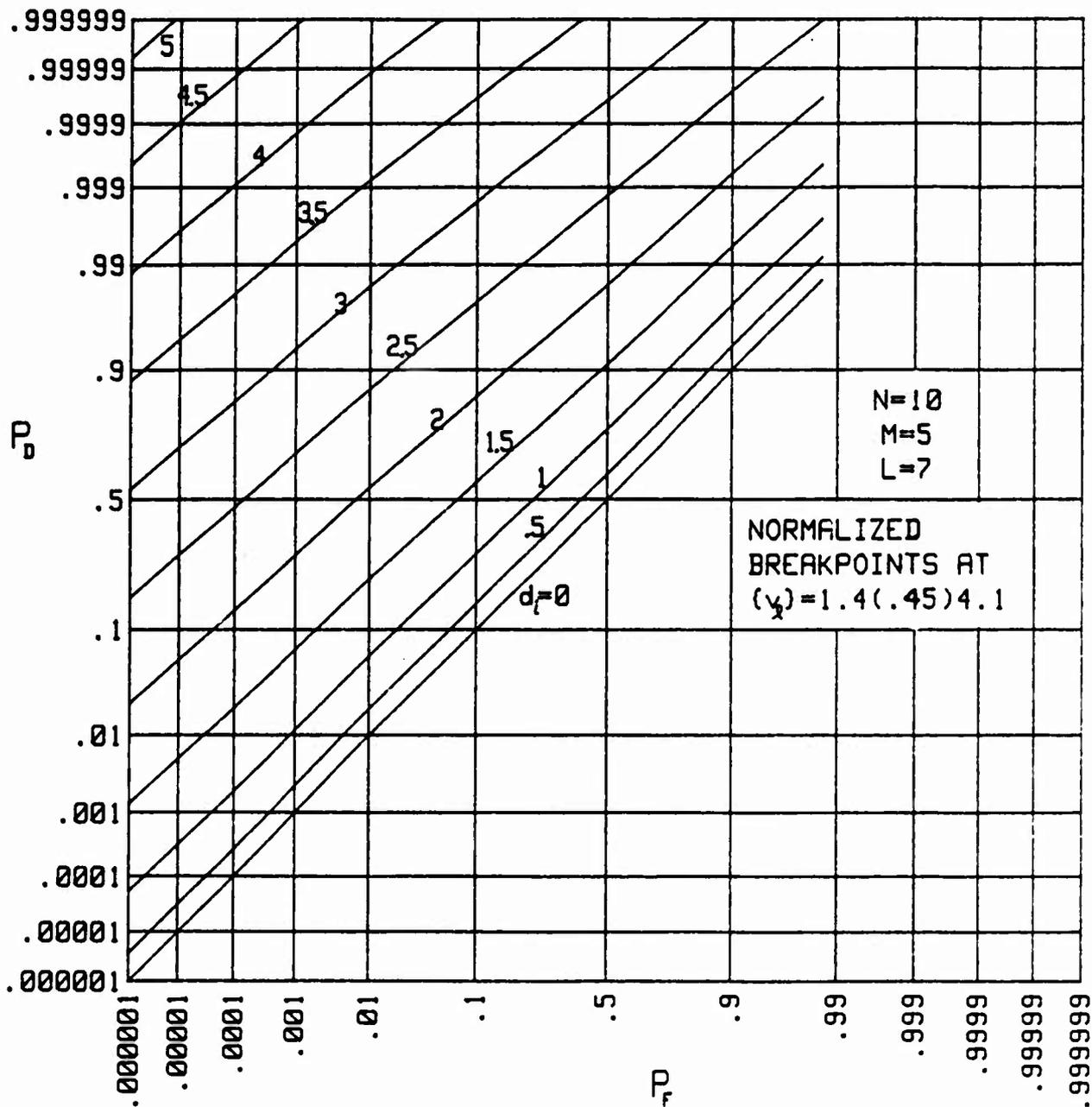
Figure 26. Operating Characteristics for N = 10
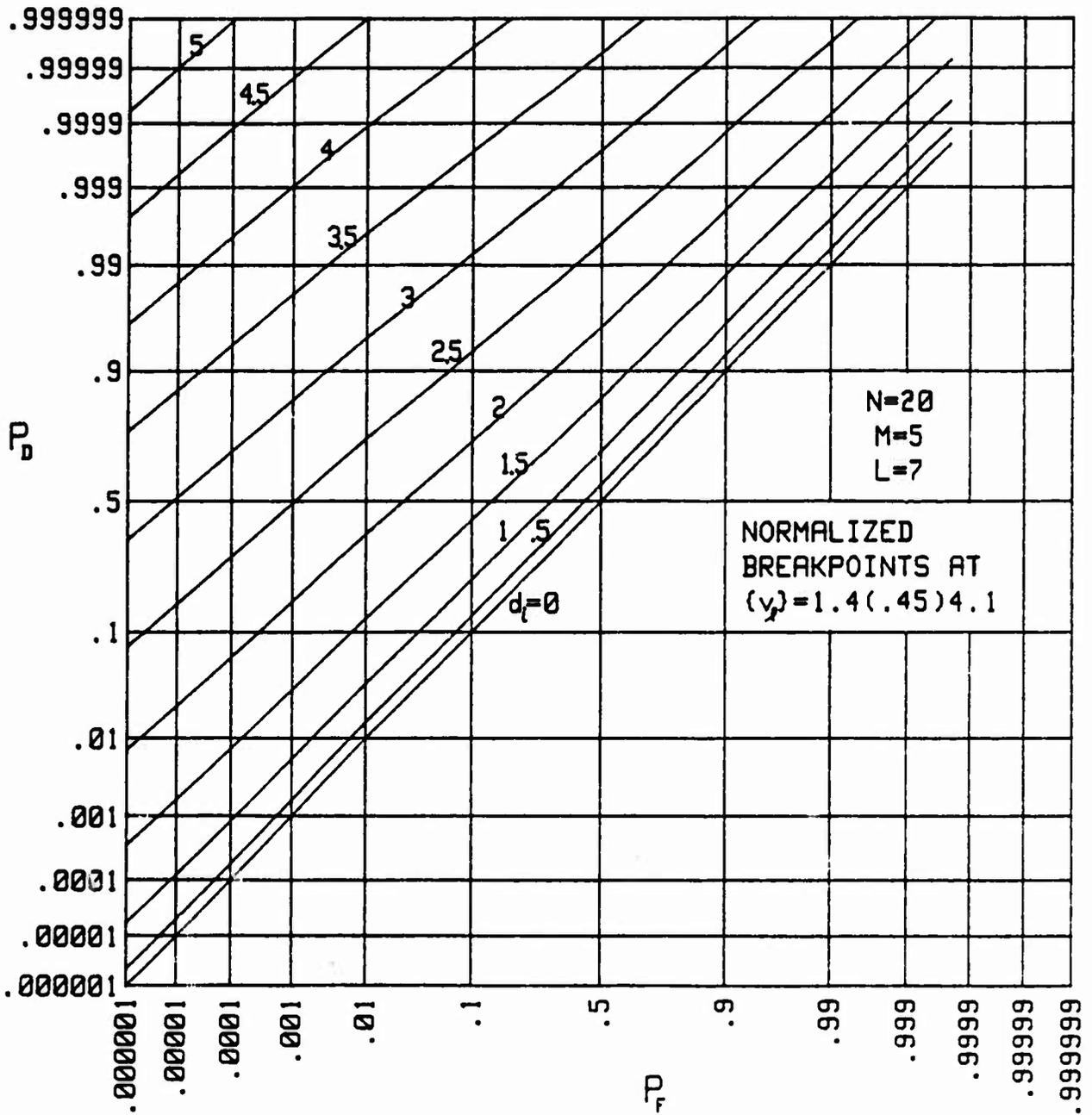
Figure 27. Operating Characteristics for N = 20
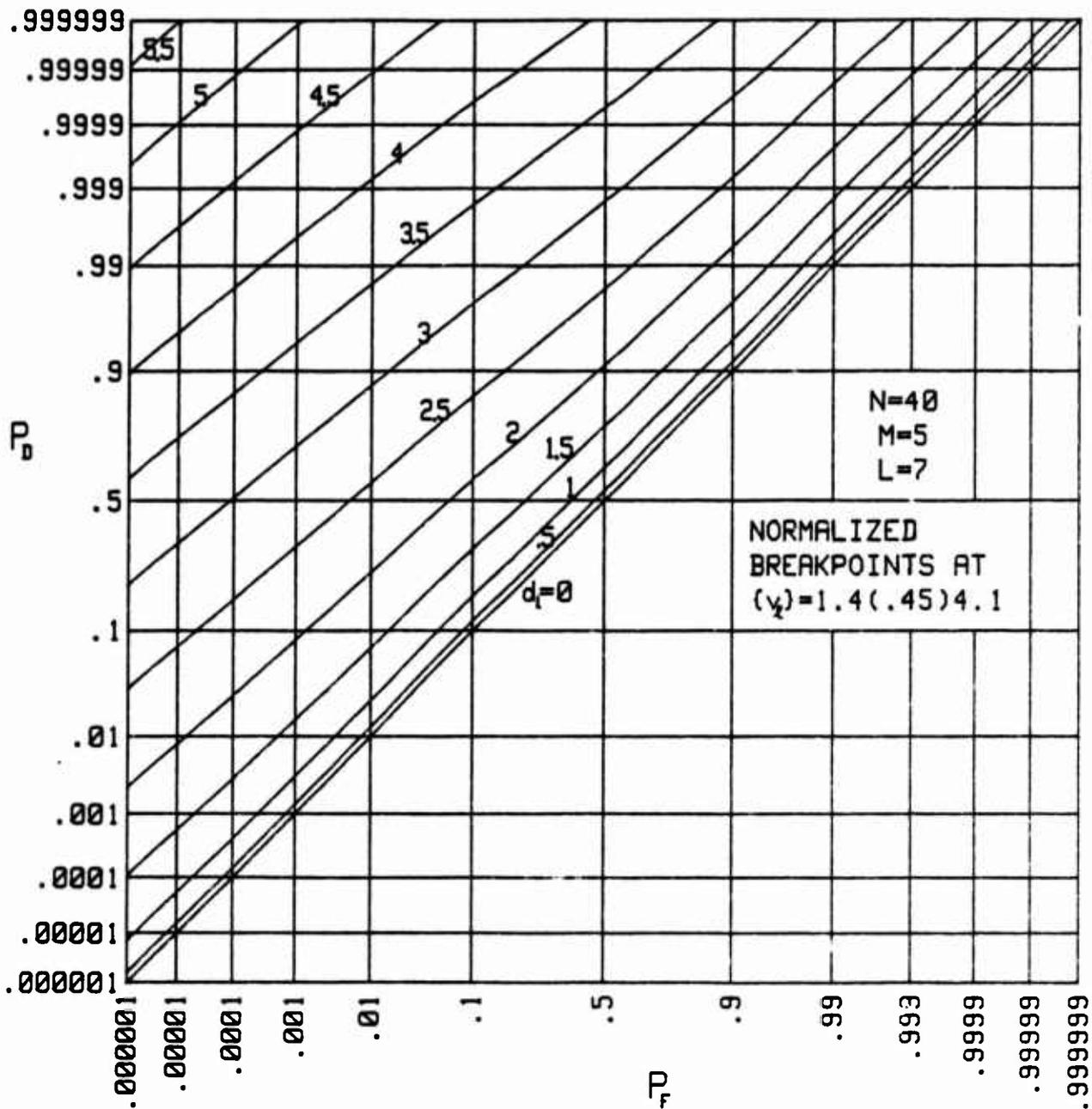
Figure 28.   Operating Characteristics for N = 40

The last series investigates the effect of varying M. In figures 29 through 33, M takes on the values 1, 2, 5, 10, and 20, respectively. The effect of increasing M is to realize better performance with smaller input signal-to-noise ratios. The decrease in d_i allowed, in order to maintain $P_F = 1 - P_D \approx 10^{-3}$, is approximately 2.4/6.8 = -9 dB for M=20 vs M=1. When M is made large, the approximate analysis in ref. 2 can be used with confidence; this is fortunate, since in the case of very large M, the DFT size, $N_f \geq$ ML+1, required here may not be easily attainable on some computer setups.
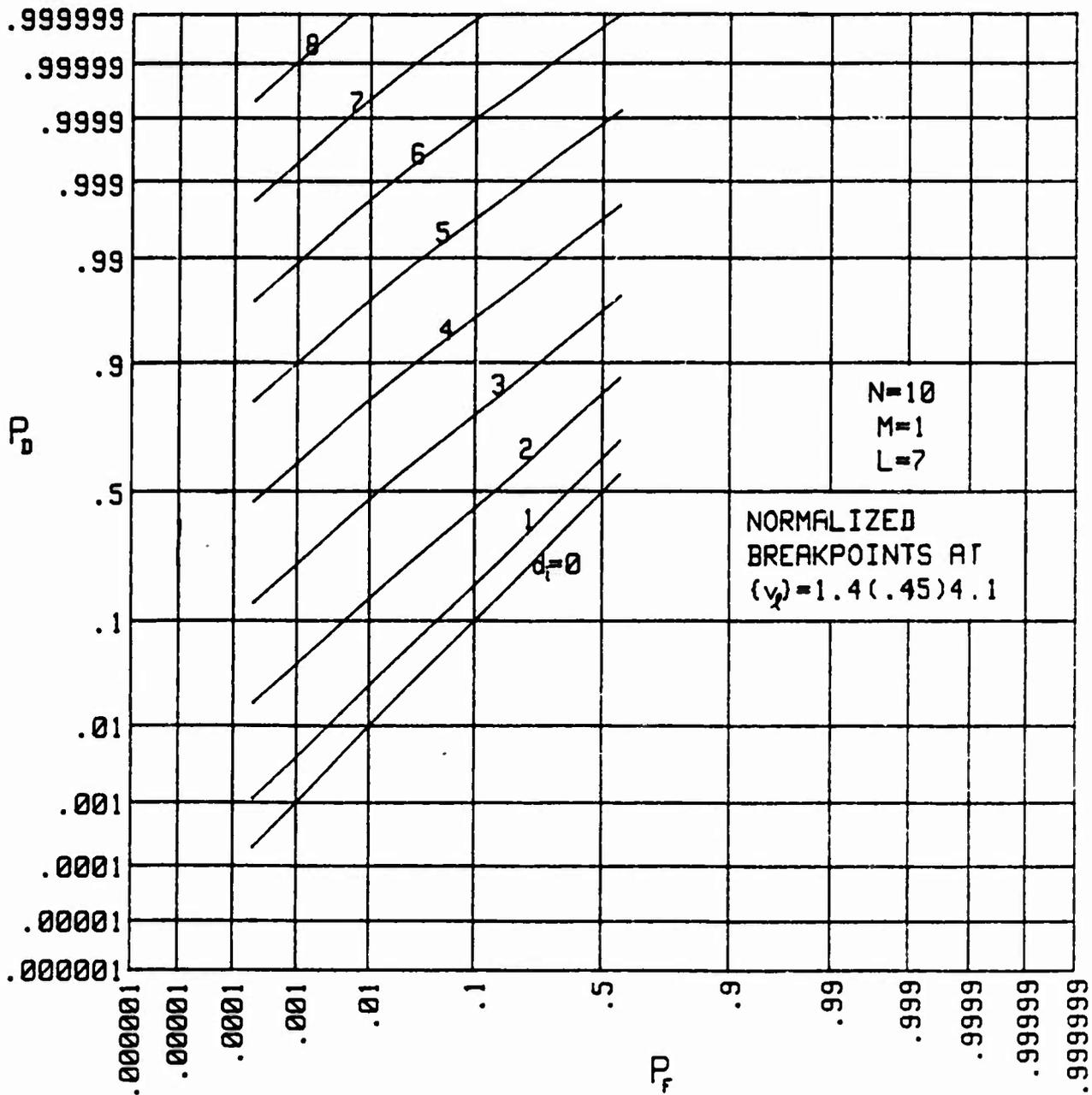
Figure 29. Operating Characteristics for M = 1
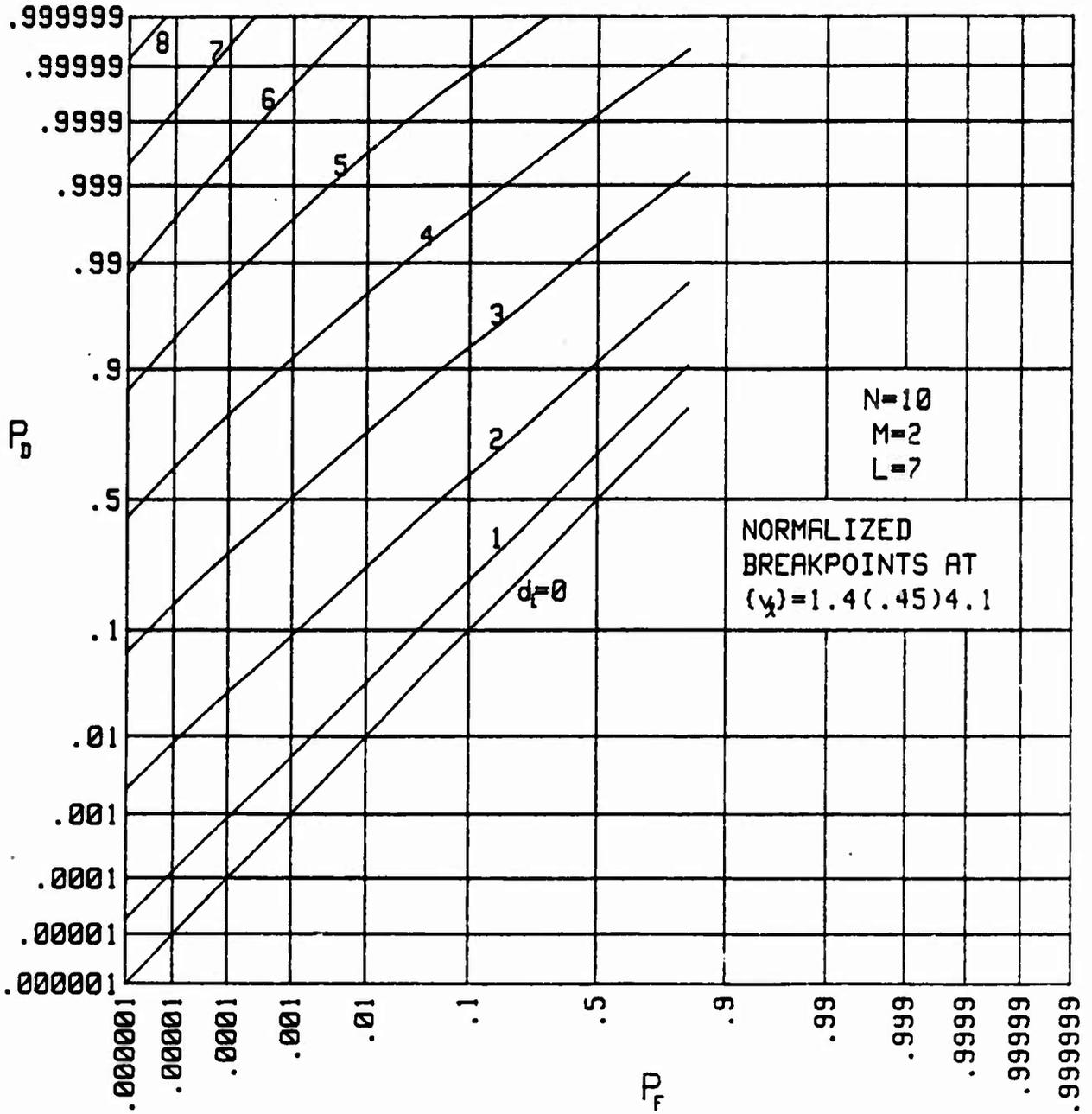
Figure 30. Operating Characteristics for M = 2
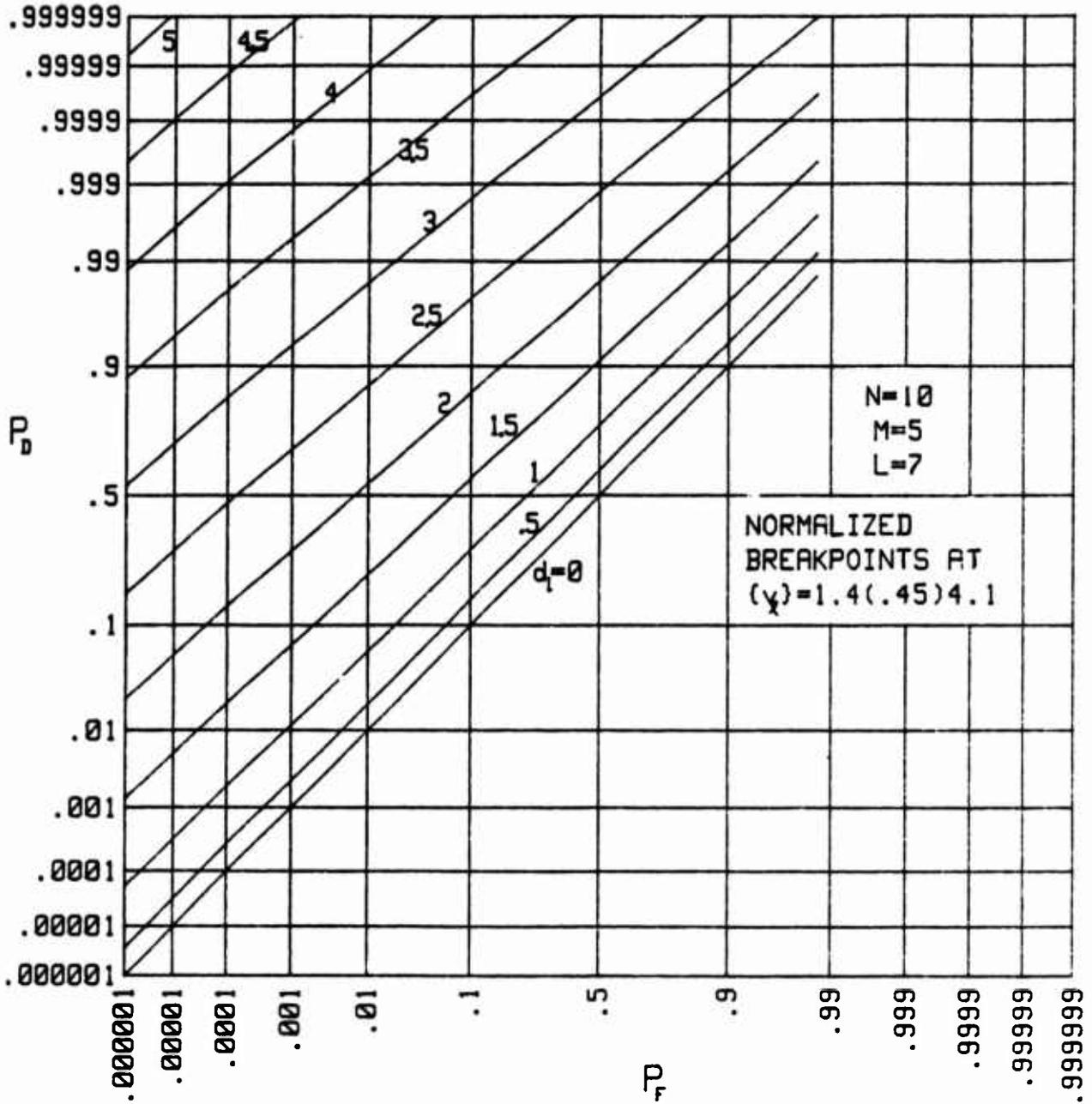
Figure 31. Operating Characteristics for M = 5
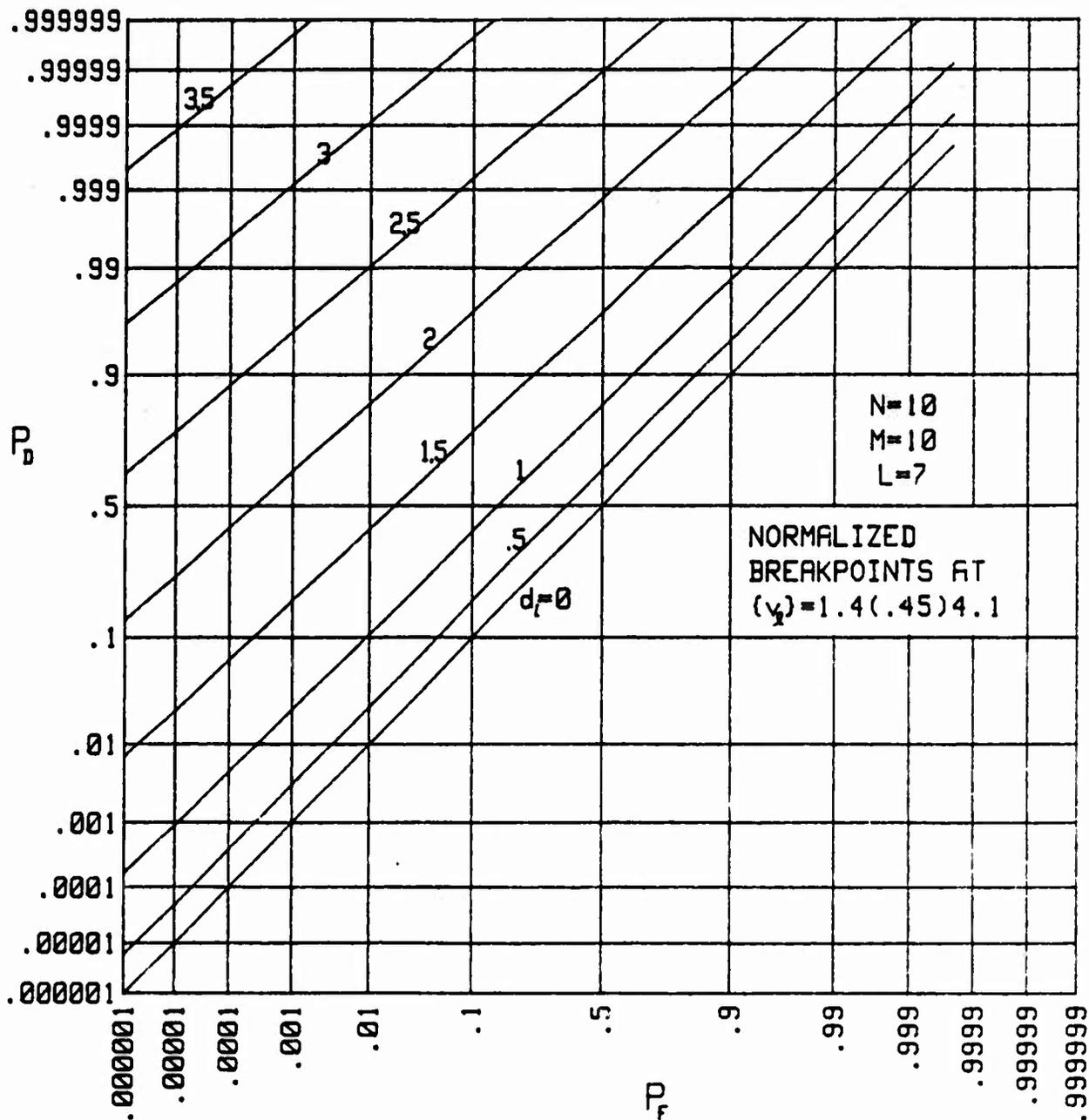
Figure 32.   Operating Characteristics for M = 10
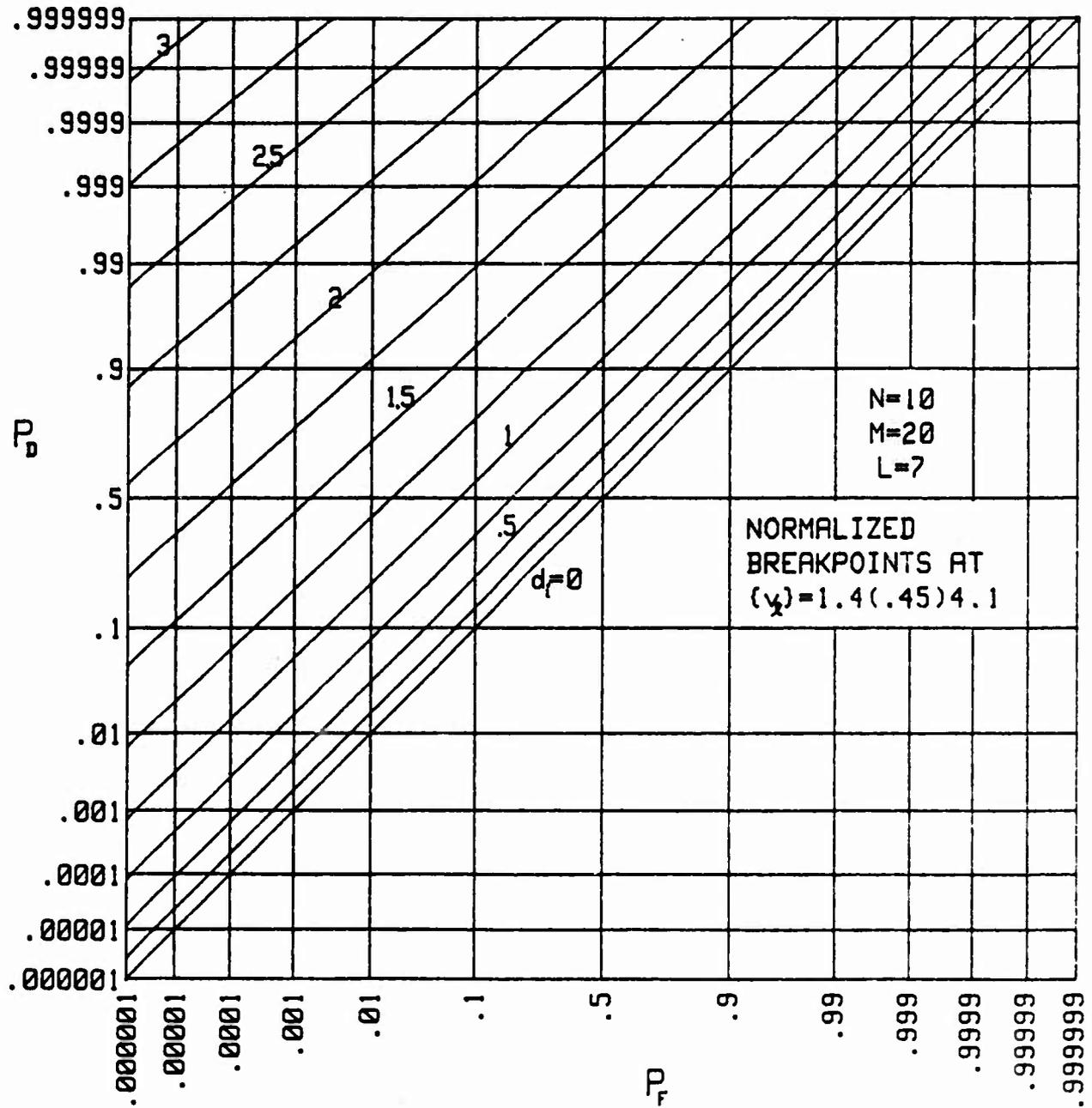
Figure 33. Operating Characteristics for M = 20

DISCUSSION

Tne large number of possible combinations of values of N, M, L, and $\{y_\chi\}_1^L$ precludes an exhaustive compilation of results. Instead we have presented some representative examples and give a program in appendix C by which the user can investigate his particular situation and alternatives. This program gives exact results for any quantizer, provided only that ML is not too large and that round-off errors do not get out of hand. For extremely large M, a Gaussian approximation for the decision variable is justified and the analysis of ref. 2 is applicable.

If the quantizer is not specialized to the equal ordinate spacings of figure 3, but is of the general form depicted in figure 2, the performance analysis is more difficult. However, the characteristic function of the decision variable is still capable of a closed form expression; see appendix D. Evaluation of tne cumulative distribution function of the decision variable is possible via one FFT, according to the methods given in refs. 6 and 7. No numerical investigation nas been undertaken of this case.

In appendix E, the form of the optimum processor operating on N input cnannels, of whicn only one may contain a signal, is derived and then specialized to the Gaussian input example of (16). This optimum processor, in general, requires knowledge of the absolute levels of the input signal and noise. However, for $d_i > 2$, tne form of the optimum processor approaches that of figures 1 or 5, where the quantizer is replaced by a linear device, and tne absolute level knowledge is no longer required. Tnus the performance of tne system considered here should be nearly optimum for large L and well-placed breakpoints.

Exact analysis of the system of figures 1 or 5, with a linear gain instead of a quantizer, is <u>more</u> difficult than that given here. A simplified second-moment analysis was presented in ref. 1; an exact analysis is possible and will be presented by the author in a future report.

A detailed comparison of the exact performance results obtainable via this report, with the second-moment results given in refs. 1-5, has revealed excellent agreement over a wide range of parameter values. However, until an extensive thorough investigation of the two approaches is made for a wide range of values of N, M, L, $\{v_\ell\}$, P_F, and P_D, it is difficult to state exactly where the earlier approximate analyses can be used with full confidence. This time-consuming investigation has not been undertaken; however, the program in appendix C affords the mechanism whereby this comparison can be conducted. The statements here regarding large M and moderate P_F can only be made quantitative after this study is completed.

Appendix A

INTERRELATIONS BETWEEN CHARACTERISTIC FUNCTIONS AND
PROBABILITY DENSITY FUNCTIONS OF DISCRETE RANDOM VARIABLES

Suppose random variable x is limited to the integer values 0, 1, ..., N, and that the probability of taking on value n is a_n. That is, the probability density function of x is

$$p_x(u) = \sum_{n=0}^{N} a_n \, \delta(u-n) \quad . \tag{A-1}$$

The characterisic function of x is then

$$f_x(\xi) = \int_{-\infty}^{\infty} du \, \exp(i\xi u) \, p_x(u) = \sum_{n=0}^{N} a_n \, \exp(i\xi n) \quad ; \tag{A-2}$$

this function has period 2π in ξ.

Now let integer M be selected such that

$$M \geq N+1 \quad . \tag{A-3}$$

Then consider M samples of characteristic function f_x at increment $2\pi/M$; that is, consider the set of samples

$$f_x(m2\pi/M) \quad \text{for } 0 \leq m \leq M-1 \quad . \tag{A-4}$$

Now let us take an M-point DFT of these samples, and scale by 1/M; that is, for $0 \leq k \leq M-1$,

$$\frac{1}{M} \sum_{m=0}^{M-1} \exp(-i2\pi mk/M) f_x(m2\pi/M)$$

$$= \frac{1}{M} \sum_{m=0}^{M-1} \exp(-i2\pi mk/M) \sum_{n=0}^{N} a_n \exp(i2\pi nm/M)$$

$$= \sum_{n=0}^{N} a_n \frac{1}{M} \sum_{m=0}^{M-1} \exp(i2\pi(n-k)m/M)$$

$$= \sum_{n=0}^{N} a_n \delta_{n-k}^{(M)} = \left\{ \begin{array}{ll} a_k & \text{for } 0 \le k \le N \\[2mm] 0 & \text{for } N < k \le M-1 \end{array} \right\} . \tag{A-5}$$

Here, in the second line, we used expression (A-2) for the characteristic function, and in the last line, we used (A-3).

Expression (A-5) states that the scaled M-point DFT of the set of characteristic function samples, (A-4), yields precisely the areas $\{a_k\}_0^N$ of the impulses in probability density function p_x, provided that $M \ge N+1$. The values returned by the DFT for $a_{N+1}, \ldots, a_{M-1}$ should all be zero.

More generally, if random variable x is limited to the values

$$h_0, \ h_0 + \Delta h, \ \ldots, \ h_0 + N \ \Delta h, \tag{A-6}$$

the characteristic function takes the form

$$f_x(\xi) = \exp(i\xi h_0) \sum_{n=0}^{N} a_n \exp(i\xi n \ \Delta h) \quad . \tag{A-7}$$

In this case, the sample set that must be subjected to an M-point DFT is not (A-4), but rather

$$f_x\left(\frac{2\pi m}{M\Delta h}\right) \exp\left(-i \ \frac{2\pi m}{M\Delta h} \ n_0\right) \quad \text{for } 0 \le m \le M-1 \quad . \tag{A-8}$$

When scaled by 1/M, this DFT yields areas $\{a_n\}$ directly.

Appendix B

ALTERNATIVE INPUT STATISTICS

Here we give the equations for two other typical random processes that could serve as candidates for inputs to the nonlinear system of interest. No numerical results have been evaluated for either of the following.

CHI-SQUARED RANDOM VARIABLES

This case could correspond to Gaussian noise with additive Gaussian signal, after passage through a square-law device and summation. That is, under H_1, consider the signal-bearing channel input to be the sum

$$x(m) = \sum_{k=1}^{2D} \left[s_k(m) + n_k(m) \right]^2 \quad , \tag{B-1}$$

where $\{s_k(m)\}$ and $\{n_k(m)\}$ are independent zero-mean Gaussian random variables with variances σ_s^2 and σ_n^2 for the signal and noise, respectively. x is a Chi-squared variate with 2D degrees of freedom.

Letting $\sigma_1^2 = \sigma_s^2 + \sigma_n^2$, the probability density function of x is given by

$$p_x^{(1)}(u) = \frac{u^{D-1}}{(D-1)! \left(2\sigma_1^2 \right)^D} \exp\left(-\frac{u}{2\sigma_1^2} \right) \quad \text{for } u > 0 \quad . \tag{B-2}$$

The cumulative distribution function of x is, via repeated integration by parts, given by

$$P_x^{(1)}(u) = 1 - \exp\left(-\frac{u}{2\sigma_1^2} \right) e_{D-1}\left(\frac{u}{2\sigma_1^2} \right) \quad \text{for } u > 0 \quad , \tag{B-3}$$

where (ref. 8, 6.5.11)

$$e_{D-1}(t) = \sum_{n=0}^{D-1} \frac{1}{n!} t^n \tag{B-4}$$

is the first D terms of $\exp(t)$.

The cumulative distribution function of x under H_0 follows immediately from (B-3) by setting $\sigma_s^2 = 0$ and identifying $\sigma_0^2 = \sigma_n^2$:

$$p_x^{(0)}(u) = 1 - \exp\left(-\frac{u}{2\sigma_0^2}\right) e_{D-1}\left(\frac{u}{2\sigma_0^2}\right) \quad \text{for } u > 0 \quad . \tag{B-5}$$

In fact, (B-3) and (B-5) could be used as input statistics for the analysis contained here in the main body of the report, without the need for interpretation (B-1); (B-1) merely lends the physical interpretation of x as the sum of a number of diversity inputs.

According to (5) and (6), we need the quantities

$$\left.\begin{aligned}
p_x^{(0)}\left(b_\ell\right) &= 1 - \exp\left(-v_\ell\right) e_{D-1}\left(v_\ell\right) \\[2mm]
p_x^{(1)}\left(b_\ell\right) &= 1 - \exp\left(-\frac{\sigma_0^2}{\sigma_1^2} v_\ell\right) e_{D-1}\left(\frac{\sigma_0^2}{\sigma_1^2} v_\ell\right)
\end{aligned}\right\} \text{for } 1 \le \ell \le L \quad , \tag{B-6}$$

where now the normalized breakpoints are defined as

$$v_\ell = \frac{b_\ell}{2\sigma_0^2} \quad \text{for } 1 \le \ell \le L \quad . \tag{B-7}$$

Of course, we should always select $b_\ell \ge 0$, since x is never negative.

NONCENTRAL CHI-SQUARED RANDOM VARIABLES

This case corresponds to Gaussian noise with additive deterministic signal, after passage through a squarer and summation. Under H_1, let the input be composed as follows:

$$x(m) = \sum_{k=1}^{20} \left[c_k(m) + n_k(m)\right]^2 \quad , \tag{B-8}$$

where $\{c_k(m)\}$ are constants, and noises $\{n_k(m)\}$ are independent zero-mean Gaussian random variables with variance σ_n^2. For example, x(m) corresponds to

the sum of D squared-envelopes of the output of a narrowband filter subjected to a sinewave and Gaussian noise. The probability density function of x in (B-8) is

$$p_x^{(1)}(u) = \frac{1}{2\sigma_n^2}\left(\frac{\sqrt{u}}{a\sigma_n}\right)^{D-1} \exp\left(-\frac{a^2}{2} - \frac{u}{2\sigma_n^2}\right) I_{D-1}\left(\frac{a\sqrt{u}}{\sigma_n}\right) \quad \text{for } u > 0, \quad \text{(B-9)}$$

where

$$a \equiv \frac{1}{\sigma_n}\left[\sum_{k=1}^{2D} c_k^2(m)\right]^{1/2} \quad \text{(B-10)}$$

is a measure of the total signal-to-noise ratio of random variable $x(m)$. (a is generally a function of m.) The cumulative distribution function is

$$P_x^{(1)}(u) = 1 - Q_D\left(a, \frac{\sqrt{u}}{\sigma_n}\right) \quad \text{for } u > 0 \quad , \quad \text{(B-11)}$$

where the Q-function is (ref. 9)

$$Q_D(a,b) \equiv \int_b^\infty dt \, t\left(\frac{t}{a}\right)^{D-1} \exp\left(-\frac{t^2 + a^2}{2}\right) I_{D-1}(at) \quad . \quad \text{(B-12)}$$

The probability density function and cumulative distribution function under H_0 follow from (B-9) and (B-11) by setting $a=0$, where we presume that $\{c_k(m)\}$ represent the signal components in (B-8); there results

$$p_x^{(0)}(u) = \frac{u^{D-1}}{(D-1)! \left(2\sigma_n^2\right)^D} \exp\left(-\frac{u}{2\sigma_n^2}\right) \quad \text{for } u > 0 \quad \text{(B-13)}$$

and

$$P_x^{(0)}(u) = 1 - \exp\left(-\frac{u}{2\sigma_n^2}\right) e_{D-1}\left(\frac{u}{2\sigma_n^2}\right) \quad \text{for } u > 0 \quad . \quad \text{(B-14)}$$

According to (5) and (6), we need

$$\left.\begin{array}{l} P_x^{(0)}\left(b_\ell\right) = 1 - \exp\left(-v_\ell\right) e_{D-1}\left(v_\ell\right) \\[2mm] P_x^{(1)}\left(b_\ell\right) = 1 - Q_D\left(a, \sqrt{2v_\ell}\right) \end{array}\right\} \quad \text{for } 1 \le \ell \le L \quad , \quad \text{(B-15)}$$

where normalized breakpoints

$$v_\ell \equiv \frac{b_\ell}{2\sigma_n^2} \quad \text{for } 1 \le \ell \le L \quad . \quad \text{(B-16)}$$

Again $b_\ell \ge 0$ since x in (B-8) can never be negative.

Appendix C

PROGRAM FOR DETECTION AND FALSE ALARM PROBABILITIES

```
10    ! DETECTION CHARACTERISTICS FOR QUANTIZERS, GREATEST-OF, AND ACCUMULATOR
20    ! QUANTIZER OUTPUT (ORDINATE) LEVELS ARE SET AT 0,1,...,L;  L+1 LEVELS
30    ! QUANTIZER ABSCISSA BREAKPOINTS ARE ARBITRARY; L BREAKPOINTS
40      M=5           ! NUMBER OF TIME SAMPLES ACCUMULATED: M>=1
50      N=3           ! NUMBER OF INPUT CHANNELS SUBJECT TO OR-ING: N>=1
60      L=7           ! NUMBER OF NON-ZERO QUANTIZER OUTPUT LEVELS: L>=1
70    ! NORMALIZED ABSCISSA BREAKPOINTS OF QUANTIZER (L NUMBERS):
80      DATA 1.4,1.85,2.3,2.75,3.2,3.65,4.1
90      Dimax=4.5   ! MAXIMUM VALUE OF Di OF INTEREST
100     Distep=.5   ! INCREMENTS IN Di OF INTEREST
110     Np=6        ! SMALLEST PROBABILITY OF INTEREST IS 10^(-Np)
120     DIM X(1:1024),Y(1:1024),Axis(1:19)
130     DIM V(1:10),S(1:10),Pw(1:10),Pfa(1:1024)
140     REDIM V(1:L),S(1:L),Pw(1:L),Pfa(1:M*L)
150     READ V(*)
160     PRINTER IS 0
170     PRINT "NUMBER OF TIME SAMPLES ACCUMULATED:         M =";M
180     PRINT "NUMBER OF INPUT CHANNELS SUBJECT TO OR-ING: N =";N
190     PRINT "NUMBER OF NON-ZERO QUANTIZER OUTPUT LEVELS: L =";L
200     PRINT "   QUANTIZER OUTPUT (ORDINATE) LEVELS ARE SET AT 0,1,...,L."
210     PRINT "   QUANTIZER NORMALIZED (ABSCISSA) BREAKPOINTS ARE AT:"
220     FOR J=1 TO L
230     PRINT V(J);
240     NEXT J
250     PRINT
260     PRINT "INPUT VOLTAGE-SNR Di VARIES FROM 0 TO";Dimax;"IN STEPS OF";Distep
270     PRINT
280     PRINT "THE GRAPH BELOW GIVES THE (FALSE ALARM) PROBABILITY THAT THE"
290     PRINT "SYSTEM OUTPUT IS GREATER THAN OR EQUAL TO J, FOR J = 1 TO ML."
300     PRINT "THE INPUT SIGNAL-TO-NOISE RATIO IS ZERO FOR THIS GRAPH."
310     M0=M*L
320     M1=M0+1
330     M2=M/2
340     N1=N-1
350     FOR J=2 TO 10
360     Nf=2^J
370     IF Nf>M0 THEN 420
380     NEXT J
390     PRINT
400     PRINT "ARRAYS  X,Y,Pfa  ARE NOT DIMENSIONED LARGE ENOUGH"
410     STOP
420     REDIM X(1:Nf),Y(1:Nf)
430     FOR J=1 TO L
440     S(J)=FNPh1(V(J))^N1
450     NEXT J
460     FOR Di=0 TO Dimax STEP Distep
470     FOR J=1 TO L
480     Pw(J)=S(J)+FNPh1(V(J)-Di)            !      EQUATIONS 5 AND 21
490     NEXT J
500     MAT X=ZER
```

```
510     MAT Y=ZER
520     X(1)=Pw(1)            !       EQUATION 6;  Alpha(I) in X(J+1)
530     FOR J=2 TO L
540     X(J)=Pw(J)-Pw(J-1)
550     NEXT J
560     X(L+1)=1-Pw(L)
570     CALL Fft(Nf,X(*),Y(*))
580     FOR J=1 TO Nf                     !       (X-iY)^M
590     T=(X(J)^2+Y(J)^2)^M2
600     A=M*FNArg(X(J),Y(J))
610     X(J)=T*COS(A)
620     Y(J)=-T*SIN(A)
630     NEXT J
640     CALL Fft(Nf,X(*),Y(*))
650     FOR J=1 TO M1
660     X(J)=X(J)/Nf                      !       BETA OF FIGURE
670     NEXT J
680     IF Di>0 THEN 1220
690     PLOTTER IS "GRAPHICS"
700     GRAPHICS
710     Np1=Np+1
720     Np2=Np*2+1
730     Axis(Np1)=0
740     FOR J=1 TO Np
750     T=FNInvphi(.1^J)
760     Axis(Np1-J)=T
770     Axis(Np1+J)=-T
780     NEXT J
790     SCALE 0,M0,Axis(1),Axis(Np2)
800     FOR J=0 TO M0 STEP 5
810     MOVE J,Axis(Np2)
820     DRAW J,Axis(1)
830     NEXT J
840     FOR J=1 TO Np2
850     MOVE 0,Axis(J)
860     DRAW M0,Axis(J)
870     NEXT J
880     PENUP
890     T=0
900     FOR J=M1 TO 2 STEP -1
910     T=T+X(J)
920     I=J-1
930     IF (T>1E-11) AND (T<1-1E-11) THEN 960
940     Pfa(I)=100
950     GOTO 1050
960     A=FNInvphi(T)
970     Pfa(I)=A
980     PLOT I,A
990  !  GOTO 1050      !    TO ELIMINATE CROSSES, INSERT THIS INSTRUCTION
1000    MOVE I,A-Axis(Np2)*.008
```

```
1010    DRAW I,A+Axis(Np2)*.008
1020    MOVE I-.004*M0,A
1030    DRAW I+.004*M0,A
1040    PLOT I,A
1050    NEXT J
1060    PENUP
1070    DUMP GRAPHICS
1080    PRINT LIN(1)
1090    PRINT "THE GRAPH BELOW IS A PLOT OF DETECTION PROBABILITY VERSUS"
1100    PRINT "FALSE ALARM PROBABILITY, FOR THE VALUES OF Di GIVEN ABOVE."
1110    GCLEAR
1120    SCALE Axis(1),Axis(Np2),Axis(1),Axis(Np2)
1130    FOR J=1 TO Np2
1140    MOVE Axis(1),Axis(J)
1150    DRAW Axis(Np2),Axis(J)
1160    NEXT J
1170    FOR J=1 TO Np2
1180    MOVE Axis(J),Axis(Np2)
1190    DRAW Axis(J),Axis(1)
1200    NEXT J
1210    PENUP
1220    T=0
1230    FOR J=M1 TO 2 STEP -1
1240    T=T+X(J)
1250    IF (T<1E-11) OR (T>1-1E-11) THEN 1360
1260    B=Pfa(J-1)
1270    IF ABS(B)>7 THEN 1360
1280    A=FNInvphi(T)
1290    PLOT B,A
1300    ! GOTO 1360      !   TO ELIMINATE CROSSES, INSERT THIS INSTRUCTION
1310    MOVE B,A-Axis(Np2)*.008
1320    DRAW B,A+Axis(Np2)*.008
1330    MOVE B-Axis(Np2)*.008,A
1340    DRAW B+Axis(Np2)*.008,A
1350    PLOT B,A
1360    NEXT J
1370    PENUP
1380    NEXT Di
1390    DUMP GRAPHICS
1400    PRINT LIN(6)
1410    PRINTER IS 16
1420    END
1430    !
1440    DEF FNArg(X,Y)               ! PRINCIPAL ARGUMENT  OF X+iY
1450    IF X=0 THEN A=.5*PI*SGN(Y)
1460    IF X<>0 THEN A=ATN(Y/X)
1470    IF X<0 THEN A=A+PI*(1-2*(Y<0))
1480    RETURN A
1490    FNEND
1500    !
```

```
1510    DEF FNPhi(X)                ! CUMULATIVE GAUSSIAN DISTRIBUTION
1520    IF ABS(X)>5.14 THEN 1780
1530    A=.282842712475*X
1540    C=COS(A)
1550    S=SIN(A)
1560    B=2*C
1570    A=B*C-1
1580    C=A*(1.2536751E-18+B*7.10005E-20+A+7.4517E-21)
1590    C=A*(1.533423425E-16+B*1.01649277E-17+C)
1600    C=A*(1.36760444757E-14+B*1.0601364636E-15+C)
1610    C=A*(8.89786526722E-13+B*8.06068838945E-14+C)
1620    C=A*(4.22616144318E-11+B*4.46968229249E-12+C)
1630    C=A*(1.46660614234E-9+B*1.80845587810E-10+C)
1640    C=A*(3.72252349369E-8+B*5.34275027603E-9+C)
1650    C=A*(6.91927520325E-7+B*1.15330990944E-7+C)
1660    C=A*(9.43281169838E-6+B*1.82066316364E-6+C)
1670    C=A*(9.44909268810E-5+B*2.10404583073E-5+C)
1680    C=A*(6.97183792408E-4+B*1.78228016255E-4+C)
1690    C=A*(3.80150767985E-3+B*1.10860645342E-3+C)
1700    C=A*(.0153985726157+B*.00507906961220+C)
1710    C=A*(.0467755234325+B*.0172439625887+C)
1720    C=A*(.108630245023+B*.0439773381941+C)
1730    C=A*(.201339747265+B*.0869894549959+C)
1740    C=A*(.330501521917+B*.144227226362+C)
1750    C=.703225002744+B*.247255168140+C
1760    Phi=.5+.0450158158079*X+.5*S*C
1770    GOTO 1920
1780    IF X>7 THEN 1910
1790    N=MAX(6,INT(69/ABS(X)),INT(525/X^2))+1
1800    A=1
1810    S=1
1820    B=1/X
1830    C=B*B
1840    FOR J=1 TO N
1850    A=(1-2*J)*A*C
1860    S=S+A
1870    NEXT J
1880    Phi=.398942280401*EXP(-.5*X*X)*ABS(B)*S
1890    IF X>0 THEN Phi=1-Phi
1900    GOTO 1920
1910    Phi=1
1920    RETURN Phi
1930    FNEND
1940    !
```

```
1950    DEF FNInvphi(Z)              ! INVERSE CUMULATIVE GAUSSIAN DISTRIBUTION
1960    X=2*Z-1
1970    DIM T(0:20),A(0:20)
1980    DATA .992885376619,.120467516143,.0169781993421,.00268670443716
1990    DATA .49963473024E-3,.988982186E-4,.2039181276E-4,.43272716E-5
2000    DATA .93808141E-6,.20673472E-6,.461597E-7,.1041668E-7,.23715E-8
2010    DATA .54393E-9,.12555E-9,.2914E-10,.679E-11,.159E-11,.37E-12
2020    DATA .912158803418,-.0162662818677,.43355647295E-3,.21443857007E-3
2030    DATA .262575108E-5,-.302109105E-5,-.1240606E-7,.6240661E-7,-.54012E-9
2040    DATA -.142321E-8,.3438E-10,.3358E-10,-.146E-11,-.81E-12,.5E-13,.2E-13
2050    DATA .956679709020,-.0231070043091,-.0043742350951,-.5765034226E-3
2060    DATA -.1096102231E-4,.2510854702E-4,.1056233607E-4,.275441233E-5
2070    DATA .43248450E-6,-.2053034E-7,-.4389154E-7,-.1768401E-7,-.399129E-8
2080    DATA -.18693E-9,.27292E-9,.13282E-9,.3183E-10,.167E-11,-.204E-11
2090    DATA -.965E-12,-.22E-12
2100    B=ABS(X)
2110    IF ABS(X)>=.8 THEN B=SQR(-LOG((1-X)*(1+X)))
2120    IF ABS(X)<.8 THEN 2220
2130    IF ABS(X)<.9975 THEN 2180
2140    Nmax=20
2150    RESTORE 2050
2160    Y=-.559457631330*B+2.28791571626
2170    GOTO 2250
2180    Nmax=15
2190    RESTORE 2020
2200    Y=-1.54881304237*B+2.56549012315
2210    GOTO 2250
2220    Nmax=18
2230    RESTORE 1980
2240    Y=X*X*3.125-1
2250    REDIM A(0:Nmax)
2260    READ A(*)
2270    Y2=Y*2
2280    T(0)=1
2290    T(1)=Y
2300    FOR N=2 TO Nmax
2310    T(N)=Y2*T(N-1)-T(N-2)
2320    NEXT N
2330    R=0
2340    FOR N=Nmax TO 0 STEP -1
2350    R=R+A(N)*T(N)
2360    NEXT N
2370    Invphi=SGN(X)*B+R*1.41421356237
2380    RETURN Invphi
2390    FNEND
2400    !
2410    SUB FFT(N,X(*),Y(*))    !    FFT SUBROUTINE HERE
```

Appendix D

ANALYSIS FOR GENERAL QUANTIZER

The characteristic function of decision variable z is still given by (8) in terms of f_y, the characteristic function of y. But now, from figures 5 and 2,

$$f_y(\xi) = \overline{\exp(i\xi y)} = \overline{\exp(i\xi q\{w\})} = \int_{-\infty}^{\infty} du\, p_w(u)\, \exp(i\xi q\{u\})$$

$$= \exp(i\xi h_o) \int_{-\infty}^{b_1} du\, p_w(u) + \ldots + \exp(i\xi h_L) \int_{b_L}^{\infty} du\, p_w(u)$$

$$= \exp(i\xi h_o)\, P_w(b_1) + \sum_{\ell=1}^{L-1} \exp(i\xi h_\ell) \left[P_w(b_{\ell+1}) - P_w(b_\ell) \right]$$

$$+ \exp(i\xi h_L) \left[1 - P_w(b_L) \right] \quad , \tag{D-1}$$

where P_w is the cumulative distribution function of random variable w, as given by (5). The inputs for this calculation of characteristic function f_z are M, N, L, $\{h_\ell\}_0^L$ and $\{b_\ell\}_1^L$. Again, input cumulative distribution functions $P_x^{(0)}$ and $P_x^{(1)}$ in (5) are arbitrary.

One problem with this quantizer is that P_D and P_F are stepwise functions of the decision threshold (which need not be integer now) at irregular points. A large DFT size would be necessary to track this behavior. However, operating characteristic plots of P_D vs. P_F would be smoother functions.

Appendix E

DERIVATION OF OPTIMUM PROCESSOR

We allow tne signal, if present, to jump randomly and independently between input channels on each time sample. Let

$$Y(m) = \left\{ x_n(m) \right\}_{n=1}^{N} \qquad \text{for } 1 \le m \le M \quad . \tag{E-1}$$

Under H_1, tne probability density function of N-vector $Y(m)$ is

$$p^{(1)}(Y(m)) = \frac{1}{N} \left[p^{(1)}\big(x_1(m)\big) \, p^{(0)}\big(x_2(m)\big) \, \cdots \, p^{(0)}\big(x_N(m)\big) \right.$$

$$\left. + p^{(0)}\big(x_1(m)\big) \, p^{(1)}\big(x_2(m)\big) \, \cdots \, p^{(0)}\big(x_N(m)\big) + \cdots \right]$$

$$= \prod_{n=1}^{N} \left\{ p^{(0)}\big(x_n(m)\big) \right\} \frac{1}{N} \sum_{n=1}^{N} \frac{p_1\big(x_n(m)\big)}{p_o\big(x_n(m)\big)} \qquad \text{for } 1 \le m \le M \quad . \tag{E-2}$$

Tnerefore the joint probability density function is

$$p^{(1)}(Y(1) \ldots Y(M)) = \prod_{m=1}^{M} \prod_{n=1}^{N} \left\{ p^{(0)}\big(x_n(m)\big) \right\} \prod_{m=1}^{M} \left\{ \frac{1}{N} \sum_{n=1}^{N} \frac{p^{(1)}\big(x_n(m)\big)}{p^{(0)}\big(x_n(m)\big)} \right\} \quad . \tag{E-3}$$

Tne likelinood ratio follows immediately as the last product in (E-3); tne log-likelihood ratio is therefore

$$\ell n \text{ likelihood ratio} = \sum_{m=1}^{M} \ell n \left(\frac{1}{N} \sum_{n=1}^{N} \frac{p^{(1)}\big(x_n(m)\big)}{p^{(0)}\big(x_n(m)\big)} \right) \quad . \tag{E-4}$$

Tnis result holds for arbitrary inputs with probability density functions $p(0)$ and $p(1)$.

When we employ tne Gaussian example in (16), (E-4) simplifies to

$$\sum_{m=1}^{M} \ell n \left(\frac{1}{N} \sum_{n=1}^{N} \exp\left(\frac{d_i}{\sigma} x_n(m) \right) \right) \quad \begin{smallmatrix} < \\ > \end{smallmatrix} \quad \text{threshold} \quad , \qquad (E-5)$$

where data-independent scale factors have been absorbed in the threshold. Exact analysis of (E-5) is conceivable, but is exceedingly tedious. Also d_i and σ must be known in order to realize (E-5).

For $d_i > 2$, it may be shown, to a good approximation, that

$$\sum_{n=1}^{N} \exp\left(\frac{d_i}{\sigma} x_n(m) \right) \cong \exp\left(\frac{d_i}{\sigma} \max_{n} \{ x_n(m) \} \right) \quad . \qquad (E-6)$$

Substitution in (E-5) yields the approximate likelihood ratio test

$$\sum_{m=1}^{M} \max_{n} \{ x_n(m) \} \quad \begin{smallmatrix} > \\ < \end{smallmatrix} \quad \text{threshold} \quad . \qquad (E-7)$$

Knowledge of d_i and σ is now not required. Processor (E-7) is just figure 1 or 5, with the quantizer replaced by a linear gain. This special case of (E-5) is very important, because decent performance can be obtained for $d_i > 2$, and this is exactly where (E-7) is virtually optimum. A good approximation to the performance of (E-7) is afforded by the results contained herein, if L is chosen large and the breakpoints are well-placed. Some approximate results via second-moment approaches are given in refs. 1, 10, 11, 12.

REFERENCES

1. A. H. Nuttall, "Signal-to-Noise Ratio Requirements for Greatest-Of Device Followed by Integrator," NUSC Technical Memorandum TC-13-75, 24 July 1975.

2. A. H. Nuttall, "Input Deflection Requirements for Quantizers Followed by Greatest-Of-Device and Integrator," NUSC Technical Memorandum TM No. 781174, 24 August 1978.

3. W. A. Struzinski, "OR-ing Loss Data for Square Law Detectors Followed by an OR-ing Device and an Accumulator," JASA, vol. 72, pp. 191-195, July 1982.

4. W. A. Struzinski, "OR-ing Loss for Quantizers Followed by an OR-ing Device and an Accumulator," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-30, no. 4, pp. 668-671, August 1982.

5. W. A. Struzinski, "Optimizing the Performance of a Quantizer," submitted to JASA, October 1982.

6. A. H. Nuttall, "Numerical Evaluation of Cumulative Probability Distribution Functions Directly from Characteristic Functions," Proc. IEEE, vol. 57, no. 11, pp. 2071-2072, November 1969; also NUSL Report No. 1032, 11 August 1969.

7. A. H. Nuttall, "Alternate forms for Numerical Evaluation of Cumulative Probability Distributions Directly from Characteristic Functions," Proc. IEEE, vol. 58, no. 11, pp. 1872-1873, November 1970; also NUSC Report No. NL-3012, 12 August 1970.

8. Handbook of Mathematical Functions, U. S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series 55, U. S. Government Printing Office, Washington DC, June 1964.

9. A. H. Nuttall, "Some Integrals Involving the Q_M-Function", NUSC Technical Report 4755, 15 May 1974.

10. "Loss of S/N Ratio in Peak Detecting N Channels of Data," ASW Technical Note, Hughes Aircraft Co., Fullerton, CA., 12 Feb. 1968.

11. J. J. Dow, B. M. Brown, N. A. Reeder, "A Performance Comparison of Three Narrowband Data Reduction Techniques," Tracor Report T70-AU-7382-C, Tracor Inc., Austin, TX, 16 July 1970.

12. D. F. Korff, "ORing," Raytheon Memo No. DFK:81/08, Raytheon Submarine Signal Division, Portsmouth, RI, 9 September 1980.

On Resonance Extraction And Waveform Fitting For Transient Data; Prony's Method

A. H. Nuttall

ABSTRACT

This report explains the basic philosophy and mathematics of waveform fitting with complex exponentials, when the available data points are more than the number needed for a perfect fit. The connection with Prony's method is developed, some recent new work by Auton and Van Blaricum is summarized, and an eigenvector generalization of linear prediction is presented. Effort is still continuing in this important field of resonance estimation and extraction, and answers to some important questions on sensitivity, sampling rate, and bias are still not available.

TABLE OF CONTENTS

LIST OF SYMBOLS

N   number of data points

g_m model sequence value at time m

n   number of complex exponentials

C_k strength of k-th resonance

μ_k location of k-th resonance

α_j linear prediction coefficient

f_m measured data value

$\hat{f}_m$ predicted data value

$\hat{e}_m$ prediction error

$\hat{E}$ total squared prediction error

$\hat{w}_m, \tilde{w}_m$ weights

$\tilde{f}_m$ model data value

$\tilde{e}_m$ data error

$\tilde{E}$ total squared data error

Q,F   data matrix

D_k strength of k-th resonance for double pole

β_j auxiliary linear predictive coefficient

c_j constraint coefficient

C   constraint vector

A   coefficient vector

d_m error

D   error matrix

λ Lagrange multiplier

A_o optimum coefficient vector

S   correlation matrix

LIST OF SYMBOLS (Cont'd)

λ_j eigenvalue

Λ eigenvalue matrix

E modal matrix

b_k eigenvector coefficients

λ_0 smallest eigenvalue of S

e_0 weakest eigenvector of S

ON RESONANCE EXTRACTION AND WAVEFORM FITTING
FOR TRANSIENT DATA; PRONY'S METHOD

INTRODUCTION

The estimation of the resonances (natural frequencies) of a system, from observation of a noisy response, is an important problem of frequent occurrence in practical situations. Usually, the number of observations is considerably greater than the number of resonances, and the task of utilizing these "extra" data to reduce the errors of estimation must be accomplished without an excessive amount of computational effort or trial-and-error. Accordingly, the original exact-fit procedure by Prony has to be generalized to a least-squares approach. In this manner, the amount of data processing is minimized, with all the nonlinear processing being concentrated in the solution for the roots of a polynomial.

The purpose of this report is to develop and explain this least-squares solution and to show its close connection to linear prediction. The first section, on Mathematical Details, sets up the problem definition and introduces the terms necessary to interpret recent work by Auton and Van Blaricum [1] described in the next section. Some important points about the waveform-fitting technique are explained, and some possible alternative approaches are mentioned. A more general model is considered in appendix A, and a generalization to linear prediction is developed in appendix B, which subsumes forward prediction, backward prediction, and a weighted linear combination in general.

MATHEMATICAL DETAILS

IDEAL EXPONENTIAL MODEL

Suppose a sequence $\{g_m\}_0^{N-1}$ of N points is given exactly by the model*

$$g_m = \sum_{k=1}^{n} C_k \exp(a_k m) \equiv \sum_{k=1}^{n} C_k \mu_k^m \quad \text{for} \quad 0 \leq m \leq N-1. \qquad (1)$$

That is, sequence $\{g_m\}_0^{N-1}$ is a sum of n complex exponentials. Without loss of generality, we presume that all the $\{C_k\}$ are nonzero for $1 \leq k \leq n$.

Consider the error (in linear prediction) of attempting to represent g_m in terms of its past n values; that is, for $n \leq m \leq N-1$, consider linear prediction error (where $\alpha_0 \equiv -1$)

$$g_m - \sum_{j=1}^{n} \alpha_j g_{m-j} = -\sum_{j=0}^{n} \alpha_j g_{m-j} = -\sum_{j=0}^{n} \alpha_j \sum_{k=1}^{n} C_k \mu_k^{m-j}$$

$$= \sum_{k=1}^{n} C_k \mu_k^{m-n} \sum_{j=0}^{n} (-\alpha_j \mu_k^{n-j}) = \sum_{k=1}^{n} C_k \mu_k^{m-n} [\mu_k^n - \alpha_1 \mu_k^{n-1} - \ldots - \alpha_{n-1} \mu_k - \alpha_n],$$

$$(2)$$

where we substituted (1) and interchanged summations. Now we choose the n linear coefficients $\{\alpha_j\}_1^n$ such that

$$\mu_k^n - \alpha_1 \mu_k^{n-1} - \ldots - \alpha_{n-1} \mu_k - \alpha_n = 0 \quad \text{for} \quad 1 \leq k \leq n. \qquad (3)$$

This requires solution of n linear equations for the n unknowns $\{\alpha_j\}_1^n$, presuming that the n quantities $\{\mu_k\}_1^n$ are known. In fact, the general solution is

$$\alpha_j = (-1)^{j-1} \{\text{sum of all possible products of j different } \mu\text{'s}\}$$

$$\text{for} \quad 1 \leq j \leq n; \qquad (4a)$$

---

*This can be generalized to include terms like $C\mu^m + Dm\mu^m$; see appendix A.

that is,

$$\alpha_1 = \mu_1 + \mu_2 + \ldots + \mu_n$$

$$\alpha_2 = -(\mu_1\mu_2 + \mu_1\mu_3 + \ldots + \mu_1\mu_n + \mu_2\mu_3 + \mu_2\mu_4 + \ldots + \mu_{n-1}\mu_n)$$

$$\vdots$$

$$\alpha_n = (-1)^{n-1} \mu_1\mu_2\cdots\mu_n. \tag{4b}$$

With this choice of $\{\alpha_j\}_1^n$, (2) and (3) yield

$$g_m - \sum_{j=1}^{n} \alpha_j \, g_{m-j} = 0 \quad \text{for} \quad n \le m \le N-1, \tag{5a}$$

or

$$g_m = \sum_{j=1}^{n} \alpha_j \, g_{m-j} \quad \text{for} \quad n \le m \le N-1, \tag{5b}$$

That is, when sequence $\{g_m\}_0^{N-1}$ is generated as a sum of n complex exponentials according to (1), the sequence value g_m can be determined exactly as a forward linear combination of the previous n values, provided that $n \le m \le N-1$. The restriction of m to this range is due to the fact that g_m is presumed unknown for $m < 0$ and for $m > N-1$; thus only the "valid," or available, data are employed in (2) and (5b).

It is important to observe that the n linear predictive coefficients $\{\alpha_j\}_1^n$ in (4b) depend on $\{\mu_k\}_1^n$ but are completely independent of the values of the exponential strengths, or "residues," $\{C_k\}_1^n$ in (1). Also, if the $\{\alpha_j\}_1^n$ were known instead of the $\{\mu_k\}_1^n$, then (3) can be solved for the $\{\mu_k\}_1^n$ as the n roots of an n-th order polynomial.

A more general approach to linear prediction is developed in appendix B. It subsumes the forward prediction (given above), backward prediction, and a weighted linear combination in general.

ACTUAL MEASURED DATA

Now suppose that some arbitrary data sequence $\{f_m\}_0^{N-1}$ has been measured or is available, and we want to choose the $2n$ parameters in the exponential model (1) such that the error of representing data $\{f_m\}_0^{N-1}$ by this model is minimized in some sense. Guided by (5b), we first let linearly predicted value

$$\hat{f}_m \equiv \sum_{j=1}^{n} \alpha_j \, f_{m-j} \quad \text{for} \quad n \le m \le N-1, \tag{6}$$

where the linear coefficients $\{\alpha_j\}_1^n$ are to be selected. In particular, we define the prediction error sequence (called the equation error in [1])

$$\hat{e}_m \equiv f_m - \hat{f}_m = f_m - \sum_{j=1}^{n} \alpha_j \, f_{m-j} \quad \text{for} \quad n \le m \le N-1. \tag{7}$$

This is also called Prony's difference equation. We then define the total squared prediction error as*

$$\hat{E} \equiv \sum_{m=n}^{N-1} \hat{w}_m \, \hat{e}_m^2 = \sum_{m=n}^{N-1} \hat{w}_m \left(f_m - \sum_{j=1}^{n} \alpha_j \, f_{m-j} \right)^2, \tag{8}$$

where $\{\hat{w}_m\}_n^{N-1}$ are a set of $N-n$ positive weights. $\hat{E}$ is called the quadratic error in [1].

Minimization of total squared prediction error $\hat{E}$ by choice of coefficients $\{\alpha_j\}_1^n$ is accomplished by setting

$$\frac{\partial \hat{E}}{\partial \alpha_k} = 0 \quad \text{for} \quad 1 \le k \le n. \tag{9}$$

This results in n linear equations in the n unknowns $\{\alpha_k\}_1^n$. We solve these equations for the $\{\alpha_k\}_1^n$ that minimize prediction error $\hat{E}$.

We must point out an alternative approach to the minimization of $\hat{E}$. One could instead minimize the Chebyshev error; that is, we could choose the $\{\alpha_j\}_1^n$ in (7) so as to minimize the quantity

$$\max_{n \le m \le N-1} \left| f_m - \sum_{j=1}^{n} \alpha_j \, f_{m-j} \right|. \tag{10}$$

That is, the maximum error in prediction is minimized. Although this approach yields nonlinear equations in the $\{\alpha_j\}_1^n$, efficient linear programming techniques exist for this problem. How well this minimax error criterion compares with the total squared error criterion is not known.

_____

*We are presuming real data sequences here; generalization to complex data is possible.

Given the values for $\{\alpha_k\}_1^n$, whether obtained via (9) or (10), we can now solve (3) for the $\{\mu_k\}_1^n$. Some of these latter values may be complex, even though all the $\{\alpha_k\}_1^n$ are real for real data $\{f_m\}_0^{N-1}$; this situation is treated in [2], p. 380.

Guided now by (1), we next let model data value*

$$\tilde{f}_m \equiv \sum_{k=1}^n C_k \, \mu_k^m \quad \text{for} \quad 0 \le m \le N-1. \tag{11}$$

Then we define data error sequence (called the true error in [1])

$$\tilde{e}_m \equiv f_m - \tilde{f}_m = f_m - \sum_{k=1}^n C_k \, \mu_k^m \quad \text{for} \quad 0 \le m \le N-1. \tag{12}$$

In a similar fashion to (8), we also define the total squared data error as

$$\tilde{E} \equiv \sum_{m=0}^{N-1} \tilde{w}_m \, \tilde{e}_m^2 = \sum_{m=0}^{N-1} \tilde{w}_m \left(f_m - \sum_{k=1}^n C_k \, \mu_k^m \right)^2 , \tag{13}$$

where $\{\tilde{w}_m\}_0^{N-1}$ are a set of N positive weights. To minimize total error $\tilde{E}$, we set

$$\frac{\partial \tilde{E}}{\partial C_j} = 0 \quad \text{for} \quad 1 \le j \le n, \tag{14}$$

thereby obtaining n linear equations in the n unknowns $\{C_j\}_1^n$. (The quantities $\{\mu_k\}_1^n$ are already known at this point; see the discussion preceding (11)). We solve these n equations for the $\{C_j\}_1^n$ that minimize $\tilde{E}$.

An alternative approach to the minimization of $\tilde{E}$ is to minimize the Chebyshev error; that is, choose the $\{C_k\}_1^n$ in (12) so as to minimize the quantity

$$\max_{0 \le m \le N-1} \left| f_m - \sum_{k=1}^n C_k \, \mu_k^m \right| . \tag{15}$$

_____

*This presumes that all the roots $\{\mu_k\}_1^n$ are distinct; if on the other hand, we had, for example, $\mu_1 = \mu_2$, then we need $C_1 \mu_1^m + C_2 m \mu_1^m$ rather than $C_1 \mu_1^m + C_2 \mu_2^m$.

5

Again, the performance quality of (10) and (15) is not known.

At this point, we have a "fitted" waveform,

$$\sum_{k=1}^{n} C_k \, \mu_k^m \quad \text{for} \quad 0 \leq m \leq N-1, \tag{16}$$

to the original given data sequence $\{f_m\}_0^{N-1}$. However, it should be observed that the fit was obtained via a two-stage <u>sequential</u> procedure. Namely, we first minimized total prediction error $\hat{E}$ to find the linear coefficients $\{\alpha_k\}_1^n$, and from them, solved the polynomial of (3) for its roots $\{\mu_k\}_1^n$. (These latter quantities are called the resonances in [1]). Then, with these known values for $\{\mu_k\}_1^n$, total data error $\tilde{E}$ was minimized, thereby determining the strengths (residues) $\{C_k\}_1^n$ of each of the known exponential components $\{\mu_k^m\}_{k=1}^n$.

Both error definitions, (7)-(8) and (12)-(13), utilize and "fit" the available data sequence $\{f_m\}_0^{N-1}$, but in two different senses, the first via linear prediction, and the second via an exponential model. The worst non-linear data processing encountered in this two-stage procedure is the solution of an n-th order polynomial, (3), for all its roots $\{\mu_k\}_1^n$. This sequential procedure will not realize as small an error as direct minimization of

$$\sum_{m=0}^{N-1} w_m \left(f_m - \sum_{k=1}^{n} C_k \, \mu_k^m \right)^2 \tag{17}$$

via <u>simultaneous</u> choice of $\{C_k\}_1^n$ and $\{\mu_k\}_1^n$. However, this latter approach is highly nonlinear in the $\{\mu_k\}_1^n$, and no direct (nonrecursive) solution is known. Of course, a gradient search on (17) could be employed, using as starting values, those obtained above via the two-stage sequential procedure.

SOME RECENT WORK

The source of the following results and comments is the work by Auton and Van Blaricum [1]. The solution for the coefficients $\{\alpha_j\}_1^n$ in (9) is called the reduced or inhomogeneous solution; see [1], vol. I, p. 2-5. This traditional solution, unfortunately, tends to zero as the white (independent) noise component in $\{f_m\}_0^{N-1}$ gets larger. A remedy to this undesired behavior is furnished by employing instead, the weakest eigenvector of the matrix Q^TQ, where Q is the data matrix formed by arranging the given data $\{f_m\}_0^{N-1}$ in columns in a particular fashion; see [1], vol. I, p. 2-2. (An equivalent interpretation is that Q^TQ or Q are approximated by matrices of lower rank, i.e., singular matrices.) It has been found that the weakest eigenvector of Q^TQ is less dependent on the absolute noise level and can furnish more useful values for the resonances $\{\mu_k\}_1^n$ than can the inhomogeneous solution. Physically, the "best" linear prediction of a noisy waveform tends to zero, whereas an eigenvector can maintain all its components nonzero, regardless of the absolute noise level. At present, the weakest eigenvector solution is judged to be the best of all iterative and noniterative methods for estimating the resonances $\{\mu_k\}_1^n$; see [1], vol. I, p. 2-28.

When the number of resonances, n, in (1) is unknown, its determination or estimation must be made from the available data $\{f_m\}_0^{N-1}$. If k is the true (unknown) number of resonances, and n is the hypothesized number, there are $n-k$ extraneous resonance estimates produced. A maximum likelihood procedure developed in [1] and applied to the ℓ smallest eigenvalues (for various values of ℓ) has been found to give reasonable estimates of k. An alternative approach, employing time reversal of the data sequence, seems to separate extraneous resonances, but more study is suggested; see [1], vol. I, p. 3-26.

CONCLUSIONS

The usual problems associated with Prony's method, regarding sensitivity to noise, have been attributed to dense sampling and bias. If both of these problems are treated properly and the weakest eigenvector is employed, Prony's method produces excellent estimates of the resonances, even from data with high noise levels; see [1], vol. I, p. 4-8.

Studies on some of these still-unanswered questions about alternative procedures for order selection and resonance estimation will continue. Certainly, further improvements in the procedures and performance will ensue. Applications to real measured data have yet to be made, however; see [1], vol. I, pp. 5-2 and 5-3.

Appendix A

A MORE GENERAL MODEL

Instead of (1) of the main text, suppose that sequence value

$$g_m = \sum_{k=1}^{n} C_k \, \mu_k^m + \sum_{k=1}^{p} D_k \, m\mu_k^m \quad \text{for} \quad 0 \leq m \leq N-1, \tag{A.1}$$

where p can be larger or smaller than n. Then for $n+p \leq m \leq N-1$, consider linear prediction error

$$g_m - \sum_{j=1}^{n} \alpha_j \, g_{m-j} - \sum_{j=1}^{p} \beta_j \, g_{m-n-j}$$

$$= -\sum_{j=0}^{n} \alpha_j \left[\sum_{k=1}^{n} C_k \, \mu_k^{m-j} + \sum_{k=1}^{p} D_k (m-j)\mu_k^{m-j} \right] \qquad (\alpha_o \equiv -1)$$

$$- \sum_{j=1}^{p} \beta_j \left[\sum_{k=1}^{n} C_k \, \mu_k^{m-n-j} + \sum_{k=1}^{p} D_k (m-n-j)\mu_k^{m-n-j} \right]$$

$$= -\sum_{k=1}^{n} C_k \, \mu_k^m \left[\sum_{j=0}^{n} \alpha_j \, \mu_k^{-j} + \sum_{j=1}^{p} \beta_j \, \mu_k^{-n-j} \right]$$

$$- \sum_{k=1}^{p} D_k \, \mu_k^m \left[\sum_{j=0}^{n} \alpha_j (m-j)\mu_k^{-j} + \sum_{j=1}^{p} \beta_j (m-n-j)\mu_k^{-n-j} \right]. \tag{A.2}$$

The quantities in brackets can be made zero for $n+p \leq m \leq N-1$, by setting both

$$\sum_{j=0}^{n} \alpha_j \, \mu_k^{-j} + \sum_{j=1}^{p} \beta_j \, \mu_k^{-n-j} = 0 \quad \text{for} \quad 1 \leq k \leq n \tag{A.3}$$

and

$$\sum_{j=0}^{n} \alpha_j (m-j)\mu_k^{-j} + \sum_{j=1}^{p} \beta_j (m-n-j)\mu_k^{-n-j} = 0 \quad \text{for} \quad 1 \leq k \leq p. \tag{A.4}$$

This combination constitutes $n+p$ linear equations in the $n+p$ unknowns $\{\alpha_j\}_1^n$ and $\{\beta_j\}_1^p$; $\alpha_0 = -1$. These equations can be put in the form

$$\alpha_0 \mu_k^{n+p} + \alpha_1 \mu_k^{n+p-1} + \ldots + \alpha_n \mu_k^p + \beta_1 \mu_k^{p-1} + \ldots + \beta_p = 0 \quad \text{for} \quad 1 \leq k \leq n, \tag{A.5}$$

$$\alpha_1 \mu_k^{n+p-1} + \ldots + \alpha_n n \mu_k^p + \beta_1(n+1)\mu_k^{p-1} + \ldots + \beta_p(n+p) = 0 \quad \text{for} \quad 1 \leq k \leq p. \tag{A.6}$$

So sequence value g_m can be determined exactly as a linear combination of its previous $n+p$ values, for $n+p \leq m \leq N-1$. Notice that coefficients $\{\alpha_j\}_1^n$ and $\{\beta_j\}_1^p$ depend on $\{\mu_k\}_1^q$ (where $q = \max(n,p)$), but not on strengths $\{C_k\}_1^n$ or $\{D_k\}_1^p$. See also [3], pp. 174-175.

Appendix B

EIGENVECTOR GENERALIZATION OF LINEAR PREDICTION

IDEAL MODEL

The starting point is again (1) of the main text. We now generalize (2) of the main text to the form

$$e_m \equiv \sum_{j=0}^{n} \alpha_j \, g_{m-j} \quad \text{for} \quad n \le m \le N-1, \tag{B.1}$$

where all the $\{\alpha_j\}_0^n$ are arbitrary for the moment. It follows, from substitution of (1) of the main text in (B-1), that

$$e_m = \sum_{j=0}^{n} \alpha_j \sum_{k=1}^{n} C_k \, \mu_k^{m-j} = \sum_{k=1}^{n} C_k \, \mu_k^{m} \sum_{j=0}^{n} \alpha_j \, \mu_k^{-j}$$

$$= \sum_{k=1}^{n} C_k \, \mu_k^{m-n} \sum_{j=0}^{n} \alpha_j \, \mu_k^{n-j} \quad \text{for} \quad n \le m \le N-1. \tag{B.2}$$

Now let us set

$$\sum_{j=0}^{n} \alpha_j \, \mu_k^{n-j} = \alpha_0 \, \mu_k^{n} + \ldots + \alpha_{n-1} \, \mu_k + \alpha_n = 0 \quad \text{for} \quad 1 \le k \le n, \tag{B.3}$$

by choice of $\{\alpha_j\}_0^n$. Since there are only n equations in (B.3), but n+1 unknowns, we will not get a unique solution for the $\{\alpha_j\}_0^n$ unless we restrict them somehow. Also, we must disallow the zero solution.

Observe that if we had used only n coefficients $\{\alpha_j\}_0^{n-1}$ in (B.1), we would have obtained, instead of (B.3), n equations in n unknowns. However, the only solution to these equations is the zero solution $\alpha_j = 0$ for all j, which is useless.

Before we consider the restriction on $\{\alpha_j\}_0^n$, observe that substituting (B.3) in (B.2) yields

$$e_m = \sum_{j=0}^{n} \alpha_j \, g_{m-j} = 0 \quad \text{for} \quad n \leq m \leq N-1. \tag{B.4}$$

That is, we can find an infinite number of linear combinations of n+1 adjacent values of sequence $\{g_m\}_0^{N-1}$ generated via (1) of the main text, which are identically zero for all possible locations of the (n+1)-long average within the record of length N.

Now to get back to the solution of (B.3) for the coefficients $\{\alpha_j\}_0^n$, we observe that the linear predictive approach considered in (2) et seq. of the main text amounts to choosing $\alpha_0 = -1$; this results in a unique solution for the n linear equations (B.3) in the remaining n unknowns $\{\alpha_j\}_1^n$, and is called forward prediction by virtue of form (5b) of the main text. An obvious alternative would be to select $\alpha_n = -1$, in which case (B.3) and (B.4) would yield a unique solution for $\{\alpha_j\}_0^{n-1}$, and

$$g_{m-n} = \alpha_0 g_m + \ldots + \alpha_{n-1} \, g_{m-n+1} \quad \text{for} \quad n \leq m \leq N-1. \tag{B.5}$$

That is, we are doing backward linear prediction to obtain the sequence values. But observe that both of these cases are specializations of the linear constraint

$$C^T A = 1 \tag{B.6}$$

on the coefficients $\{\alpha_j\}_0^n$, where

$$C = \begin{bmatrix} c_0 \\ c_1 \\ . \\ . \\ . \\ c_n \end{bmatrix} , \quad A = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ . \\ . \\ . \\ \alpha_n \end{bmatrix} \tag{B.7}$$

are column matrices. Constraint (B.6) prevents the zero solution, and when combined with (B.3), gives a unique solution for A. We can normalize the matrix of constants, C, such that

$$C^T C = 1 \qquad (\text{or } K \text{ if desired}), \tag{B.8}$$

without loss of generality. Forward or backward prediction, respectively, corresponds to choosing all the $\{c_j\}_0^n$ equal to zero except for edge elements c_0 or c_n, respectively, equal to -1. So, generally, we can realize the linear combination.

$$\sum_{j=0}^{n} \alpha_j \, g_{m-j} = 0 \quad \text{for} \quad n \leq m \leq N-1, \tag{B.9}$$

subject to $\{\alpha_j\}_0^n$ satisfying the linear constraint (B.6), which guarantees a nonzero solution. C is <u>any</u> vector satisfying (B.8).

ACTUAL MEASURED DATA

Now consider that measured data $\{f_m\}_0^{N-1}$ are available. Instead of linear prediction (6) of the main text, consider the more general linear combination (as in (B.1))

$$d_m = \sum_{j=0}^{n} \alpha_j \, f_{m-j} \quad \text{for} \quad n \leq m \leq N-1, \tag{B.10}$$

where set $\{\alpha_j\}_0^n$ is not yet specified. Define error and data matrices

$$D = \begin{bmatrix} d_n \\ d_{n+1} \\ . \\ . \\ . \\ d_{N-1} \end{bmatrix}, \quad F = \begin{bmatrix} f_n & f_{n-1} & \cdots & f_0 \\ f_{n+1} & \cdots & & f_1 \\ . & & & . \\ . & & & . \\ . & & & . \\ f_{N-1} & \cdots & & f_{N-1-n} \end{bmatrix} \quad (N-n) \times (n+1). \tag{B.11}$$

Then (B.10) can be expressed as

$$D = FA \tag{B.12}$$

where we used (B.7).

Now we want to minimize the total quadratic error of (B.10), namely,

$$\sum_{m=n}^{N-1} d_m^2 = D^T D = A^T F^T F A \tag{B.13}$$

by selection of A, but subject to linear constraint (B.6) on A, which guarantees a nonzero solution. C is an arbitrary, yet-unspecified matrix. Accordingly, we use a Lagrange multiplier 2λ and look for an extremum of

$$A^T S \, A - 2\lambda \, C^T A, \tag{B.14}$$

where we have defined

$$S = F^T F \qquad (n+1) \times (n+1) \text{ matrix.} \tag{B.15}$$

S is easily seen to be a nonnegative definite matrix; it generally has full rank when $N > 2n$. Completing the square in (B.14), we rewrite it as

$$(A - \lambda S^{-1} C)^T S(A - \lambda S^{-1} C) - \lambda^2 C^T S^{-1} C. \tag{B.16}$$

The extremum is then obviously realized for coefficient matrix

$$A_0 = \lambda S^{-1} C. \tag{B.17}$$

To evaluate λ, we have to satisfy the linear constraint (B.6):

$$\lambda C^T S^{-1} C = 1, \quad \lambda = \frac{1}{C^T S^{-1} C} . \tag{B.18}$$

The best coefficient set is then, from (B.17),

$$A_0 = \frac{S^{-1} C}{C^T S^{-1} C} . \tag{B.19}$$

(Thus the best coefficients are proportional to the first column of S^{-1} for forward linear prediction, or to the last column for backward linear prediction.) The corresponding minimum value of the total quadratic error (B.13), is

$$A_0^T S A_0 = \frac{C^T S^{-1} S S^{-1} C}{(C^T S^{-1} C)^2} = \frac{1}{C^T S^{-1} C} . \tag{B.20}$$

(This denominator reduces to the 0,0 element of S^{-1} for forward linear prediction, or to the n,n element of S^{-1} for backward linear prediction.)

But this result, (B.20), obviously depends on the particular values assigned to the constraint vector C in (B.6). The question then arises as to what constraint vector would yield further reduction of error (B.20). To determine this, let matrix S, defined in (B.15), have eigenvalue matrix

$$\Lambda = \begin{bmatrix} \lambda_0 & & & \\ & \lambda_1 & & 0 \\ & & \cdot & \\ & & & \cdot \\ & 0 & & & \cdot \\ & & & & \lambda_n \end{bmatrix} , \quad \lambda_0 < \lambda_1 < \ldots < \lambda_n , \tag{B.21}$$

and modal (eigenvector) matrix

$$E = \begin{bmatrix} | & | & & | \\ e_0 & e_1 & \cdots & e_n \\ | & | & & | \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}. \tag{B.22}$$

Then

$$SE = E\Lambda \tag{B.23}$$

or

$$Se_k = \lambda_k e_k \qquad \text{for} \quad 0 \leq k \leq n. \tag{B.24}$$

By taking the inverse of (B.23), and pre- and post-multiplying by E, we obtain

$$S^{-1} E = E\Lambda^{-1} \tag{B.25}$$

or

$$S^{-1} e_k = \lambda_k^{-1} e_k \quad \text{for} \quad 0 \leq k \leq n, \tag{B.26}$$

which we will need below. The inverse matrix has the same eigenvectors but the inverse eigenvalues of S.

Now any n+1 column matrix can be expressed in terms of the eigenvectors of S. In particular, suppose we let

$$C = \sum_{k=0}^{n} b_k e_k . \tag{B.27}$$

Recalling normalization (B.8), we have the constraint on the $\{b_k\}_0^n$:

$$\sum_{k,\ell=0}^{n} b_k b_\ell e_k^T e_\ell = \sum_{k=0}^{n} b_k^2 = 1 , \tag{B.28}$$

since the eigenvectors $\{e_k\}_0^n$ are orthonormal. If we substitute (B.27) in (B.20), the denominator is given by

$$C^T S^{-1} C = \sum_{k,\ell=0}^{n} b_k b_\ell e_k^T S^{-1} e_\ell = \sum_{k,\ell=0}^{n} b_k b_\ell e_k^T \lambda_\ell^{-1} e_\ell$$

$$= \sum_{k,\ell=0}^{n} b_k b_\ell \lambda_\ell^{-1} \delta_{k\ell} = \sum_{k=0}^{n} b_k^2/\lambda_k, \tag{B.29}$$

where we employed (B.26) and the orthonormality of the eigenvectors. Now since we want to minimize (B.20), we must maximize (B.29), but subject to (B.28). Obviously the best choice of $\{b_k\}_0^n$ is given by

$$b_0 = \pm 1, \qquad b_k = 0 \quad \text{for} \quad 1 \leq k \leq n, \tag{B.30}$$

where λ_0 is the smallest eigenvalue of S; see (B.21). Thus

$$\text{Minimum total quadratic error} = \min_C\left\{A_0^T S A_0\right\} = \lambda_0 , \tag{B.31}$$

which is the smallest eigenvalue of S defined in (B.15).

Now we can employ result (B.30) in (B.27) and (B.19) to find the best coefficient set A_0. We have $C = \pm e_0$, and (B.19) becomes

$$A_0 = \frac{\pm S^{-1} e_0}{e_0^T S^{-1} e_0} = \pm \frac{\lambda_0^{-1} e_0}{e_0^T \lambda_0^{-1} e_0} = \pm e_0 , \tag{B.32}$$

where we used (B.26). Thus both the constraint vector and the best linear weighting of the data in (B.10) are equal to the weakest eigenvector of the matrix $S = F^T F$, where F is the data matrix defined in (B.11).

We can now return to (B.3) to solve for the $\{\mu_k\}_1^n$, where we use the components of the weakest eigenvector of S for the $\{\alpha_j\}_0^n$; that is, we use

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{bmatrix} = \pm \begin{bmatrix} e_{00} \\ e_{01} \\ \cdot \\ \cdot \\ \cdot \\ e_{0n} \end{bmatrix} . \tag{B.33}$$

What we have done is to find the best linear constraint such that the total quadratic error (B.13) is minimized. The end result is the same as if we had minimized (B.13) directly, subject only to constraint

$$A^T A = \sum_{j=0}^n \alpha_j^2 = 1. \tag{B.34}$$

This latter interpretation corresponds to the best A vector in (n+1)-space, with its tip on the unit sphere, that minimizes the total quadratic error.

REFERENCES

1. J. R. Auton and M. L. Van Blaricum, "Investigation of Procedures for Automatic Resonance Extraction from Noisy Transient Electromagnetics Data," Final Report for Contract N00014-80-C-0299, Effects Technology Inc., Santa Barbara, CA, 17 August 1981.

2. F. B. Hildebrand, Introduction to Numerical Analysis, McGraw-Hill Book Co., New York, 1956.

3. M. L. Van Blaricum, "Problems and Solutions Associated with Prony's Method for Processing Transient Data," IEEE Transactions on Antennas and Propagation, vol. AP-26, no. 1, pp. 174-182, January 1978.

On Generation of Random Numbers With Specified Distributions Or Densities

A. H. Nuttall

ABSTRACT

Generation of random numbers with specified probability density functions or cumulative distribution functions is reviewed and employed to generate some standard random variables with common densities and distributions. Combinations of random variables then afford a quick method of generating variables with more-involved distribution functions. Application to the Q_M cumulative distribution function is made as an example. Timing results for all cases are listed. Numerous statistical tests on the Hewlett-Packard 9845 Desk Calculator confirm it as a reliable generator of uniformly distributed statistically independent random numbers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF SYMBOLS

x random variable uniformly distributed over $(0,1)$

$g(x)$ monotonic nonlinear distortion of x

y result of distortion g

t nonrandom threshold

$\text{Prob}\{\}$ probability of $\{\}$

P_y cumulative distribution function of random variable y

p_y probability density function of random variable y

f_y characteristic function of random variable y

$\bar{h}$ inverse function to h

Φ Gaussian cumulative distribution function; (18)

E expectation; ensemble average

d additive constant; (36)

Q_M Q_M function; (50)

K_ν modified Bessel function of second kind and order ν; (54)

R_k sample correlation; (65)

μ_k sample moment; (73)

q_k sample probability; (77)

DFT   discrete Fourier transform

ON GENERATION OF RANDOM NUMBERS WITH SPECIFIED
DISTRIBUTIONS OR DENSITIES

INTRODUCTION

The generation of random numbers with a specified probability density function or a specified cumulative distribution function is a frequent occurrence in the simulation of signal processing techniques that are difficult or impossible to evaluate analytically. Accordingly, it is of interest to be able to generate easily and efficiently such random numbers. It is assumed that a uniform random number generator is already available; that is, independent random variables with a probability density function equal to unity over the range (0,1) can be generated. Validation of this assumption for a particular random number generator must be confirmed by numerical tests, several of which are tested herein. This report is an extension and elaboration of [1].

NONLINEAR DISTORTION OF A SINGLE UNIFORM RANDOM VARIABLE

Suppose that random variable x is uniformly distributed on (0,1). Let the nonlinear distortion of x yield the random variable y given by

$$y = g(x) = \begin{cases} g_i(x) \\ \text{or} \\ g_d(x) \end{cases} , \tag{1}$$

where

$g_i(\ )$ is a monotonically increasing function of its argument,

$g_d(\ )$ is a monotonically decreasing function of its argument. $\qquad$ (2)

This restriction of the nonlinear distortion to be monotonic (either increasing or decreasing) is not necessary; it is done here only for simplicity. Then since random variable x is uniformly distributed on (0,1), for a fixed (nonrandom) t in the range (0,1),

$$t = \text{Prob}\left\{x < t\right\} = \text{Prob}\begin{cases} g_i(x) < g_i(t) \\ \text{or} \\ g_d(x) > g_d(t) \end{cases}$$

$$= \text{Prob}\begin{cases} y < g_i(t) \\ \text{or} \\ y > g_d(t) \end{cases} = \begin{cases} P_y(g_i(t)) \\ \text{or} \\ 1 - P_y(g_d(t)) \end{cases} , \tag{3}$$

where we used (2), (1), and denoted the cumulative distribution function of random variable y by $P_y(\ )$. We rewrite (3) as

$$t = P_y(g_i(t))$$

$$\text{or} \tag{4}$$

$$1 - t = P_y(g_d(t)) \quad ,$$

and observe that cumulative distribution function $P_y(\ )$ is a monotonically increasing function of its argument.

Now, for arbitrary monotonic function h( ), we denote its inverse function by $\tilde{h}(\ )$; that is,*

$$\tilde{h}(h(u)) = u, \qquad h(\tilde{h}(v)) = v \quad . \tag{5}$$

Applying monotonic function $\tilde{P}_y(\ )$ to both sides of (4), and using (5), there follows

$$g_i(t) = \tilde{P}_y(t)$$
$$\text{or} \tag{6}$$
$$g_d(t) = \tilde{P}_y(1 - t) \quad .$$

Employing (6) in (1) finally yields the random variable

$$y = g(x) = \begin{cases} g_i(x) = \tilde{P}_y(\dot{x}) \\ \qquad \text{or} \\ g_d(x) = \tilde{P}_y(1 - x) \end{cases} \tag{7}$$

as the desired result of this nonlinear transformation of random variaole x. If cumulative distribution function $P_y(\ )$ of random variaole y is specified and desired, and if random variable x is uniformly distributed on (0,1), (7) tells us that we must evaluate the inverse of the given cumulative distribution function and use it as the nonlinear transformation of either random variable x or random variable 1-x, depending on wnether we want a monotonically increasing or monotonically decreasing transformation, respectively. The key element to this approach is the ease with wnich the inverse function $\tilde{P}_y(\ )$ can be computed.

---

* Some additional relationships among functions and their inverses are presented in appendix A.

EXAMPLES OF SINGLE-VARIATE DISTORTION

The following examples have been adjusted to a convenient scale, such as zero mean or unit variance. By addition of a constant and/or multiplication by a scale factor, alternative desired ranges can be realized. The times of execution given below were obtained for the Hewlett-Packard 9845B Desk Calculator equipped with the Fast Processor Upgrade Kit; they include the time required to generate the uniform random variable x. Results were obtained by averaging over 1000 independent trials. Loop counters were declared INTEGER for maximum speed.

EXPONENTIAL DENSITY

The desired probability density function of random variable y is exponential*:

$$p_y(u) = \exp(-u) \quad \text{for} \quad u > 0 \quad . \tag{8}$$

The corresponding cumulative distribution function is

$$P_y(t) = \int_{-\infty}^{t} du \, p_y(u) = 1 - \exp(-t) \quad \text{for} \quad t > 0 \quad . \tag{9}$$

The characteristic function of random variable y is (for future reference)

$$f_y(\xi) = \int_{-\infty}^{+\infty} du \, \exp(i\xi u) \, p_y(u) = (1 - i\xi)^{-1} \quad . \tag{10}$$

In order to find the inverse function to (9), we let

$$u = P_y(t) = 1 - \exp(-t) \quad . \tag{11}$$

Solving the left relation for t, and then solving the outside relation for t, we get, respectively,

---

* The probability density functions and cumulative distribution functions are zero in the unspecified regions; for example, $p_y(u) = 0$ for $u < 0$.

$$t = \tilde{P}_y(u), \qquad t = -\ln(1 - u) \quad . \tag{12}$$

Combining these two, there follows for the inverse function

$$\tilde{P}_y(u) = -\ln(1-u) \quad \text{for} \quad 0 < u < 1 \quad . \tag{13}$$

The nonlinear transformation, by reference to (7), is then

$$y = \left\{ \begin{array}{c} -\ln(1-x) \\ \text{or} \\ -\ln x \end{array} \right\} \quad . \tag{14}$$

The time of execution for generation of a random variable y via (14) is 1.9 msec.

RAYLEIGH DENSITY

$$p_y(u) = u \exp(-u^2/2) \qquad \text{for} \quad u > 0 \quad ,$$

$$P_y(u) = 1 - \exp(-u^2/2) \quad \text{for} \quad u > 0 \quad . \tag{15}$$

Via a procedure similar to (11) and (12), there follows

$$\tilde{P}_y(u) = \sqrt{-2 \ln(1-u)} \quad \text{for} \quad 0 < u < 1 \quad . \tag{16}$$

The nonlinear transformation that yields Rayleigh random variables is

$$y = \left\{ \begin{array}{c} \sqrt{-2 \ln(1-x)} \\ \text{or} \\ \sqrt{-2 \ln x} \end{array} \right\} \quad . \tag{17}$$

The time of execution of (17) is 2.5 msec.

$$p_y(u) = (2\pi)^{-1/2} \exp(-u^2/2) \quad \text{for all } u,$$

$$P_y(u) = \int_{-\infty}^{u} dt \ (2\pi)^{-1/2} \exp(-t^2/2) \equiv \bar{\Phi}(u) \quad , \tag{18}$$

$$f_y(\xi) = \exp(-\xi^2/2) \quad .$$

The inverse function is [2; 26.2, 26.2.22, 26.2.23]

$$\tilde{P}_y(u) = \tilde{\bar{\Phi}}(u) = -\tilde{\bar{\Phi}}(1-u) \text{ for } 0 < u < 1 \quad . \tag{19}$$

The nonlinear transformation that yields Gaussian random variables from uniformly distributed ones is then, from (7) and (19),

$$y = \left\{ \begin{array}{c} \tilde{\bar{\Phi}}(x) \\ \text{or} \\ -\tilde{\bar{\Phi}}(x) \end{array} \right\} \quad . \tag{20}$$

The time of execution of (20) is 13.2 msec.

A much better approach, for this particular case of generation of Gaussian random variables, is as follows: let x_1 and x_2 be two independent random variables, uniformly distributed on (0,1). Then according to the previous example, the two independent random variables

$$r = \sqrt{-2 \ln x_1} \quad , \quad \theta = 2\pi x_2 \quad , \tag{21}$$

have, respectively, the probability density functions

$$p_r(u) = u \exp(-u^2/2) \quad \text{for} \quad u > 0 \quad , $$

$$p_\theta(u) = (2\pi)^{-1} \qquad \text{for} \quad 0 < u < 2\pi \quad . \tag{22}$$

Now define two new random variables by the nonlinear transformations

$$y_1 = r \cos \theta, \quad y_2 = r \sin \theta \quad .$$
$$(23)$$

The joint characteristic function of y_1 and y_2 is [2; 9.1.21 and 11.4.29]

$$E\{\exp(i\xi y_1 + i\mu y_2)\} = E\{\exp(i\xi r \cos \theta + i\mu r \sin \theta)\}$$

$$= \iint\limits_{-\infty}^{+\infty} du\, dv\, p_r(u)\, p_\theta(v)\, \exp(i\xi u \cos v + i\mu u \sin v)$$

$$= \int\limits_{0}^{+\infty} du\, u \exp(-u^2/2) \int\limits_{0}^{2\pi} dv\, \frac{1}{2\pi} \exp(iu(\xi \cos v + \mu \sin v))$$

$$= \int\limits_{0}^{+\infty} du\, u \exp(-u^2/2)\, J_0\left(u\sqrt{\xi^2 + \mu^2}\right)$$

$$= \exp\left(-\frac{1}{2}\xi^2 - \frac{1}{2}\mu^2\right) \quad .$$
$$(24)$$

Thus y_1 and y_2 are independent Gaussian random variables, each with zero-mean and unit variance. The time of execution of (21) and (23) is 5.4 msec per random variable (actually 10.7 msec for a pair of independent Gaussian random variables.) This 5.4 msec is considerably less than the 13.2 msec required of (20), which also requires a special function definition. A more general distortion than (21)-(23) is considered in appendix B.

Another alternative for the generation of approximately Gaussian random variables is to sum N independent random variables, uniformly distributed over (0,1). By subtraction of the constant N/2 and scaling by $\sqrt{12/N}$, a zero-mean unit-variance random variable can be generated which, however, is limited to the finite range $(-\sqrt{3N}, \sqrt{3N})$. A table of execution times and range values is given below.

Table 1. Execution Times for Sum of N Independent Random Variables

N	Execution Time (msec)	Range of Values
4	2.8	-3.46, 3.46
5	3.3	-3.87, 3.87
6	3.9	-4.24, 4.24
7	4.4	-4.58, 4.58
8	4.9	-4.90, 4.90
9	5.4	-5.20, 5.20
10	5.9	-5.48, 5.48
20	11.0	-7.75, 7.75
30	16.2	-9.49, 9.49

The characteristic function of this random variable is $[\sin(R\xi)/(R\xi)]^N$ where $R = \sqrt{3/N}$, and the normalized fourth-cumulant of the random variable generated by this summation procedure is $-1.2/N$. If the non-Gaussianness can be tolerated, this summation procedure is then a viable alternative to (21) and (23) in terms of execution time, for $N \leq 9$.

CAUCHY DENSITY

$$p_y(u) = \frac{1}{\pi} \frac{1}{1 + u^2} \quad \text{for all } u,$$

$$P_y(u) = \frac{1}{2} + \frac{1}{\pi} \arctan(u),$$

$$f_y(\xi) = \exp(-|\xi|) \quad . \tag{25}$$

This is the probability density function of the ratio of two zero-mean independent Gaussian random variables. The inverse function to the cumulative distribution function is

$$\tilde{p}_y(u) = \tan\left[\pi\left(u - \frac{1}{2}\right)\right] \quad \text{for} \quad 0 < u < 1 \quad , \tag{26}$$

and the nonlinear transformation is

$$y = \begin{Bmatrix} \tan\left[\pi\left(x - \tfrac{1}{2}\right)\right] \\ \text{or} \\ -\tan\left[\pi\left(x - \tfrac{1}{2}\right)\right] \end{Bmatrix}. \tag{27}$$

The execution time of (27) is 2.8 msec. Equivalently, the transformation $\tan(\pi x)$ would realize the desired probability density function, although it is not monotonic in x over (0,1).

RECTIFIED CAUCHY DENSITY

$$p_y(u) = \frac{2}{\pi}\frac{1}{1 + u^2} \qquad \text{for } u > 0 \quad,$$

$$P_y(u) = \frac{2}{\pi}\text{ arc tan}(u) \quad \text{for } u > 0 \quad,$$

$$\tilde{P}_y(u) = \tan\left(\frac{\pi}{2}u\right) \qquad \text{for } 0 < u < 1 \quad,$$

$$y = \begin{Bmatrix} \tan\left(\frac{\pi}{2}x\right) \\ \text{or} \\ \cot\left(\frac{\pi}{2}x\right) \end{Bmatrix}. \tag{28}$$

The execution time of (28) is 2.7 msec.

REALIZATION OF DISTRIBUTION VIA COMBINATION OF SEVERAL RANDOM VARIABLES

If several independent random variables are added, their characteristic functions multiply. So if a specified probability density function or cumulative distribution function has a characteristic function which can be broken down into easily realizable terms, this observation can be utilized to generate complicated distributions with relative ease. For example, consider the following.

CHI-SQUARED DENSITY WITH 2N DEGREES OF FREEDOM

This variate is normally thought of as being generated by summing the squares of 2N zero-mean unit-variance independent Gaussian random variables. One half of this sum has the probability density function

$$p_y(u) = \frac{u^{N-1} \exp(-u)}{(N-1)!} \quad \text{for } u > 0 \quad . \tag{29}$$

The cumulative distribution function is

$$P_y(u) = 1 - \exp(-u) \sum_{n=0}^{N-1} u^n/n! \quad \text{for } u > 0 \quad . \tag{30}$$

The inverse function, $\tilde{P}_y(\ )$, to (30) is not available in closed form for $N \geq 2$, so that recourse to (7) is not reasonable.

However, the characteristic function corresponding to (29) is

$$f_y(\xi) = (1 - i\xi)^{-N} \quad . \tag{31}$$

But this is (10) multiplied by itself N times. Then (14) reveals that we can generate y according to a sum of N independent variates:

$$y = \sum_{n=1}^{N} \left\{ -\ln x_n \right\} = -\ln \left\{ \prod_{n=1}^{N} x_n \right\} \quad . \tag{32}$$

The latter form in (32) is preferable computationally; it involves N-1 multiplies and only one logarithm, and is obviously much quicker than

10

generating 2N Gaussian random variables and summing their squares. The time of execution of (32) is given below.

Table 2. Execution Times for Chi-Squared Variate of 2N Degrees of Freedom

N	Execution Time (msec)
10	8.0
20	14.1
30	20.2

If we were to use the nonlinear distortion procedure in (20) for generating Gaussian random variables, and square and add 2N such numbers to generate a Chi-squared variate, the time of execution would be 2N (13.2) msec. For the examples in table 2, these times are 264, 528, and 792 msec respectively, which are far greater than the execution times for (32).

A closely related random variable to (32) is

$$r = \sqrt{2y} = \left[-2 \ln \left\{ \prod_{n=1}^{N} x_n \right\} \right]^{\frac{1}{2}} \equiv g(y) \quad . \tag{33}$$

The probability density function of random variable r is

$$p_r(u) = p_y(\breve{g}(u)) \frac{d}{du} \breve{g}(u) = p_y(u^2/2) \, u$$

$$= \frac{u^{2N-1} \exp(-u^2/2)}{2^{N-1}(N-1)!} \quad \text{for} \quad u > 0 \quad . \tag{34}$$

The cumulative distribution function of r is

$$P_r(u) = 1 - \exp(-u^2/2) \sum_{n=0}^{N-1} \frac{(u^2/2)^n}{n!} \quad \text{for} \quad u > 0 \quad . \tag{35}$$

11

The time of execution of (33) is given in table 3.

Table 3. Execution Times for Square Root of Chi-Squared Variate

N	Execution Time (msec)
10	8.6
20	14.7
30	20.8

Another example of this combination of variates is furnished by the following.

Q DISTRIBUTION AND RICE DENSITY

Suppose that y_1 and y_2 are two independent zero-mean unit-variance Gaussian random variables, as generated via (21) and (23). Then for a constant d, the random variable

$$z = \frac{1}{2}\left[(y_1 + d)^2 + y_2^2\right] = \frac{1}{2}\left[d^2 + 2dy_1 + y_1^2 + y_2^2\right]$$

$$= \frac{1}{2}\left[d^2 + 2dr\cos\theta + r^2\right] \qquad (36)$$

has characteristic function [2; 9.1.21 and 11.4.29]

$$f_z(\xi) = E\{\exp(i\xi z)\} = E\left\{\exp\left(i\xi\left(\frac{d^2}{2} + dr\cos\theta + \frac{r^2}{2}\right)\right)\right\}$$

$$= \int_0^{+\infty} du\, u\, \exp(-u^2/2)\, \exp\left[\frac{i\xi}{2}(d^2 + u^2)\right] \int_0^{2\pi} dv\, \frac{1}{2\pi}\exp(i\xi\, d\, u\cos v)$$

$$= \exp(i\xi d^2/2) \int_0^{+\infty} du\, u\, \exp\left[-\frac{u^2}{2}(1-i\xi)\right] J_0(\xi\, d\, u)$$

$$= (1-i\xi)^{-1}\exp\left(\frac{d^2}{2}\frac{i\xi}{1-i\xi}\right) \qquad . \qquad (37)$$

12

The corresponding probability density function is

$$p_z(u) = \exp\left(-\frac{d^2}{2} - u\right) I_0\left(d\sqrt{2u}\right) \quad \text{for} \quad u > 0 \quad , \tag{38}$$

as may be confirmed by direct Fourier transformation (according to (10)) of (38); see [2; 11.4.29]. The cumulative distribution function of z is

$$P_z(u) = 1 - Q(d, \sqrt{2u}) \quad \text{for} \quad u > 0 \quad , \tag{39}$$

where Marcum's Q function is defined as

$$Q(a,b) = \int_b^{+\infty} dx \; x \; \exp\left(-\frac{a^2 + x^2}{2}\right) I_0(ax) \quad . \tag{40}$$

The time of execution of (36) and (21) is 7.2 msec; there is no need to employ (23).

A closely related random variable to (36) is

$$y = \sqrt{2z} = \left(d^2 + 2dr \cos\theta + r^2\right)^{1/2} . \tag{41}$$

The probability density function of y is the Rice density

$$p_y(u) = p_z(u^2/2) \; u$$

$$= u \exp\left(-\frac{d^2 + u^2}{2}\right) I_0(du) \quad \text{for} \quad u > 0 \quad , \tag{42}$$

and the cumulative distribution function of y is

$$P_y(u) = 1 - Q(d, u) \quad \text{for } u > 0 \quad . \tag{43}$$

The time of execution of (41) and (21) is 7.5 msec; there is no need to employ (23).

These two examples, (36) and (41), would again be analytically very difficult to realize by use of (7), because the inverse functions to cumulative distribution functions (39) and (43) are not available in closed form. Furthermore, different values of d are easily accommodated in (36) and (41), whereas the inverses to (39) and (43) would necessarily involve d as a parameter.

A more general situation is encountered for the following.

Q_M DISTRIBUTION

This example is a combination of the last two. Let, as in (36) and (21),

$$v = \frac{1}{2}\left[(y_1 + d)^2 + y_2^2\right] = \frac{1}{2}[d^2 + 2dr\cos\theta + r^2] \quad, \tag{44}$$

where

$$r = \sqrt{-2\ln x_1} \quad, \quad \theta = 2\pi x_2 \quad . \tag{45}$$

And as in (32), let

$$w = -\ln\left\{\prod_{m=3}^{M+1} x_m\right\} \quad . \tag{46}$$

All the random variables, $\left\{x_m\right\}_1^{M+1}$, are independent and uniformly distributed on (0,1). Then let random variable z be the sum of the above two:

$$z = v+w = \frac{1}{2}[d^2 + 2dr\cos\theta + r^2] - \ln\left\{\prod_{m=3}^{M+1} x_m\right\} \quad . \tag{47}$$

The characteristic function of random variable z follows upon use of (37) and (31):

$$f_z(\zeta) = (1-i\zeta)^{-M} \exp\left(\frac{d^2}{2} \frac{i\zeta}{1-i\zeta}\right) \quad . \tag{48}$$

The corresponding probability density function and cumulative distribution function for z are

$$p_z(u) = \left(\frac{u}{d^2/2}\right)^{\frac{M-1}{2}} \exp\left(-\frac{d^2}{2} - u\right) I_{M-1}(d\sqrt{2u}) \quad \text{for} \quad u > 0 \quad ,$$

$$P_z(u) = 1 - Q_M(d,\sqrt{2u}) \quad \text{for} \quad u > 0 \quad , \tag{49}$$

where the Q_M function is defined by [3]

$$Q_M(a,b) = \int_b^{+\infty} dx \left(\frac{x}{a}\right)^{M-1} x \exp\left(-\frac{a^2 + x^2}{2}\right) I_{M-1}(ax) \quad . \tag{50}$$

The execution time of (45) and (47) is given in table 4 for several values of M. The use of (7) would have required the inverse of (49), a rather formidable analytical task involving the two parameters d and M.

Table 4. Execution Times for Q_M Variate

M	Execution Time (msec)
10	14.6
20	20.6
30	26.7

A closely related random variable to (47) is

$$y = \sqrt{2z} = \left[d^2 + 2dr \cos\theta + r^2 - 2\ln\left\{\prod_{m=3}^{M+1} x_m\right\}\right]^{1/2} \quad . \tag{51}$$

There follows

$$p_y(u) = u\left(\frac{u}{d}\right)^{M-1} \exp\left(-\frac{d^2 + u^2}{2}\right) I_{M-1}(du) \text{ for } u > 0 \quad ,$$

$$P_y(u) = 1 - Q_M(d,u) \quad \text{for } u > 0 \quad . \tag{52}$$

The execution times for (45) and (51) were 0.4 msec larger than those given in table 4.

PRODUCTS OF RANDOM VARIABLES

We can now take products of some of the random variables above and generate additional cases of complicated probability density functions. For example, if we take the product of two random variables as given by (32),

$$z = y_1 y_2 = \ln\left\{\prod_{n=1}^{N_1} x_n^{(1)}\right\} \ln\left\{\prod_{n=1}^{N_2} x_n^{(2)}\right\} \quad , \tag{53}$$

where $\left\{x_n^{(1)}\right\}$ and $\left\{x_n^{(2)}\right\}$ are uniformly distributed, the probability density function of the product random variable is, from (29) and [4; 3.471 9],

$$p_z(u) = \int_{-\infty}^{+\infty} \frac{dt}{|t|} \, p_{y_1}(t) \, p_{y_2}\left(\frac{u}{t}\right)$$

$$= \int_0^{+\infty} dt \, \frac{t^{N_1-N_2-1} \, u^{N_2-1}}{(N_1-1)! \, (N_2-1)!} \, \exp\left(-t - \frac{u}{t}\right)$$

$$= \frac{2 \, u^{\frac{N_1+N_2}{2}-1}}{(N_1-1)! \, (N_2-1)!} \, K_{N_1-N_2}(2\sqrt{u}) \quad \text{for} \quad u > 0 \quad . \tag{54}$$

Here $K_\nu(\ )$ is a modified Bessel function of the second kind and order ν [2; section 9.6]; it is even in ν [2; 9.6.6]. Execution times for (53) can be determined from (32) and table 2.

The random variable $w = 2z^{1/2}$, where z is given by (53), has the probability density function

$$p_w(u) = \frac{2(u/2)^{N_1+N_2-1}}{(N_1-1)! \, (N_2-1)!} \, K_{N_1-N_2}(u) \quad \text{for} \quad u > 0 \quad . \tag{55}$$

The product of two independent zero-mean Gaussian random variables, as given by (21) and (23), is given more simply by

$$z = y_1 y_2 = r^2 \sin \theta \cos \theta = \frac{1}{2} r^2 \sin(2\theta) = -\ln(x_1) \sin(4\pi x_2) \quad . \qquad (56)$$

The probability density function of z can be determined exactly as in (54), with the result

$$p_z(u) = \frac{1}{\pi} K_0(|u|) \quad \text{for all } u \quad . \qquad (57)$$

The 4π sweep of the sin argument in (56) is unnecessary; the following will accomplish the same probability density function:

$$z = \ln(x_1) \cos(\pi x_2) \quad , \qquad (58)$$

where x_1 and x_2 are uniform over $(0,1)$. The time of execution of (58) is 6.0 msec.

GENERATION FROM A NON UNIFORMLY DISTRIBUTED RANDOM VARIABLE

Suppose that it is possible to generate a random variable x with cumulative distribution function $P_x(\ )$, and that the cumulative distribution function of

$$y = g(x) \tag{59}$$

is desired to be $P_y(\ )$. Then for $g = g_i$, a monotonically increasing function, the cumulative distribution function of x is

$$P_x(u) = \text{Prob}\left\{x < u\right\} = \text{Prob}\left\{g_i(x) < g_i(u)\right\}$$

$$= \text{Prob}\left\{y < g_i(u)\right\} = P_y(g_i(u)) \quad . \tag{60}$$

Applying inverse function $\tilde{P}_y$ to both sides of (60), the desired nonlinearity is

$$g_i(u) = \tilde{P}_y(P_x(u)) \quad . \tag{61}$$

Equation (7) is obviously the special case of this when x is uniformly distributed. Alternative expressions to (61) are available by use of some of the relations in appendix A.

If we want to use distortion $g = g_d$, a monotonically decreasing function, to generate y, we have

$$P_x(u) = \text{Prob}\left\{x < u\right\} = \text{Prob}\left\{g_d(x) > g_d(u)\right\}$$

$$= \text{Prob}\left\{y > g_d(u)\right\} = 1 - P_y(g_d(u)) \quad . \tag{62}$$

Inverting this relation,

$$g_d(u) = \tilde{P}_y\left(1 - P_x(u)\right) \quad . \tag{63}$$

This also reduces to (7) for a uniformly distributed random variable x. In either case, (61) or (63), the inverse function of the desired cumulative distribution function, P_y, must be realized. Then a cascade operation as dictated by (61) or (63) must be employed. The desired random variable is given by

$$y = \left\{ \begin{array}{c} \tilde{P}_y\big(P_x(x)\big) \\ \text{or} \\ \tilde{P}_y\big(1 - P_x(x)\big) \end{array} \right\} \, . \qquad (64)$$

TESTS OF UNIFORM RANDOM NUMBER GENERATOR

A critical component of the procedures described above is the generator of random variables, $\{x_n\}$, with a first-order probability density function that is flat over $(0,1)$, and with statistically independent samples. Here we will investigate several tests of the random number generator, RND, of the Hewlett-Packard 9845 Desk Calculator. Inherent in this investigation is the need to state the confidence associated with a particular measurement or estimate; for example, see [5].

CORRELATION TEST

The sample correlation of a set of N measurements $\{y_n\}_1^N$ is defined here as

$$R_k = \frac{1}{N-k} \sum_{n=k+1}^{N} y_n y_{n-k} \quad \text{for} \quad 0 \leq k < N \quad . \tag{65}$$

We presume that the $\{y_n\}$ are all independent with a flat probability density function

$$p_y(u) = \frac{1}{2\sqrt{3}} \quad \text{for} \quad |u| < \sqrt{3} \quad . \tag{66}$$

These random variables can be obtained from uniform $\{x_n\}$ according to the linear transformation

$$y_n = 2\sqrt{3}\left(x_n - \frac{1}{2}\right) ; \tag{67}$$

they have the convenient normalized values

$$E(y_n) = 0 \quad , \quad E(y_n^2) = 1 \quad . \tag{68}$$

More generally, from (66),

$$E\left\{y_n^{2m+1}\right\} = 0 \quad , \quad E\left\{y_n^{2m}\right\} = \frac{3^m}{2m+1} \quad . \tag{69}$$

21

Under these assumptions, it is readily shown, by use of (69), that R_k is unbiased, that is

$$E(R_k) = \begin{cases} 1 \text{ for } k = 0 \\ 0 \text{ otherwise} \end{cases} , \qquad (70)$$

and that the standard deviation of R_k is

$$\sigma(R_k) = \begin{cases} (1.25N)^{-1/2} & \text{for} \quad k = 0 \\ (N-k)^{-1/2} & \text{for} \quad 1 \le k < N \end{cases} . \qquad (71)$$

Thus we expect R_k to lie within 2 or 3 standard deviations, $\sigma(R_k)$, of $E(R_k)$ most of the time; that is, the normalized random variable

$$r_k = \frac{R_k - E(R_k)}{\sigma(R_k)} \qquad (72)$$

should be between $\pm$ 2 most of the time, with rare excursions to $\pm$ 3, if the $\{x_n\}$ are truly independent [5]. In table 5, the results of sample runs for N=100, 1000, and 10000 are listed. They furnish no reason for rejecting the hypothesis that $\{x_n\}$ are independent. Runs for other sets of random numbers yielded results very similar to table 5. An alternative test on the whiteness of $\{x_n\}$ is given in appendix C.

Table 5.   Sample Correlation Results for (72)

Delay k	r_k for N = 100	r_k for N = 1000	r_k for N = 10000
0	1.64	0.78	0.65
1	-1.04	-0.91	0.25
2	-0.26	-0.68	-0.02
3	-0.10	1.39	-0.13
4	1.44	0.31	0.27
5	-2.31	-0.30	0.47
6	-0.49	-1.20	-0.20
7	0.83	0.81	-1.40
8	0.17	0.61	-1.47
9	-0.61	-0.77	-0.91
10	-1.43	-0.45	0.48

MOMENT TEST

The sample moment of order k for a set of N measurements $\{y_n\}$ is

$$\mu_k = \frac{1}{N} \sum_{n=1}^{N} y_n^k \quad . \tag{73}$$

The mean and variance of μ_k are

$$E(\mu_k) = E\left(y_n^k\right), \quad Var(\mu_k) = \frac{1}{N} \, Var\left(y_n^k\right) \quad , \tag{74}$$

and can be obtained from (66) and (69). The random variable

$$m_k = \frac{\mu_k - E(\mu_k)}{\text{Std. Dev.}(\mu_k)} \tag{75}$$

should therefore lie mostly in the range ± 2. A sample result is given in table 6, which again confirms the flat probability density function hypothesis in (66). (Although these particular runs all resulted in positive numbers, other sample runs resulted in a preponderance of negative values for m_k.)

Table 6. Sample Moment Results for (75)

Moment k	m_k for N = 100	m_k for N = 1000	m_k for N = 10000
1	1.07	1.37	0.52
2	1.64	0.78	0.65
3	1.08	1.09	0.42
4	1.49	0.67	0.89
5	0.95	1.11	0.30
6	1.21	0.51	0.94

FIRST-ORDER DISTRIBUTION TEST

Let I_k be an interval of the line segment $(0,1)$. Then let c_n be a counting variable given by

$$c_n = \begin{Bmatrix} 1 \text{ if } x_n \in I_k \\ 0 \text{ otherwise} \end{Bmatrix} \quad \text{for } 1 \leq n \leq N \quad . \tag{76}$$

Then an estimate of the probability P_k, that $x_n \in I_k$, is given by the quantity

$$q_k = \frac{1}{N} \sum_{n=1}^{N} c_n \quad . \tag{77}$$

This random variable has mean and variance

$$E(q_k) = P_k, \quad Var(q_k) = P_k(1-P_k)/N \quad . \tag{78}$$

Thus the random variable

$$v_k = \frac{q_k - P_k}{\left[P_k(1-P_k)/N\right]^{1/2}} \tag{79}$$

should lie most often in the region ± 2 if x_n truly does have probability P_k of lying in I_k. In table 7 is displayed the results of sample runs for the case where interval $(0,1)$ is broken into 10 equal parts; that is,

$$I_k = \left(\frac{k-1}{10}, \frac{k}{10}\right) \quad \text{for} \quad 1 \leq k \leq 10 \quad . \tag{80}$$

Once again, there is no reason to reject the hypothesis that the random number generator is uniformly distributed over $(0,1)$.

Table 7. Sample Probability Results for (79)

Interval k	v_k for N = 1000	v_k for N = 10000	v_k for N = 100000
1	-0.74	0.67	0.50
2	1.16	-0.57	0.67
3	-1.26	-1.13	-1.02
4	-1.37	0.03	-0.96
5	-0.32	0.90	-1.26
6	-0.32	-0.73	-0.64
7	1.48	-0.27	0.53
8	-0.53	-0.13	-0.32
9	0.32	-0.07	1.87
10	1.58	1.30	0.64

SUMMARY

Several methods of generating random variables with specified cumulative distribution functions have been presented and evaluated in terms of their time of execution and efficiency of generation. They include nonlinear distortion of a single (uniformly distributed) random variable or through combinations of simply generated random variables. The former approach requires the ability to realize the inverse function to the desired cumulative distribution function, whereas the latter approach is very fruitful if the characteristic function corresponding to the specified cumulative distribution function can be broken down into a product of simpler components. Which approach to adopt depends on the particular example and the exact way that any parameters enter into the cumulative distribution function. Of course, the inverse to a cumulative distribution function can always be evaluated numerically and used as a table look-up, in order to generate transformed random variables; however, this numerical procedure would have to be repeated if any parameter of the cumulative distribution function were desired changed.

The ability to generate uniformly distributed random variables is a key component of this procedure. Several statistical tests on the Hewlett-Packard 9845 random number generator have confirmed it to have a flat probability density function and independent samples.

These techniques are useful for generating sets of random variables of size N with a specified cumulative distribution function, and then plotting the sample cumulative distribution functions as described in [6], in order to ascertain the amount of fluctuation which is typical for that set size N and in different regions of probability. Then by comparing the amount of fluctuation of a measured data set (of unknown statistics) with typical sample cumulative distribution functions of the same size and with a known specified cumulative distribution function, decisions on acceptance or rejection of a hypothesized cumulative distribution function can be made with confidence. See [6, figures 6 and 7] for illustrations of this procedure.

APPENDIX A. SOME INVERSE FUNCTION RELATIONS

In this appendix, x and y are real (non random) variables. Let g(x) be a monotonic function for x in a given range, and let

$$y = g(x) \quad . \tag{A-1}$$

The inverse relation to (A-1) is, for y in the appropriate range,

$$x = \tilde{g}(y) \quad . \tag{A-2}$$

Now suppose that we cascade nonlinearities: let

$$z = h(y) = h(g(x)) \equiv f(x) \quad , \tag{A-3}$$

to yield overall nonlinearity f. Then the outer equality in (A-3) yields

$$x = \tilde{f}(z) \quad , \tag{A-4}$$

whereas the first and third terms in (A-3) yield

$$\tilde{h}(z) = g(x), \quad \text{and} \quad \tilde{g}(\tilde{h}(z)) = x \quad . \tag{A-5}$$

Combining (A-4) and (A-5), there follows for the inverse function of the cascade, (A-3),

$$\tilde{f}(z) = \tilde{g}(\tilde{h}(z)) \quad , \tag{A-6}$$

in terms of the inverses of the individual transformations.

If we combine (A-1) and (A-2), we get

$$y = g(\tilde{g}(y)) \quad . \tag{A-7}$$

Taking a derivative with respect to y, we find

$$1 = g'(\tilde{g}(y)) \, \tilde{g}'(y) \quad ;$$

$$\tilde{g}'(y) = \frac{1}{g'(\tilde{g}(y))} \quad . \tag{A-8}$$

That is, the derivative of the inverse, $\tilde{g}$, can be found in terms of the inverse function and the derivative of the original function g.

Suppose that y is given in terms of x via transformation

$$y = g(\tilde{f}(x)) \quad , \tag{A-9}$$

but the inverse function $\tilde{f}$ is impossible or very difficult to obtain from given function f. A simple way of evaluating y vs x, then, is parametrically by letting

$$t = \tilde{f}(x), \text{ to get } x = f(t), \, y = g(t) \quad . \tag{A-10}$$

Now, as t is varied, f and g can be evaluated to determine y vs x. This transformation also allows evaluation of an integral involving an inverse function:

$$\int_a^b dx \, g(\tilde{f}(x)) = \int_{t_a}^{t_b} dt \, f'(t) \, g(t) \quad , \tag{A-11}$$

where

$$t_a = \tilde{f}(a), \quad t_b = \tilde{f}(b) \quad . \tag{A-12}$$

Another use of inverse functions in integral evaluation is afforded by the example

$$I = \int_a^b dx \, g(x) \, w(x) \quad , \tag{A-13}$$

where function $w(x)$ need not be monotonic or possess an inverse. $g(x)$ is assumed monotonic in (a,b). Letting

$$y = g(x), \quad x = \tilde{g}(y), \quad dx = dy\, \tilde{g}'(y) \quad , \tag{A-14}$$

there follows, for the integral in terms of inverse function $\tilde{g}$,

$$I = \int_{g(a)}^{g(b)} dy\, y\, \tilde{g}'(y)\, w(\tilde{g}(y)) \quad . \tag{A-15}$$

Integrating by parts, using identification

$$u = y\, w(\tilde{g}(y)), \quad v = \tilde{g}(y) \quad , \tag{A-16}$$

we get the alternative form

$$I = b\, g(b)\, w(b) - a\, g(a)\, w(a) - \int_{g(a)}^{g(b)} dy\, \tilde{g}(y)\, \frac{d}{dy}\left\{ y\, w(\tilde{g}(y)) \right\} \quad . \tag{A-17}$$

The special case of $w(x) = 1$ in (A-13) and (A-17), namely

$$\int_{a}^{b} dx\, g(x) = b\, g(b) - a\, g(a) - \int_{g(a)}^{g(b)} dy\, \tilde{g}(y) \quad , \tag{A-18}$$

has the geometrical interpretation in figure A-1. Equation (A-18) is the statement that $A_1 + A_2 + A_3 =$ total area $b\, g(b)$.

As an example of the application of (A-13) and (A-15), consider

$$g(x) = \text{arc } \sin(x), \quad \tilde{g}(y) = \sin y \quad . \tag{A-19}$$

There follows

$$\int_{a}^{b} dx\, \text{arc } \sin(x)\, w(x) = \int_{\text{arc } \sin(a)}^{\text{arc } \sin(b)} dy\, y\, \cos(y)\, w(\sin y). \tag{A-20}$$

So if w is a polynomial, the integral on the right side of (A-20) can be
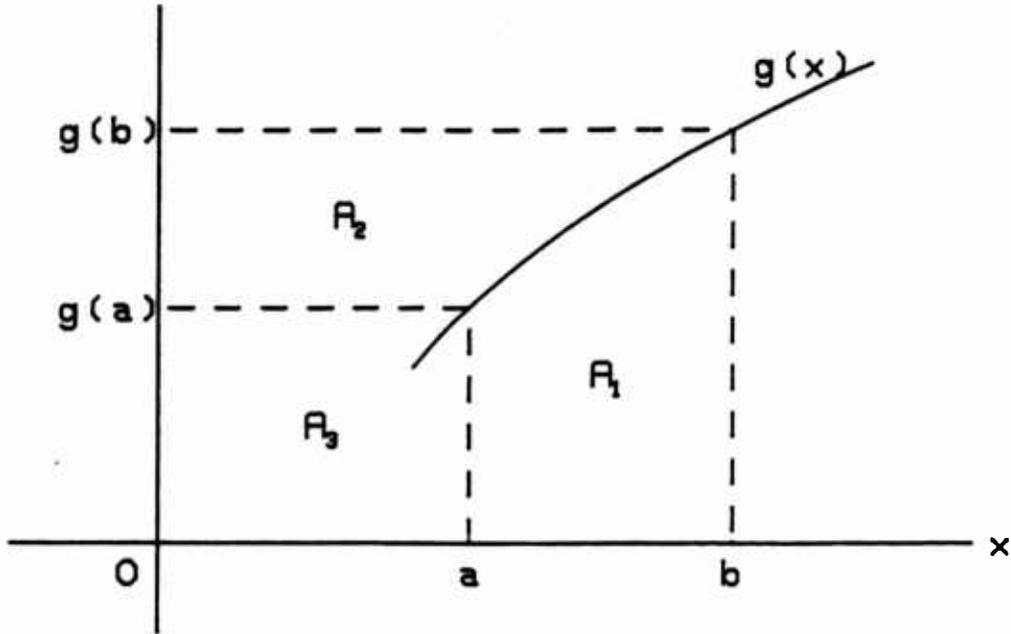evaluated in closed form.

Figure A-1.   Geometrical Interpretation of (A-18)

APPENDIX B. A MORE GENERAL DISTORTION

In (21), a random variable r with a Rayleigh probability density function and a random variable θ with a uniform probability density function were generated. These were then used in nonlinear transformation (23) to generate two independent Gaussian random variables y_1 and y_2. Here we will generalize the probability density function p_r of r in (22), and allow p_r to be arbitrary. θ is still uniform over 2π.

The two new random variables generated are again as in (23):

$$y_1 = r \cos\theta, \quad y_2 = r \sin\theta \quad . \tag{B-1}$$

Because of the uniform distribution of θ over 2π, the joint probability density function of y_1 and y_2 is of the symmetric form

$$p(y_1, y_2) = h\left(\sqrt{y_1^2 + y_2^2}\right) \text{ for all } y_1, y_2 \quad . \tag{B-2}$$

To determine h, observe that, for $t > 0$,

$$P \equiv \text{Prob}\left\{\sqrt{y_1^2 + y_2^2} < t\right\} = \iint\limits_{C_t} dy_1 \ dy_2 \ p(y_1, y_2)$$

$$= \iint\limits_{C_t} dy_1 \ dy_2 \ h\left(\sqrt{y_1^2 + y_2^2}\right) = \int_0^t d\rho \ \rho \int_{-\pi}^{\pi} d\phi \ h(\rho) = 2\pi \int_0^t d\rho \ \rho \ h(\rho) \ , \tag{B-3}$$

where C_t is a circle of radius t centered at the origin. But also, from (B-1),

$$P = \text{Prob}\left\{r < t\right\} = \int_0^t du \ p_r(u) \quad . \tag{B-4}$$

Equating (B-3) and (B-4) and taking a derivative with respect to t, there follows

$$h(t) = \frac{p_r(t)}{2\pi t} \quad \text{for } t > 0 \quad . \tag{B-5}$$

31

Reference to (B-2) then yields for the joint probability density function of y_1, y_2,

$$p(y_1, y_2) = \frac{p_r\left(\sqrt{y_1^2 + y_2^2}\right)}{2\pi\sqrt{y_1^2 + y_2^2}} \quad \text{for all } y_1, y_2 \quad . \tag{B-6}$$

y_1 and y_2 are statistically dependent in general.

EXAMPLE 1.

Our first case is the probability density function in (34), for random variable r as generated by (33):

$$p_r(u) = \frac{u^{2N-1} \exp(-u^2/2)}{2^{N-1}(N-1)!} \quad \text{for} \quad u > 0 \quad . \tag{B-7}$$

Substitution in (B-6) yields the joint probability density function

$$p(y_1, y_2) = \left[2\pi(N-1)!\right]^{-1} \left(\frac{y_1^2 + y_2^2}{2}\right)^{N-1} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \quad \text{for all } y_1, y_2 \quad . \tag{B-8}$$

The special case of N=1 reduces to the pair of Gaussian random variables already considered in (21)-(24). For N=2, (B-7) and (B-8) yield

$$p_r(u) = \frac{1}{2} u^3 \exp(-u^2/2) \quad ,$$

$$p(y_1, y_2) = \frac{y_1^2 + y_2^2}{4\pi} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \quad . \tag{B-9}$$

EXAMPLE 2.

Here random variable r is generated according to (32), and has the probability density function given by (29):

$$p_r(u) = \frac{u^{N-1} \exp(-u)}{(N-1)!} \quad \text{for} \quad u > 0 \quad . \tag{B-10}$$

Substitution of (B-10) in (B-6) yields

$$p(y_1, y_2) = [2\pi (N-1)!]^{-1} \left(y_1^2 + y_2^2\right)^{\frac{N}{2} - 1} \exp\left(-\sqrt{y_1^2 + y_2^2}\right) \text{ for all } y_1, y_2. \quad \text{(B-11)}$$

Special cases for N equal to 1 and 2 are respectively

$$p(y_1, y_2) = (2\pi)^{-1} \left(y_1^2 + y_2^2\right)^{-1/2} \exp\left(-\sqrt{y_1^2 + y_2^2}\right) \qquad \text{(B-12)}$$

and

$$p(y_1, y_2) = (2\pi)^{-1} \exp\left(-\sqrt{y_1^2 + y_2^2}\right) \ . \qquad \text{(B-13)}$$

APPENDIX C. A TEST FOR WHITENESS OF A SEQUENCE

Suppose data points $\{x_k\}_0^{K-1}$ are available. We can form a correlation estimate at delay n according to*

$$\hat{R}_n = \frac{1}{K} \sum_k x_k \, x_{k-n}^* \quad \text{for all n} \quad , \tag{C-1}$$

where summations without limits are over the range of nonzero summands. $\hat{R}_n$ is nonzero only for $|n| < K$. If process $\{x_n\}$ were white, we would expect to have

$$\left|\hat{R}_n\right| << \hat{R}_0 \quad \text{for all } n \neq 0 \quad . \tag{C-2}$$

Therefore a measure of nonwhiteness is afforded by the ratio $E/\hat{R}_0^2$, where error measure E is defined as the sum of squares for $n \neq 0$,

$$E = \sum_{n \neq 0} \left|\hat{R}_n\right|^2 = \sum_n \left|\hat{R}_n\right|^2 - \hat{R}_0^2 \quad . \tag{C-3}$$

However, (C-3) is very time-consuming to calculate via (C-1), because of all the multiplications and additions required. A much more practical evaluation of (C-3) is afforded by the following procedure (the derivation is presented later in this appendix).

Define an M-point DFT of the K data points:

$$X_m = \sum_{k=0}^{K-1} x_k \, \exp(-i2\pi km/M) \quad \text{for} \quad 0 \leq m \leq M-1 \quad . \tag{C-4}$$

Then it follows that if $M \geq 2K-1$,

$$\sum_n \left|\hat{R}_n\right|^2 = \frac{1}{K^2 M} \sum_{m=0}^{M-1} \left|X_m\right|^4 \quad , \tag{C-5}$$

---

* The following mathematical development is very similar to that in [7].

and our whiteness measure can be expressed as

$$\frac{E}{\hat{R}_0^2} = \frac{M \sum_{m=0}^{M-1} |X_m|^4}{\left(\sum_{m=0}^{M-1} |X_m|^2\right)^2} - 1 \quad . \tag{C-6}$$

A threshold test of (C-6) is equivalent to

$$Q \equiv \frac{M \sum_{m=0}^{M-1} |X_m|^4}{\left(\sum_{m=0}^{M-1} |X_m|^2\right)^2} \begin{array}{c} \text{non-white} \\ > \\ < \\ \text{white} \end{array} \quad \text{Threshold} \quad . \tag{C-7}$$

The distribution of Q could be computed from white-noise simulations, and a threshold value selected for prescribed error probability. By Schwarz's inequality, $Q \geq 1$, with equality realized if and only if $|X_m|^2$ is constant for all m.

Evaluation of Q in (C-7) requires one M-point DFT of the data $\{x_k\}_0^{K-1}$, where we must have

$$M \geq 2K - 1 \quad . \tag{C-8}$$

The subsequent calculations in (C-7) are quickly conducted.

An alternative interpretation of error measure E in (C-3) is very illuminating and lends additional credence to (C-7) as an appropriate statistic. If we define spectral estimate

$$\hat{G}_m = \sum_n \hat{R}_n \exp(-i2\pi mn/M) \quad \text{for all } m \quad , \tag{C-9}$$

then we find

$$\hat{G}_m = \frac{1}{K} |X_m|^2 \quad \text{for } 0 \leq m \leq M-1 \quad , \tag{C-10}$$

and the error measure E in (C-3) becomes

$$E = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^2 - \left(\frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m \right)^2 \quad . \qquad (C-11)$$

But notice that if we define the sample mean of the set of spectral estimates $\left\{ \hat{G}_m \right\}_0^{M-1}$ as

$$\mu = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m \quad , \qquad (C-12)$$

then the sample variance becomes

$$\sigma^2 = \frac{1}{M} \sum_{m=0}^{M-1} (\hat{G}_m - \mu)^2$$

$$= \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^2 - \left(\frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m \right)^2 = E \quad . \qquad (C-13)$$

That is, error measure E is equal to the sample variance of the set of spectral estimates. This latter quantity is a very natural measure for deducing whether a sequence is white, since σ^2 would be expected to be smaller for a truly white process.

This also suggests that a meaningful measure, for determining whether a sequence is white over a limited band of the zero-to-Nyquist range, is the sample variance of the spectral estimates over that particular band in question.

For a real sequence $\left\{ x_k \right\}_0^{K-1}$, the sums in (C-7) need only be conducted over half of the range, by virtue of the conjugate symmetry:

$$X_{M-m} = X_m^* \quad \text{for} \quad 1 \le m \le M-1 \quad \text{if} \quad \left\{ x_k \right\} \text{ real} \quad . \qquad (C-14)$$

Thus then (for M even)

$$\sum_{m=0}^{M-1} |X_m|^4 = X_0^4 + X_{\frac{M}{2}}^4 + 2 \sum_{m=1}^{\frac{M}{2}-1} |X_m|^4 \quad \text{if} \quad \{x_k\} \text{ real.} \quad (C-15)$$

DERIVATIONS

Define spectral estimate

$$\hat{G}_m^{(M)} = \sum_n \hat{R}_n \exp(-i2\pi nm/M) \quad \text{for all } m \quad . \quad (C-16)$$

The superscript specifically indicates the period in the subscript variable. Substituting (C-1) in (C-16), there follows

$$\hat{G}_m^{(M)} = \sum_n \exp(-i2\pi nm/M) \frac{1}{K} \sum_k x_k\, x_{k-n}^{\ \star}$$

$$= \frac{1}{K} \sum_k x_k \exp(-i2\pi km/M) \left[\sum_n x_{k-n} \exp(-i2\pi(k-n)m/M) \right]^\star$$

$$= \frac{1}{K} \left| X_m^{(M)} \right|^2 \quad , \quad (C-17)$$

where

$$X_m^{(M)} = \sum_k x_k \exp(-i2\pi km/M) \quad \text{for all } m \quad . \quad (C-18)$$

Now the inverse DFT of $\left\{ \hat{G}_m^{(M)} \right\}$ is

$$\hat{R}_n^{(M)} \equiv \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)} \exp(i2\pi mn/M) = \quad \ldots + \hat{R}_{n-M} + \hat{R}_n + \hat{R}_{n+M} + \ldots \quad (C-19)$$

for all n,

which is an infinitely-aliased version of $\left\{ \hat{R}_n \right\}$. However, if

$$M \geq 2K-1, \quad (C-20)$$

38

there is no overlap of terms in (C-19). Henceforth we presume (C-20) to be true; that is, the size of the transform (C-18) must be at least twice as large as the number of data points. Then there follows from (C-19),

$$\hat{R}_n = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)} \exp(i2\pi nm/M) \quad \text{for } |n| < K \quad .$$

(C-21)

Therefore

$$\sum_n |\hat{R}_n|^2 = \sum_{n=-K+1}^{K-1} |\hat{R}_n|^2 = \sum_{\substack{n:\text{one period} \\ \text{of length } M}} \left|\hat{R}_n^{(M)}\right|^2$$

$$= \sum_{n=0}^{M-1} \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{p=0}^{M-1} \hat{G}_m^{(M)} \hat{G}_p^{(M)} \exp(i2\pi(m-p)n/M) = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)^2} \quad .$$

(C-22)

Then from (C-3) and (C-19),

$$E = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)^2} - \left(\frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)}\right)^2 \quad .$$

(C-23)

Although (C-18) is defined for all relative sizes of M and K, the relation (C-23) holds only if (C-20) holds.

REFERENCES

1. A. H. Nuttall, "Some Techniques for Generation of Random Numbers with Specified Distributions or Densities," NUSC Technical Memorandum 811135, 1 October 1981.

2. Handbook of Mathematical Functions, U. S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series No. 55, U. S. Government Printing Office, June 1964.

3. A. H. Nuttall, "Some Integrals Involving the Q_M-Function," NUSC Technical Report 4755, 15 May 1974.

4. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, Academic Press, Inc., N. Y., 1980.

5. K. D. Senne, "Machine Independent Monte Carlo Evaluation of the Performance of Dynamic Stochastic Systems," Journal of Stochastics, Vol. 1, pp. 215-238, September 1974.

6. A. H. Nuttall, "On Characterization of the Probability Distribution of Measured Data," NUSC Technical Memorandum 811184, 10 December 1981.

7. A. H. Nuttall, "Reconstruction of Power Spectral Estimates," NUSC Technical Memorandum 781051, 6 March 1978.

An Aid in Steepest Descent Evaluation Of Integrals

A. H. Nuttall

ABSTRACT

A simple technique for determining the steepest descent contours of integrands is explained and programmed in BASIC, and then applied to a variety of examples, some of which are very difficult to ascertain analytically, including one with an infinite number of saddle points. The amount of programming is minimal, requiring numerical evaluation only of the derivative of an analytic function.

Table of Contents

List of Illustrations

List of Symbols

z	complex variable
C	contour in z-plane
$g(z)$	amplitude function in integral
$w(z)$	exponent function in integral
λ	large parameter
z_s	saddle point location
x,y	real and imaginary parts of z
u,v	real and imaginary parts of w
∇	gradient operator
$(\ )_b^{1/2}$	particular branch of square root
$(\ )^{1/2}$	principal square root
z_o	zero location

Introduction

The evaluation of contour integrals with analytic integrands is often accomplished by moving the contour to an equivalent one, such that the integrand is better behaved or can be approximated more easily. In particular, movement of the original contour to one that takes advantage of saddle points of the integrand or paths of descent or steepest descent is a very fruitful procedure. It is presumed that the reader is familiar with this technique; see references 1-3, for example. However, one of the difficulties of this procedure is determining the locations of the steepest descent contours (reference 1, p. 263). For complicated integrands, especially those involving branches of multivalued functions, exact determination of steepest descent contours is virtually impossible analytically, and recourse to some type of computer aid is recommended. The procedure given here does not require solution of nonlinear equations, but does give a very good indication of steepest descents with a minimum of analytical and programming effort.

Explanation of Technique

Suppose we wish to evaluate the contour integral

$$\int_C dz \; g(z) \; \exp(\lambda w(z)) \quad , \tag{1}$$

where $g(z)$ and $w(z)$ are analytic functions of z, except for isolated singularities such as poles, essential singularities, and branch points; and C is a contour (finite or infinite) in the complex z-plane. The saddle points of the exponential in (1) occur where (reference 1, p. 258)

$$w'(z_s) = 0 \quad . \tag{2}$$

For λ real, the standard method of determining the paths of steepest descent out of a saddle point is to keep the imaginary part of $w(z)$ constant and equal to its value at z_s (reference 1, p. 255). This generally leads to difficult transcendental equations that must be analytically investigated approximately or solved numerically.

An alternative procedure for finding the steepest descent directions at any point in the z-plane is as follows: Let

$$z = x + iy \quad ,$$
$$w(x + iy) = u + iv \quad . \tag{3}$$

Then the magnitude of the exponential in (1) is $\exp(\lambda u)$, and its direction of steepest descent at x, y is proportional to the negative of the gradient (reference 1, p. 254):

$$\nabla \exp(\lambda u) = \lambda \; \exp(\lambda u) \; \nabla u \quad , \tag{4}$$

where

$$\nabla u = \frac{\partial u}{\partial x} \vec{a}_x + \frac{\partial u}{\partial y} \vec{a}_y \quad , \tag{5}$$

and $\vec{a}_x$ and $\vec{a}_y$ are unit vectors in the positive x- and y-directions, respectively. The explicit evaluation of (5) requires that one analytically evaluate $u = \mathrm{Re}\{w(x + iy)\}$ and then analytically derive $\partial u/\partial x$ and $\partial u/\partial y$. This can be tedious and is liable to human error.

An alternative simpler procedure is possible: By the Cauchy-Riemann conditions applied to a function analytic at z, the derivative

$$w'(z) = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}$$
$$= \frac{\partial v}{\partial y} - i \frac{\partial u}{\partial y} \quad . \tag{6}$$

That is, we can express the desired partial derivatives as

$$\frac{\partial u}{\partial x} = \mathrm{Re}\{w'(z)\}, \quad \frac{\partial u}{\partial y} = -\mathrm{Im}\{w'(z)\} \quad . \tag{7}$$

Then the negative of the gradient in (5) can be written as

$$-\nabla u = -\mathrm{Re}\{w'(z)\}\ \vec{a}_x + \mathrm{Im}\{w'(z)\}\ \vec{a}_y \quad . \qquad (8)$$

So if we evaluate $w'(z)$, the direction of steepest descent at any point z has components in the x, y directions proportional to

$$-\mathrm{Re}\{w'(z)\}, \ \mathrm{Im}\{w'(z)\} \qquad (9)$$

for $w'(z) \neq 0$. The only analytical calculation necessary is that of derivative $w'(z)$, a task generally easily accomplished, and indeed necessary for evaluation of saddle point locations anyway. A computer program can then be written to numerically evaluate $w'(z)$ in terms of its components (9) at all points of interest in the z-plane. A program for this procedure is given in appendix A, along with the specific examples displayed later in this report.

Presentation of this steepest descent information for human interpretation is accomplished here by drawing a short standard-size line through each point $z = x + iy$, centered on the point and with an arrowhead pointing in the direction of steepest descent. The magnitude of the rate of steepest descent is discarded; only the direction is preserved. How effective this procedure is will be demonstrated by the following examples.

Examples

Airy Function

The Airy function, Ai, is proportional to (1) when $g(z) = 1$,

$$w(z) = z - \frac{1}{3}z^3 \quad , \tag{10}$$

and C is an infinite contour starting anywhere in the angular range $-5\pi/6 < \arg(z) < -\pi/2$ and ending in $\pi/2 < \arg(z) < 5\pi/6$; see reference 1, pp. 52 and 266. Then

$$w'(z) = 1 - z^2 \quad . \tag{11}$$

A computer need only evaluate the complex product $z \cdot z$ and subtract it from 1 in order for (11) to be used in (9). An example of this procedure is given in figure 1. The arrows clearly indicate the steepest descent paths from any point in the z-plane. The two solutions of (11) equal to zero, namely, saddle points $z_s = \pm 1$, have arrows pointing both inward and outward at these points, reflecting the very nature of a saddle point. Movement of the original contour C to the steepest descent contours (solid lines) out of the saddle point at $z_s = -1$ is easily accomplished; no singularities of the integrand of (1) are crossed in the movement process.

The Airy function for complex argument (reference 1, section 7.3) has, more generally,

$$w(z) = z \exp(i\vartheta) - \frac{1}{3}z^3 \quad . \tag{12}$$

Figure 1 corresponded to $\theta = 0$. For $\theta \neq 0$, the determination of steepest descent paths is analytically difficult (ibid.). However, since

$$w'(z) = \exp(i\theta) - z^2 \quad , \tag{13}$$

computer evaluation of (13) and (9) is trivial. The steepest descent directions for $\theta = 3\pi/4$, for example, are depicted in figure 2. The solid lines in the neighborhood of the saddle points were hand-drawn upon observation of the descent arrows. The interpretation of figure 2 is much easier than its counterpart in reference 1, figure 7.3.3. Also the determination of an equivalent contour to C is easily achieved by reference to figure 2. First let the new contour come from $\infty \exp(-i2\pi/3)$ along a steepest ascent to the saddle point at $z_s = \exp(-i5\pi/8)$. Then let it continue in a northeasterly direction along the steepest descent direction out of this saddle point to the y-axis. Next proceed due north to a point near $z = i$ and then follow a steepest descent path out to $\infty \exp(i2\pi/3)$. The vertical portion of this new contour is not a path of steepest descent, but it is obviously a path of descent because the projections of the arrows on this vertical section all point in the upward direction of travel. This discussion also points out that the other saddle point at $z_s = \exp(i3\pi/8)$ does not enter into the asymptotic development of Ai, at least for this value of θ.

An alternative equivalent contour to C is the pair of steepest descent contours passing through the saddle points and connecting $\infty \exp(-i2\pi/3)$ to $+\infty$ and $+\infty$ to $\infty \exp(i2\pi/3)$, respectively; see figure 2. Observe that the movement of C to this new

pair of paths is not an approximation; it is an alternative exact representation of the original integral. If one now approximates these two path integrals by their contributions near their peaks at the saddle points, the saddle point at $z_s = \exp(i3\pi/8)$ will yield exponentially small contributions relative to that at $z_s = \exp(-i5\pi/8)$. This can be seen from figure 2 by drawing a straight line between the two saddle points; all projections of arrows along this line point at the upper-right saddle point, meaning that the value of $|\exp(\lambda w(z))|$ is smaller there.
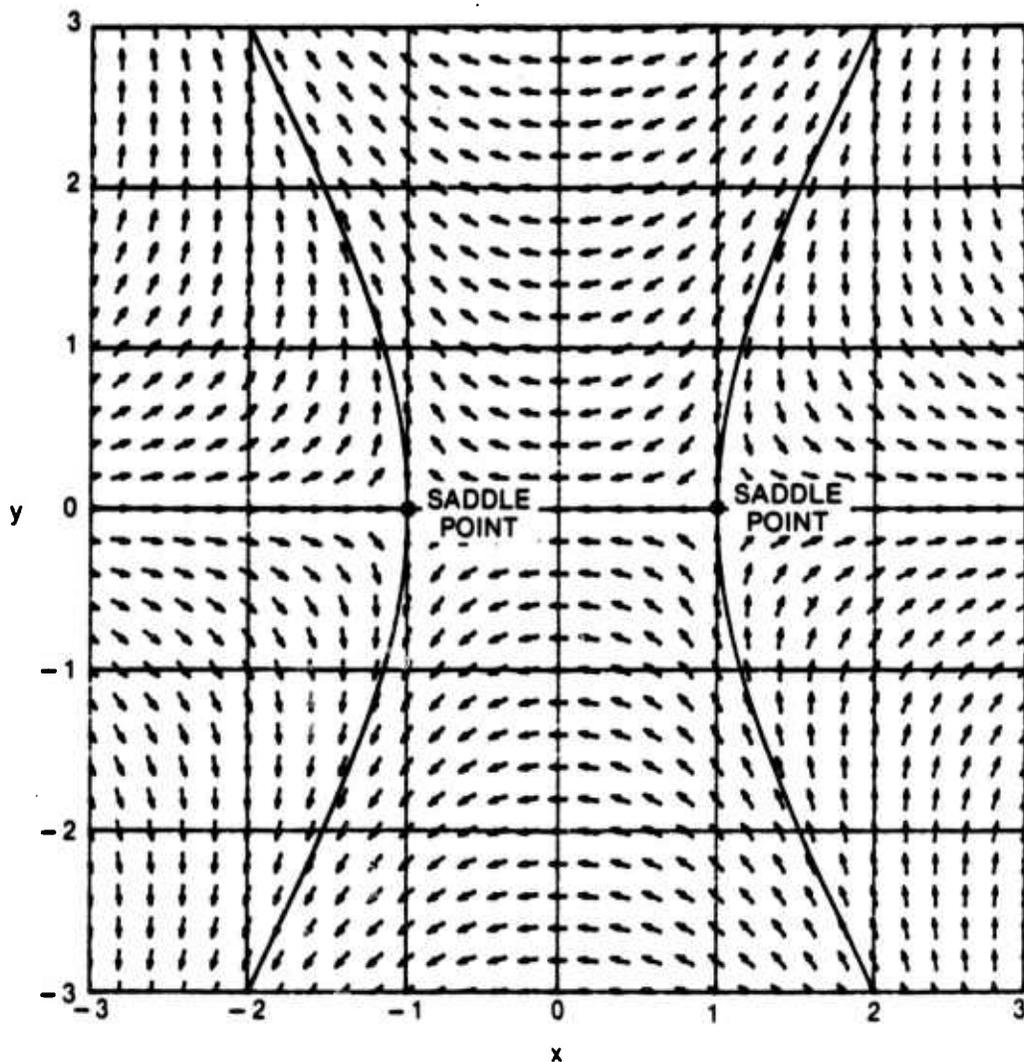

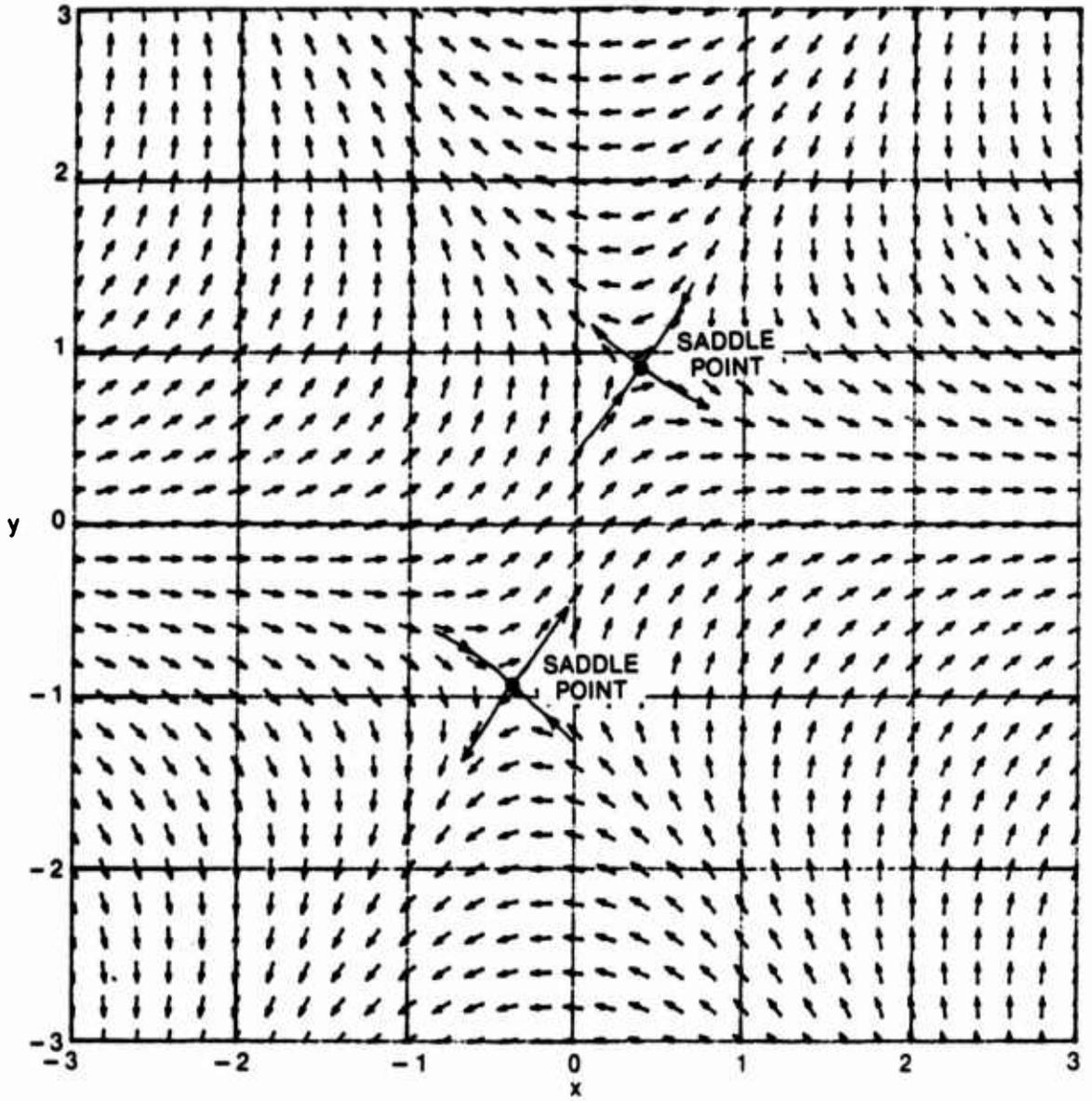**Figure 1. Steepest Descent Directions for Airy Function, $\theta = 0$**

Figure 2. Steepest Descent Directions for Airy Function, $\theta = 3\pi/4$

When we continue on to the case $\theta = \pi$ in (12), the character of steepest descents is as depicted in figure 3. Now the straight line connecting the two saddle points has all arrows perpendicular to it; thus this vertical line is a contour of constant $|\exp(\lambda w(z))|$. This means that both saddle points contribute equally to the value of integral (1). Again the pair of steepest descent contours through the saddle points (mentioned in the above paragraph) represent exactly the original integral; one could evaluate the original integral exactly by adding the total contributions of both of these paths, or an approximation can be achieved by computing the integrand near its peaks at the saddle points.
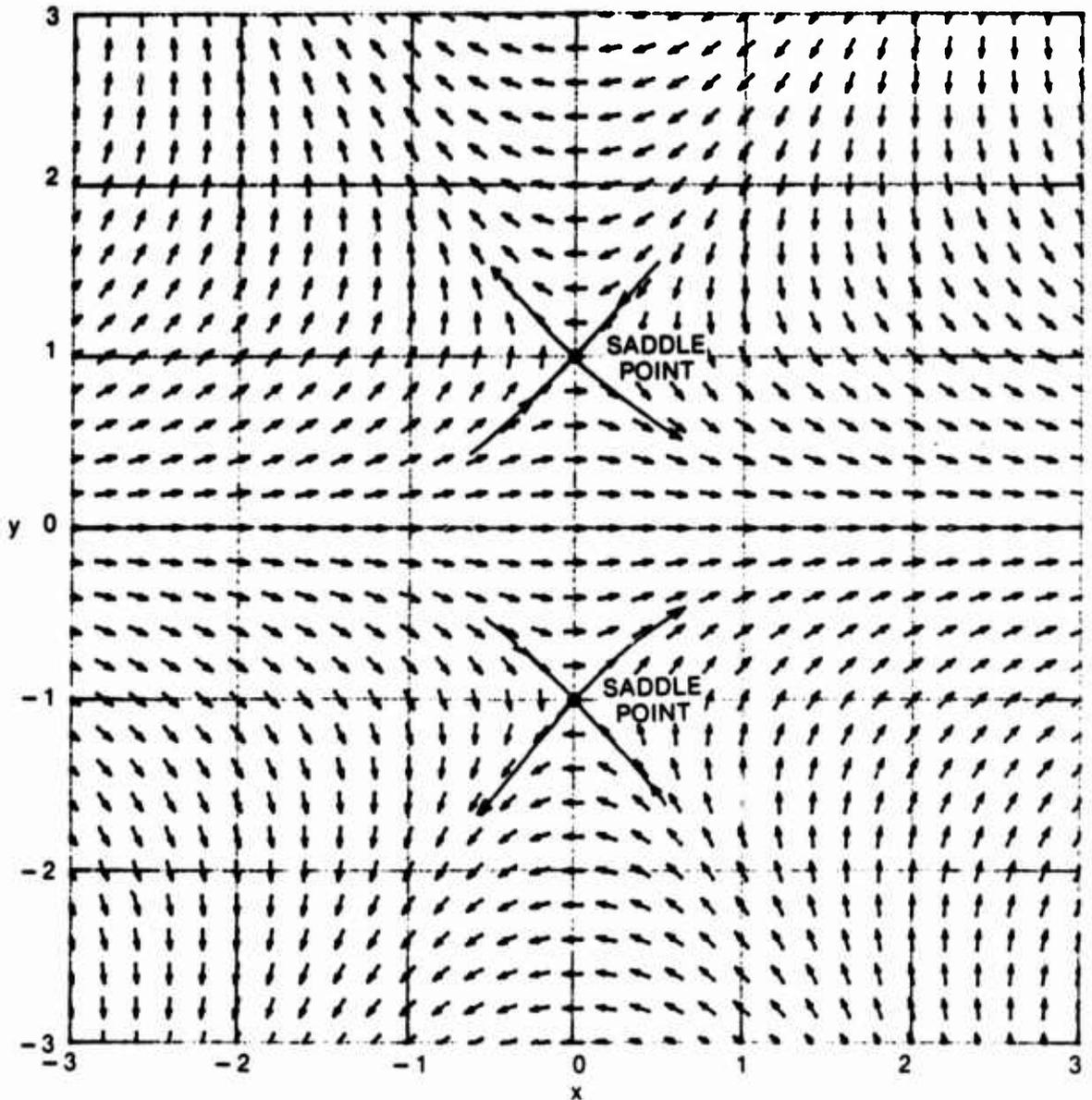

Figure 3. Steepest Descent Directions for Airy Function, $\theta = \pi$

Hankel Function

The Hankel function takes the form (reference 1, 7.2.23)

$$w(z) = i[\cos(z) + \beta(z - \frac{\pi}{2})] , \qquad (14)$$

where β is a constant. There follows immediately

$$w'(z) = i[- \sin(z) + \beta] . \qquad (15)$$

Computer evaluation of (15) requires only a trigonometric sin of a complex number, followed by subtraction and multiplication. Sample descent direction plots for $\beta = 0.5$, 1, and 1.5 are given in figures 4-6, respectively. (It is informative to compare these figures with figures 7.2.1-7.2.4 in reference 1.) Movement to equivalent contours is obvious from figures 4-6. Since (15) has period 2π in x, only a 2π strip has been plotted in figures 4-6. The character of the steepest descents in figure 5 for $\beta = 1$ is different, in that the saddle points have coalesced; however, there is no difficulty ascertaining from the plots what new contour to adopt.

Klein-Gordon Equation

This example is complicated by the presence of branch lines in the exponent function w(z). Specifically we have (reference 1, 7.5.9)

$$w(z) = i\left[\theta(z^2 - 1)_b^{1/2} - z\right] , \qquad (16)$$

where the branch of the square root is taken as positive real for $z = x > 1$, and with branch lines extending vertically downward from the branch points at $z = \pm 1$. The derivative of (16) is

$$w'(z) = i\left[\frac{\theta z}{\left(z^2 - 1\right)_b^{1/2}} - 1\right] , \qquad (17)$$

where the same square root branch must be taken as in (16).

If one has available a computer program that evaluates the principal square root of a complex number, denoted here by $z^{\cdot}$, it can be used to evaluate (17) in the following manner. Observe first that the branch line of principal square root $z^{\cdot}$ occurs where $z = -p$ for $p \geqslant 0$; i.e., p can take on all nonnegative real values. So consider the representation

$$\left(z^2 - 1\right)_b^{1/2} = i(-iz + i)^{1/2} (-iz - i)^{1/2} . \qquad (18)$$

For z real, positive, and large, the right-hand side of (18) approaches $i[\exp(-i\pi/4)z^{\cdot}] [\exp(-i\pi/4)z^{\cdot}] = z$, as desired. Furthermore, the two branch lines of (18) occur where the arguments of the two principal square roots have values

$$-iz \pm i = -p \text{ for } p \geq 0 ;$$

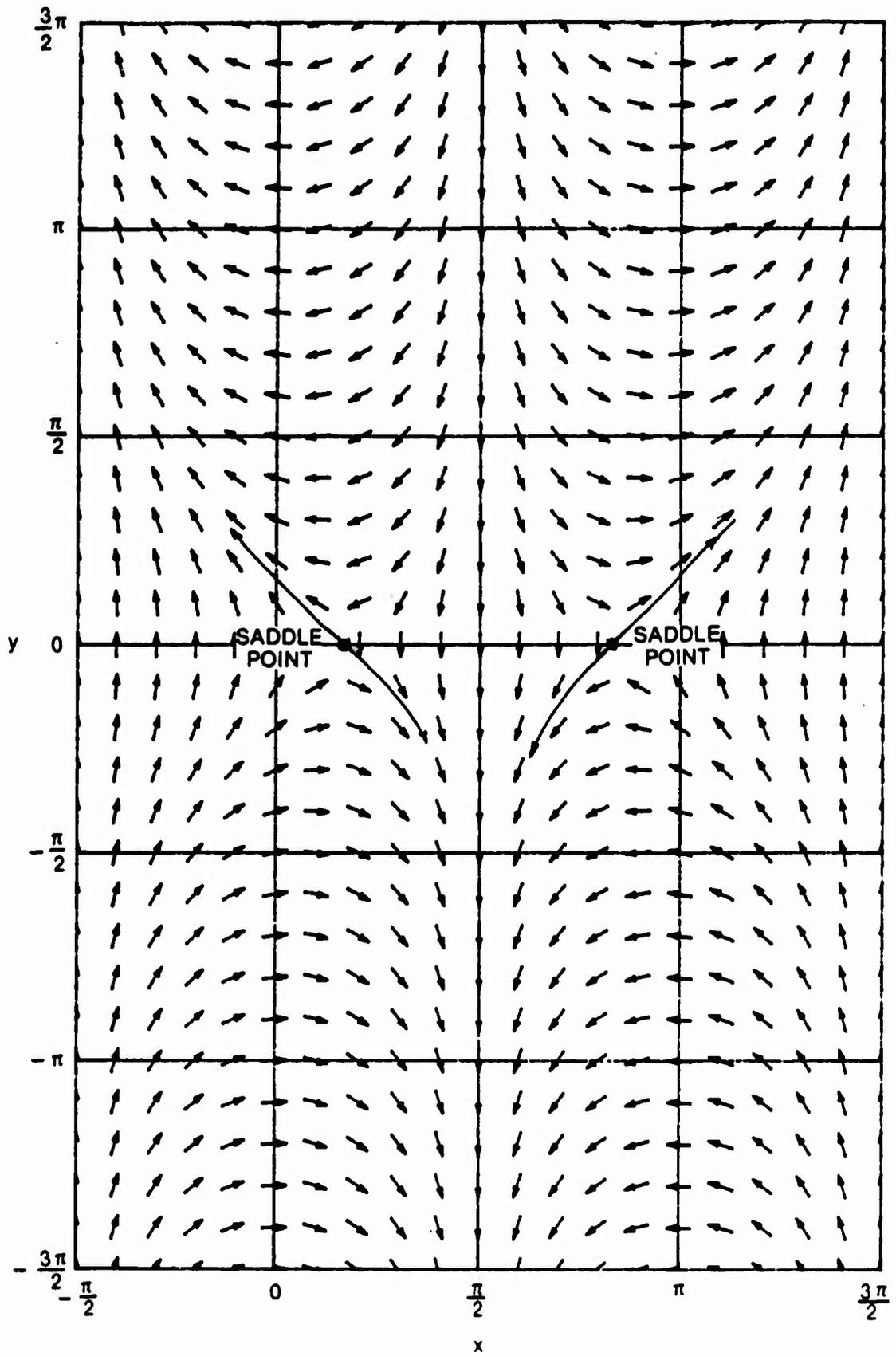$$i.e., z = \pm 1 - ip \text{ for } p \geq 0 . \qquad (19)$$

Figure 4. Steepest Descent Directions for Hankel Function, $\beta = 0.5$
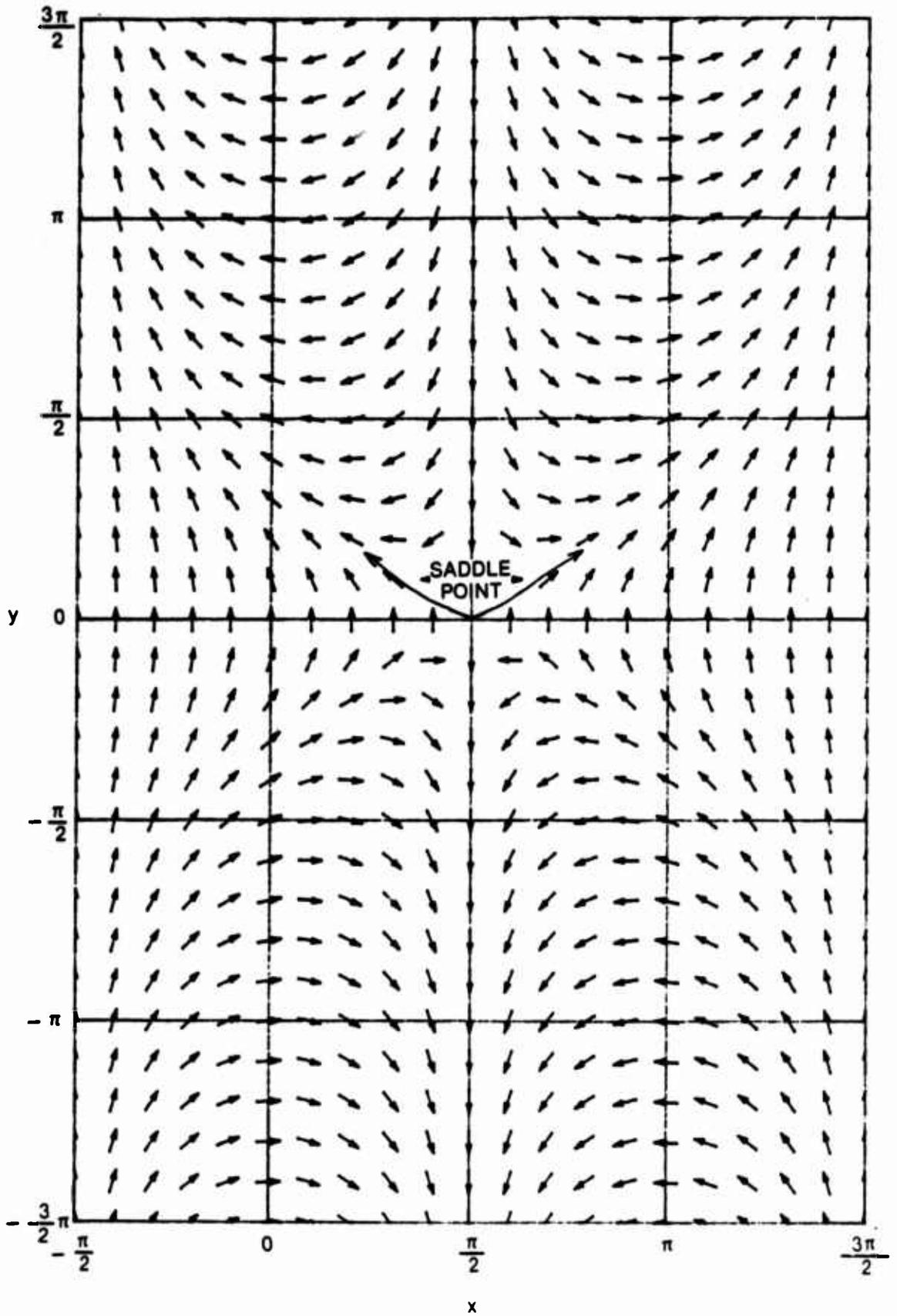
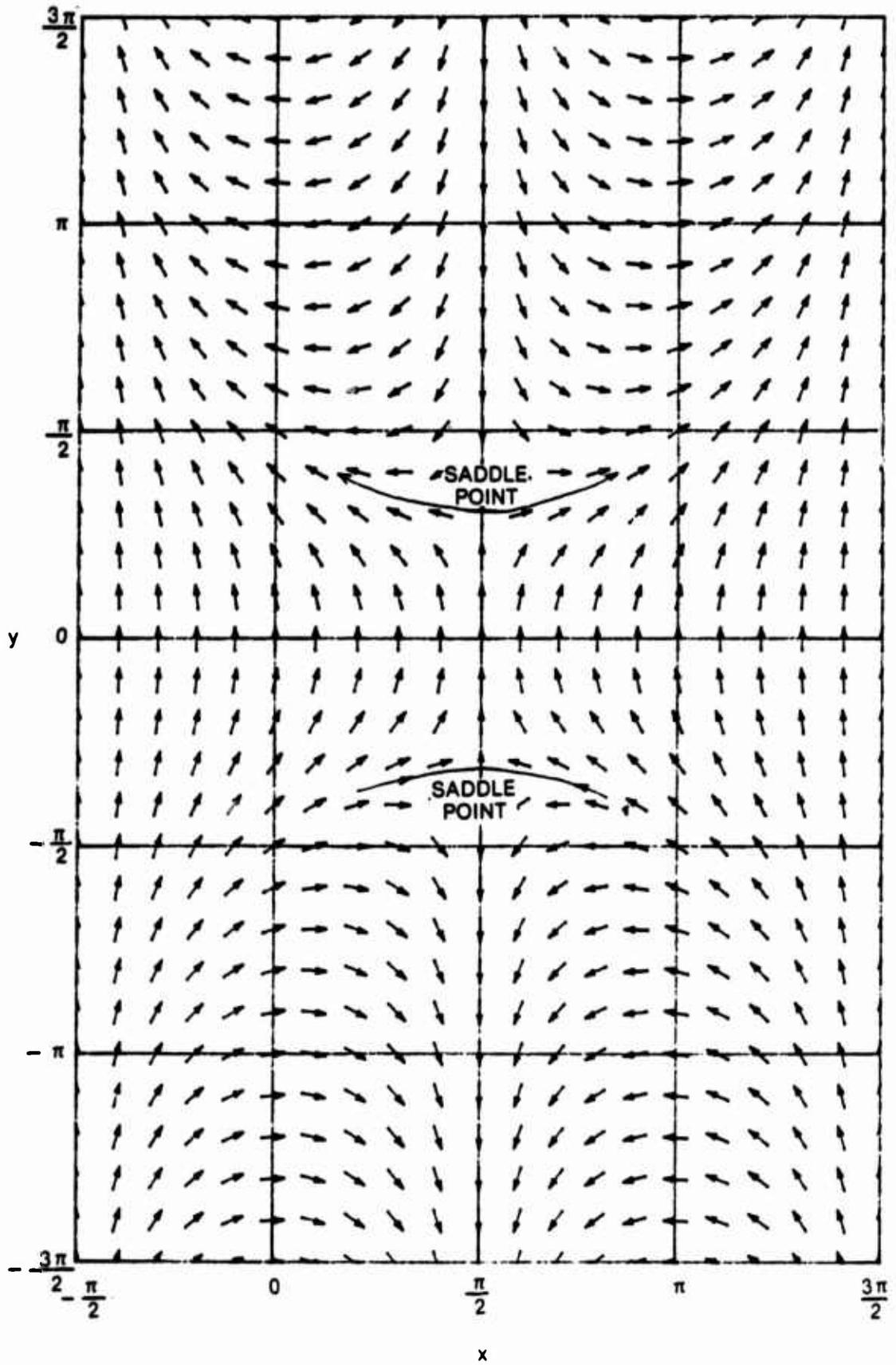Figure 5. Steepest Descent Directions for Hankel Function, $\beta = 1$

Figure 6. Steepest Descent Directions for Hankel Function, $\beta = 1.5$

These are vertically downward from $z = \pm 1$, as desired. Thus (17) can be easily evaluated by taking the two complex principal square roots indicated in (18) and performing multiplication, division, and addition of complex numbers. The specific coding is illustrated in appendix A.

The pole of $g(z)$ at $z = v_0$ for this example (reference 1, 7.5.8) has no effect on the steepest descent contours of $w(z)$. The steepest descent directions for $\theta = 0.8$ are depicted in figure 7. There are saddle points at $z_s = \pm 5/3$, and the steepest descent contours go vertically downward eventually. The steepest descent directions near the branch lines emanate from the branch lines themselves, but these branch lines have no effect on the steepest descent contours through the saddle points. However, if the branch line emanating from the branch point at $z = 1$ had been taken at angle $-\pi/6$, for example, it would have interfered with the steepest descent contours in the 4th quadrant of the z-plane. Such a choice of branch for the square root in (16) and (17) is undesirable and should be avoided, as was done in figure 7.

Function with Essential Singularity

This integrand is characterized by (reference 4)

$$\exp(w(z)) = z \, \exp\left(\frac{-K}{z - 1}\right) , \tag{20}$$

which function has a zero at $z = 0$ and an essential singularity at $z = 1$. Then

$$w(z) = \ln(z) - \frac{K}{z - 1} ,$$

$$w'(z) = \frac{1}{z} + \frac{K}{(z - 1)^2} . \tag{21}$$

The steepest descent contours for (21) are depicted in figure 8 for $K = 3$. The essential singularity generates a "dipole effect" about $z = 1$, i.e., 0 at $z = 1+$, and ∞ at $z = 1-$, for $K > 0$. The zero of (20) at $z = 0$ manifests itself as a point toward which all the arrows point, since zero is the smallest magnitude that any complex function can take on. The saddle points (roots of (21)) occur at

$$z_s = \exp(\pm i\theta), \text{ where } \theta = \text{arc } \cos\left(1 - \frac{K}{2}\right), \text{ for } 0 \leq K \leq 4 . \tag{22}$$

Utilizing the information in figure 8, we find it relatively easy to decide what the effect of moving an original contour around in the z-plane will do to integral (1). Movement across the essential singularity at $z = 1$ will necessitate consideration of the residue at this point. The zero and saddle points of (20) are not points of singularity.

Cubic Function

This example comes from reference 3, pp. 296-302; it is characterized by $g(z) = \frac{1}{2}$ and (ibid., upper line of 6.6.25)

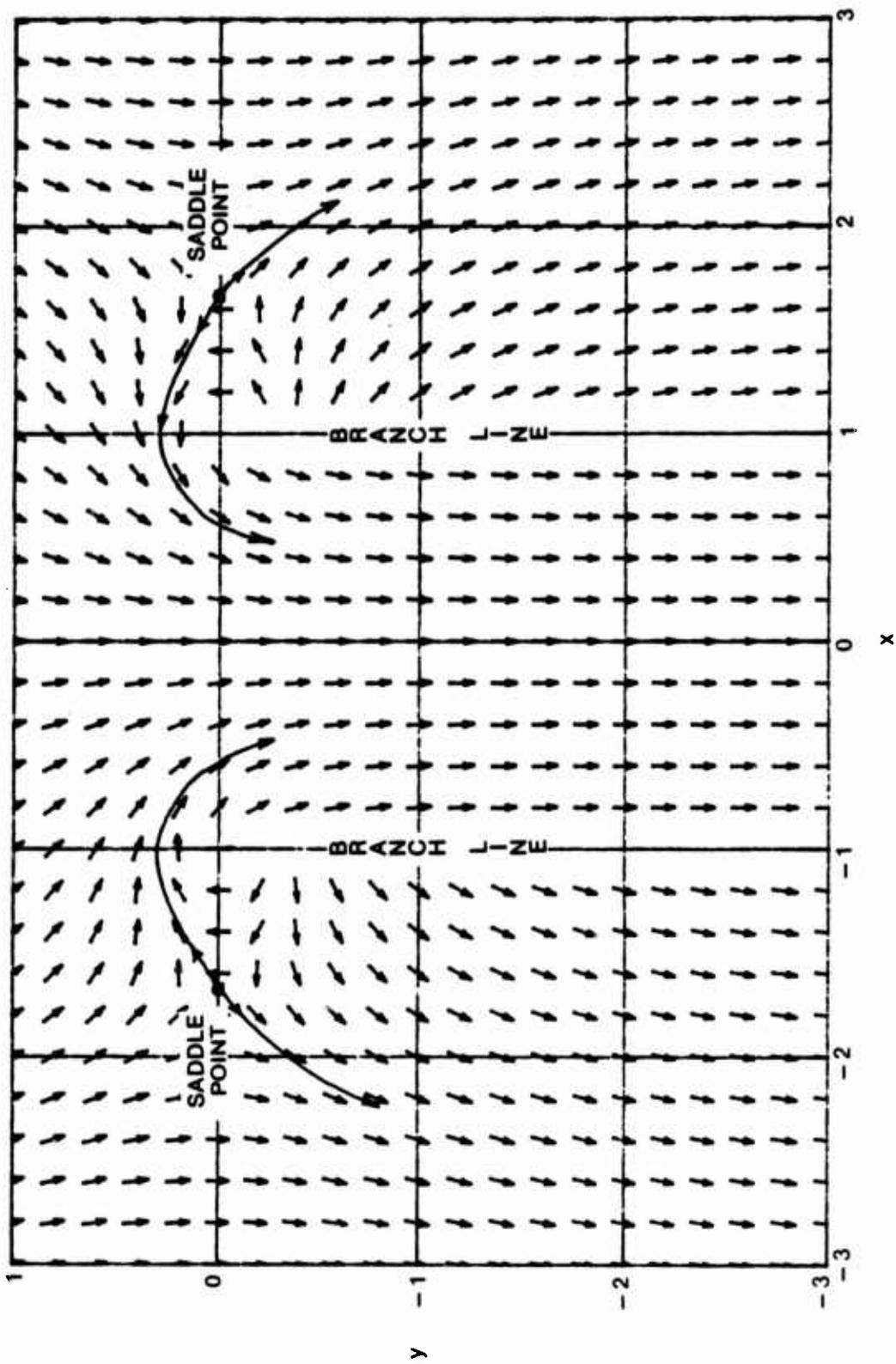$$w(z) = -e^{i\omega}\left(4z^2 - i5z + iz^3\right) . \tag{23}$$

Figure 7. Steepest Descent Directions for Klein-Gordon Equation, $\theta = 0.8$
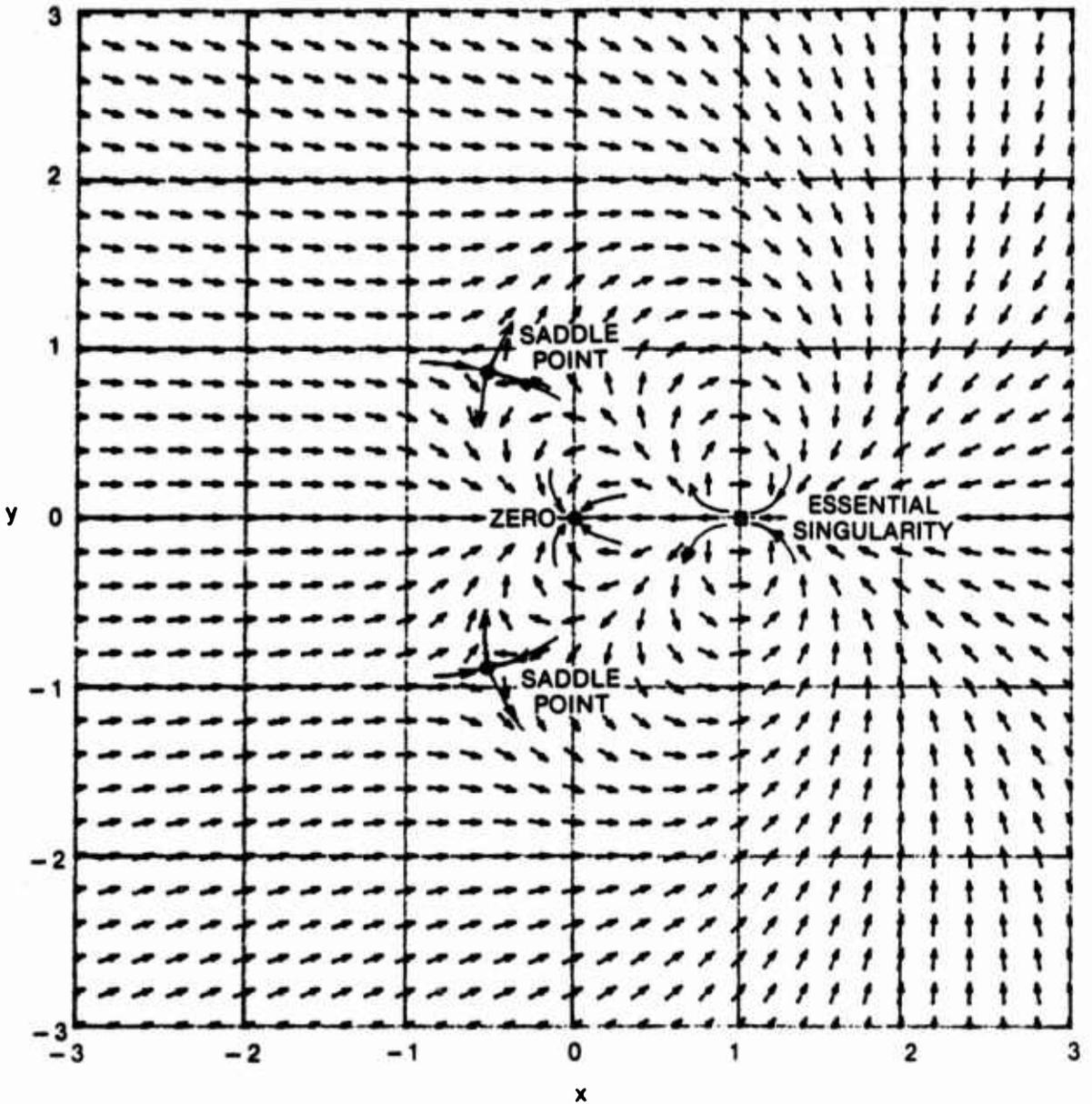
Figure 8. Steepest Descent Directions for $\exp(w(z)) = z \exp\left(\dfrac{-K}{z-1}\right)$, $K = 3$

Also now the integral is a finite one, from $z = -1$ to $z = +1$. This example exhibits a Stokes phenomenon at certain values of α, where we have represented $\lambda = |\lambda|e^{i\alpha}$. We find

$$w'(z) = -e^{i\alpha}\left(8z - i5 + i3z^2\right),\qquad(24)$$

which has zeros (saddle points) at $z_s = i$ and $i5/3$.

A plot of steepest descent directions for $\alpha = 0$ (positive real λ) is given in figure 9. It indicates that the steepest descent contours out of the limits at $z = -1$, $+1$ tend to $\infty\exp(-i5\pi/6)$ and $\infty\exp(-i\pi/6)$, respectively. But these two valleys at ∞ can be

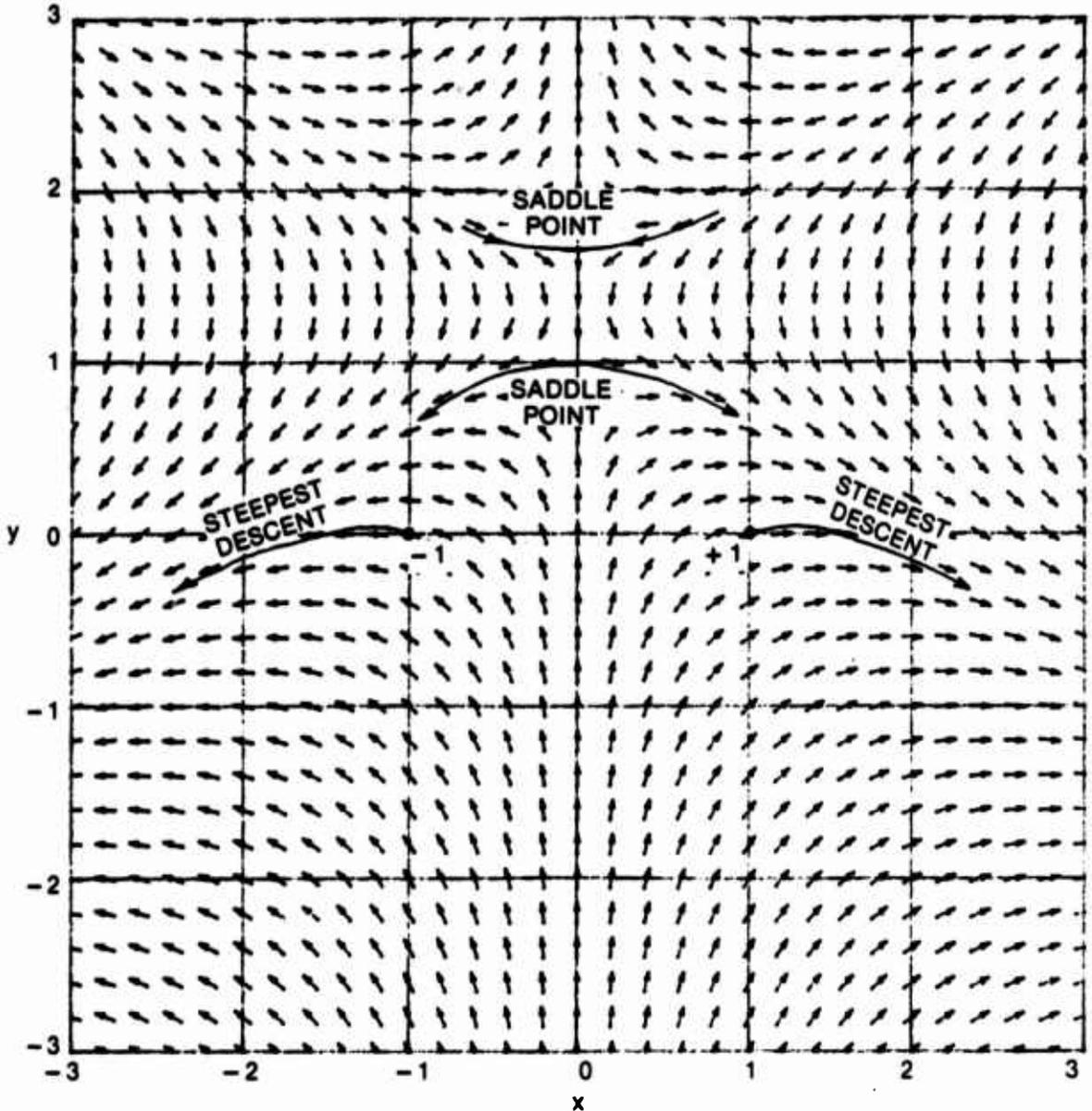**Figure 9. Steepest Descent Directions for Cubic, $\alpha = 0$**

joined by the saddle point contribution through the point $z_s = i$. Thus the integral over $(-1, 1)$ is exactly equal to the sum of these three steepest descent contours. The saddle point at $z_s = i5/3$ need not be considered. The dominant contribution is obviously that at the saddle point $z_s = i$, as may be seen by the arrow directions.

For $\alpha = 5\pi/12$, figure 10 indicates a similar behavior. Since the steepest descent contours out of -1, $+1$ tends to $-175°$ and $-55°$, respectively, the saddle point at $z_s = i$ must again be used to join them. The dominant contribution is seen to be due to $z = -1$, by making use of the arrows of descent in this figure.

For $\alpha = 3\pi/4$, however, we see from figure 11 that both of the steepest descent contours out of -1, $+1$ tends to $\infty\exp(-i5\pi/12)$. Now there is no need to employ

either of the saddle points. The integral over $(-1,1)$ is given exactly by the sum of the two steepest descent contributions. The dominant contribution is again due to $z = -1$, since we have to descend from $z = -1$ to near $z = 0$ to reach magnitude values comparable to those at $z = +1$.
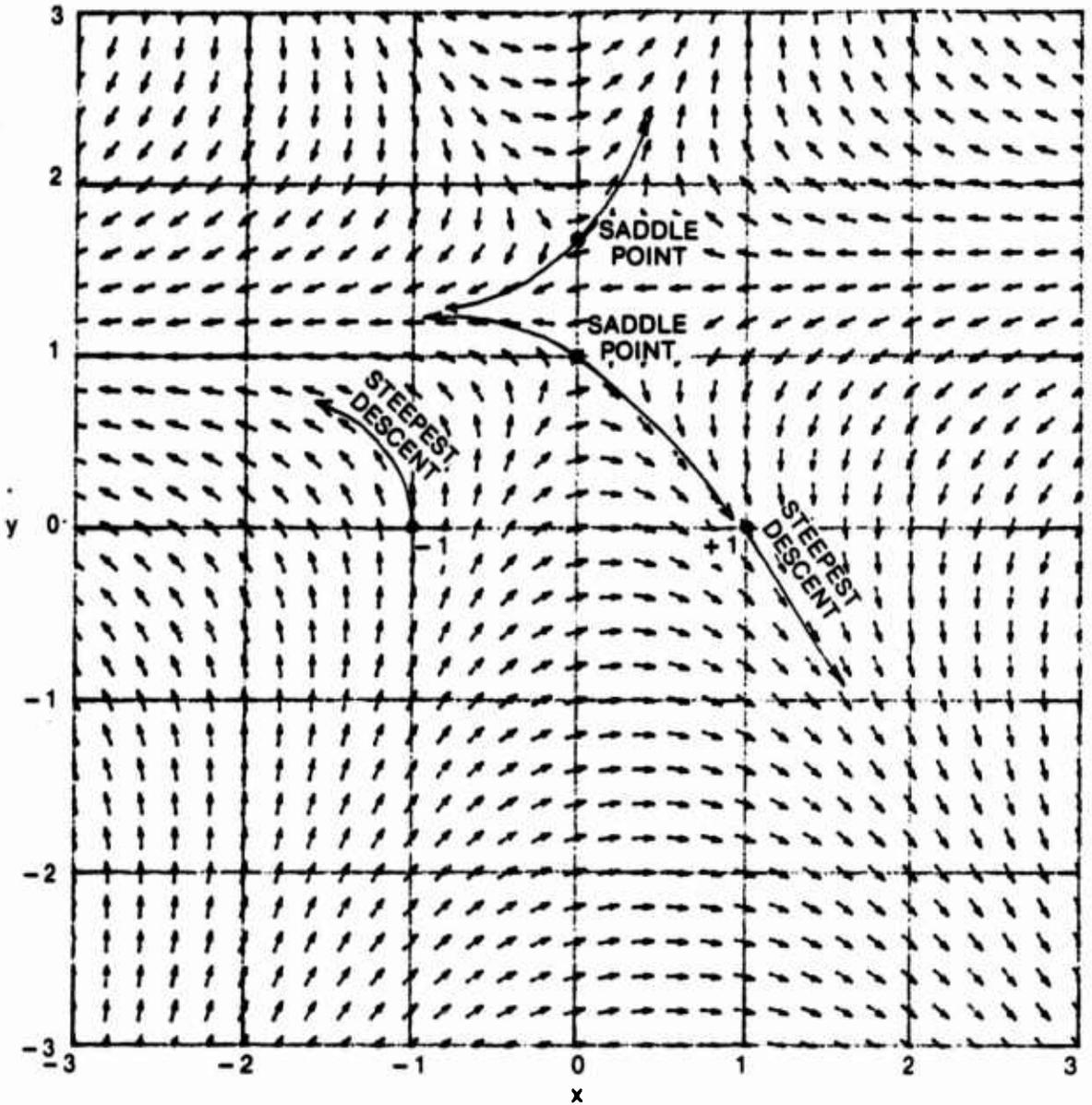
Figure 10. Steepest Descent Directions for Cubic, $\alpha = 5\pi/12$
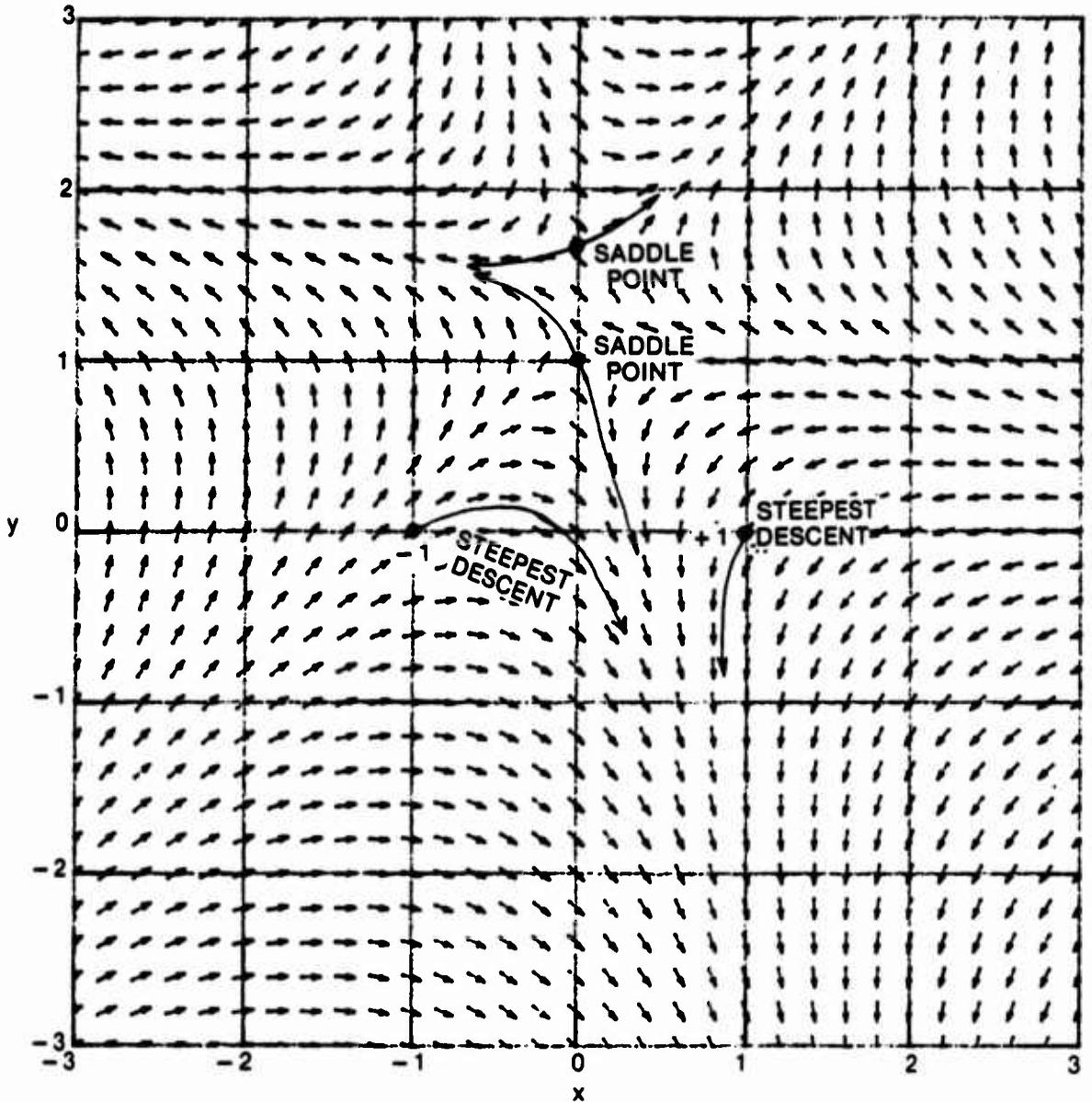
Figure 11. Steepest Descent Directions for Cubic, $\alpha = 3\pi/4$

Gaussian Exponent

The characteristic function of a particular type of impulsive noise is given in reference 5, equation (5), in the form

$$f(\zeta) = \exp(-A) \sum_{m=0}^{n} \frac{A^m}{m!} \exp\left(- \frac{1}{2} \frac{\overline{mB_o^2}}{2} \zeta^2\right) \tag{25}$$

for a purely Poisson process (no Gaussian background). This summation can be evaluated in the closed form

$$f(\xi) = \exp\left(A \exp\left(-\frac{1}{2} \frac{\overline{B_o^2}}{2} \xi^2\right) - A\right) . \tag{26}$$

The corresponding probability density function of this impulsive noise is given by the Fourier transform

$$p(v) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\xi \exp\left(-i\xi v + A \exp\left(-\frac{1}{2} \frac{\overline{B_o^2}}{2} \xi^2\right) - A\right) . \tag{27}$$

By expanding $f(\xi)$ in (26) in a power series, we note that the mean square value of the Poisson process is readily found to be $A\overline{B_o^2}/2$. Since we will be interested in values for the dimensionless parameter A of the order of 1 (e.g., A = 0.35 in reference 5, figure 3), we will normalize our random variable according to $t = v/(\overline{B_o^2}/2)^{\frac{1}{2}} \equiv v/\sigma_o$. Also, as, $\xi \to \pm\infty$, $f(\xi)$ in (26) tends to nonzero value exp(-A). Adding and subtracting this quantity and letting $\xi = z/\sigma_o$ enables (27) to be expressed as

$$p(v) = \exp(-A) \delta(v)$$

$$+ \frac{\exp(-A)}{2\pi\sigma_o} \int_{-\infty}^{+\infty} dz \exp(-itz)\left[\exp(A \exp(-z^2/2)) - 1\right] . \tag{28}$$

Although there is no obvious parameter λ in this form, it is shown in appendix B that we can still use steepest descent procedures on the logarithm of the integrand of (28); i.e., here we have

$$w(z) = -itz + \ln[\exp(a(z)) - 1] , \tag{29}$$

where we have defined

$$a(z) = A \exp(-z^2/2) . \tag{30}$$

Observe that as $z \to \infty$ with arg(z) in the two sectors within $\pi/4$ of the positive-real or negative-real axes, a(z) becomes very small and

$$w(z) \sim -itz + \ln[a(z)] = -itz + \ln A - z^2/2 ; \tag{31}$$

this means that Re $w(z) \to -\infty$ in these two sectors of the z-plane, and, therefore, the integrand of (28) tends to zero in these sectors as $z \to \infty$.

The integrand of the integral of interest in (28) is zero in the finite z-plane only when

$$h(z_o) \equiv \exp\left(Ae^{-z_o^2/2}\right) - 1 = 0 ,$$

$$Ae^{-z_o^2/2} = i2\pi n \text{ for } n \neq 0 ,$$

$$- \frac{z_0^2}{2} = \ln\left(i \frac{2\pi n}{A}\right) + i2\pi m = \ln\left(\frac{2\pi |n|}{A}\right) + i\frac{\pi}{2} \text{sgn}(n) + i2\pi m \quad ,$$

$$z_0 = \pm i \left[2 \ln\left(\frac{2\pi|n|}{A}\right) + i\pi(4m + \text{sgn}(n)) \right]^{1/2} \quad , \tag{32}$$

where $A > 0$, the square root is the principal branch, and n and m are arbitrary integers (negative, zero, or positive), except that $n \neq 0$. These zeros of $h(z)$ are important because they will be locations toward which all the steepest descent arrows must point in their neighborhoods, since zero is the smallest magnitude that a complex function can take on. These zero locations, of which there are an infinite number, depend only on A, and not on normalized variable $t = v/\sigma_0$ in expression (28) for the probability density $p(v)$. The locations of the zeros of $h(z)$ in the third quadrant closest to the origin are depicted in figure 12 for $A = 0.35$. There is symmetry in the other quadrants since (from the first line of (32))

$$h(-z) = h(z), \quad h(z^*) = h^*(z) \quad . \tag{33}$$

For purposes of evaluating steepest descent directions and saddle point locations, we note that $a(z)$ in (30) has the property

$$a'(z) = -za(z) \quad , \tag{34}$$

and so (29) yields

$$w'(z) = -it - \frac{z\, a(z)}{1 - \exp(-a(z))} \quad . \tag{35}$$

Thus the saddle points, z_s, of which there are an infinite number, are solutions of

$$\frac{z_s\, a(z_s)}{1 - \exp(-a(z_s))} = -it \quad , \tag{36}$$

and obviously depend on both A and t. The most important saddle point is at $z_s = -i\beta$, β positive real, where (using (30)),

$$\frac{iA\, \exp(\beta^2/2)}{1 - \exp(-A\, \exp(\beta^2/2))} = t \quad . \tag{37}$$

The steepest descent directions for $A = 0.35$, $t = 2$ are depicted for (35) in figure 13. The saddle point at $z_s = -i\beta = -i1.341$ satisfies (37); there are five other saddle points indicated by ✗ in the figure, in addition to the zeros carried over from figure 12. The steepest descent contour out of the saddle point $z_s = -i1.341$ is drawn as a solid line; it is asymptotic to $y = -t$ as $z \to \infty$ with $\arg(z)$ within $\pi/4$ of the positive- and negative-real axes. Movement of the original contour from the real axis in (28) to the steepest descent contour is easily justified.
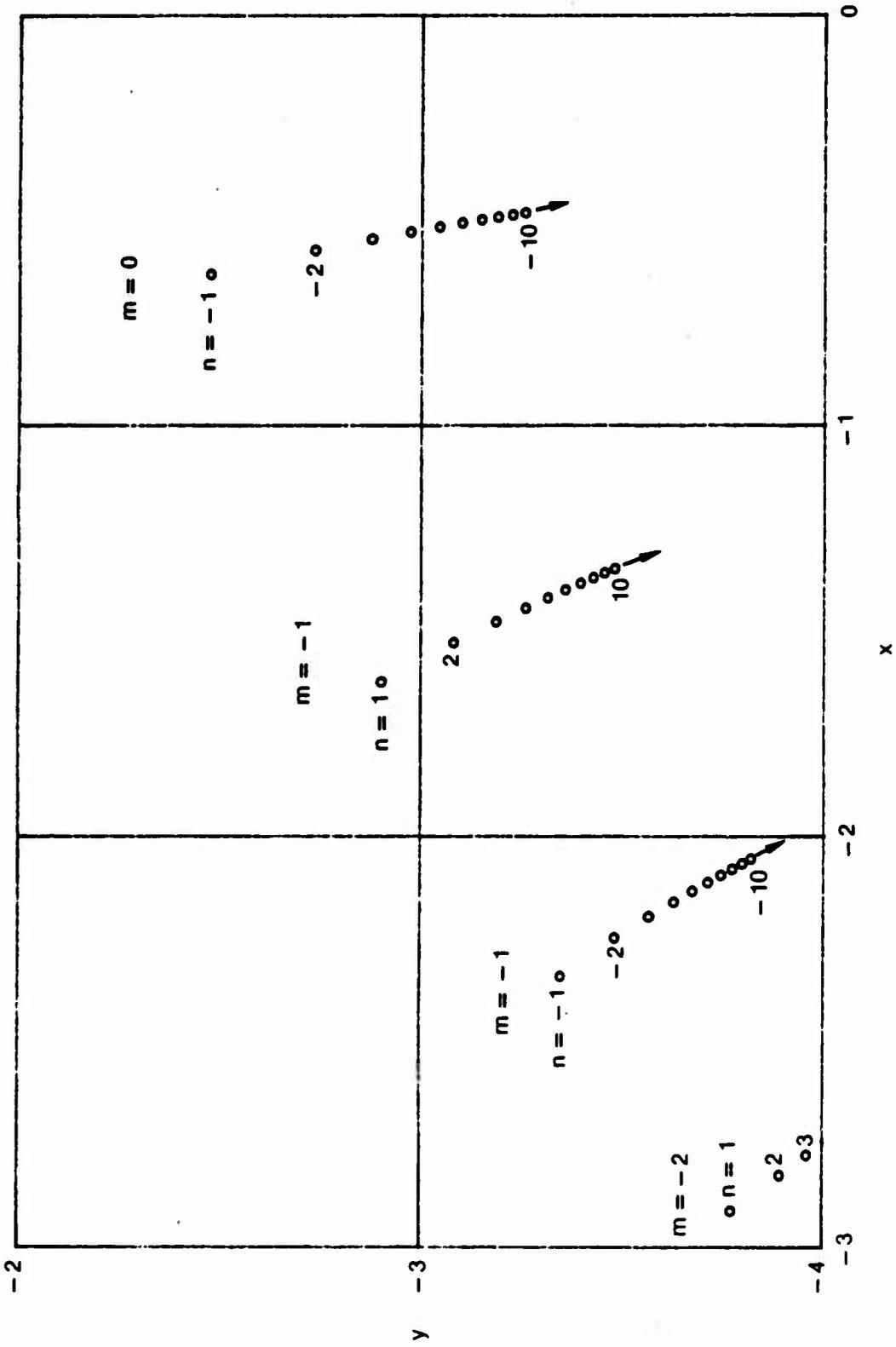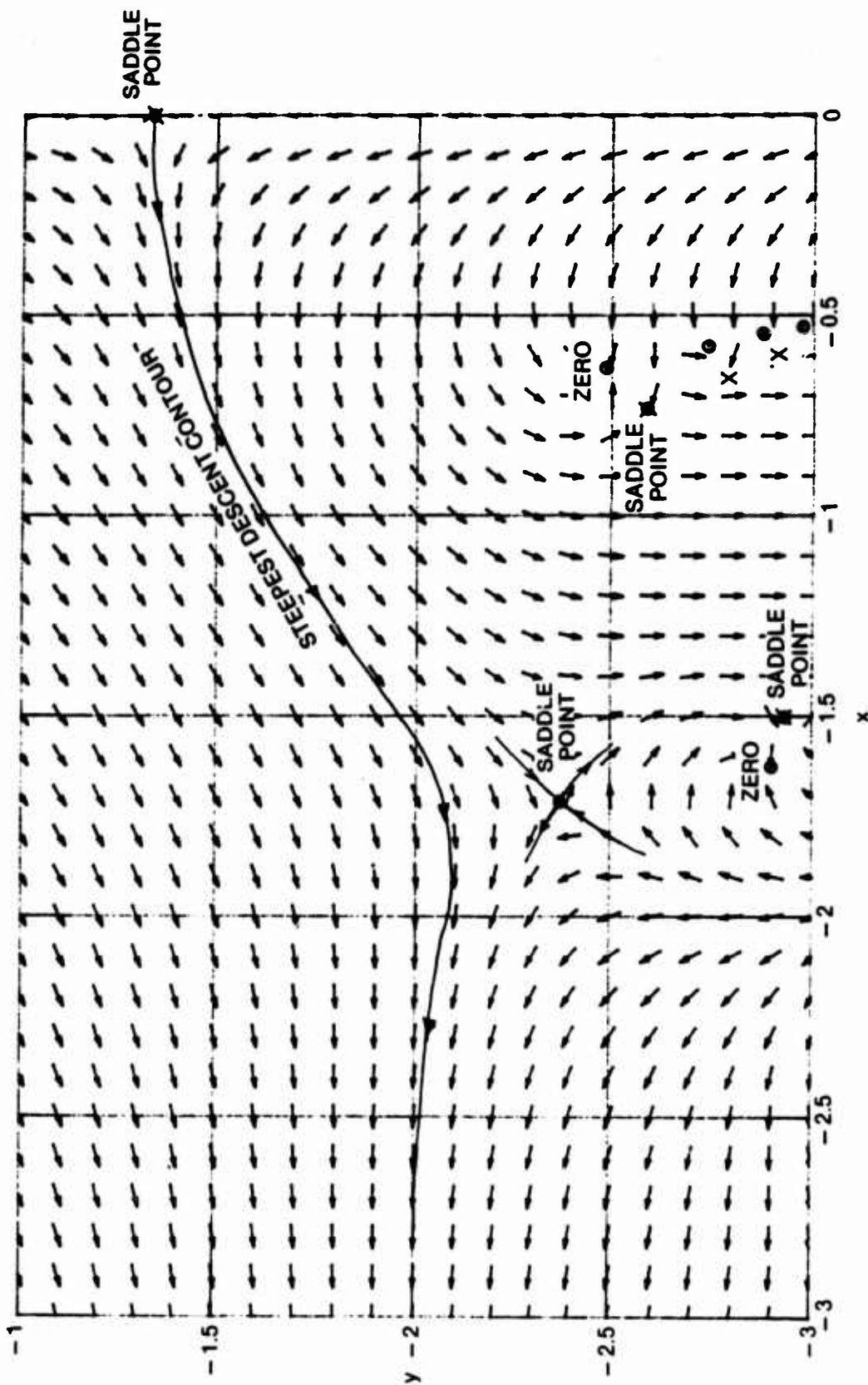
Figure 12. Zero Locations for A = 0.35

Figure 13. Steepest Descent Contour for A = 0.35, 1 = 2

To deduce the asymptotic nature of the steepest descent contours, recall (31); thus the imaginary part of w is

$$v \sim -tx - xy = -x(y + t) \text{ as } z \to \infty \tag{38}$$

in the sectors under consideration. But, since from (29),

$$w(-i\beta) = -\beta t + \ln\left[\exp(A \exp(\beta^2/2)) - 1\right] \tag{39}$$

is real, we require (38) to approach 0 as $x \to \pm\infty$. This requires that $y \to -t$, as claimed.

We observe from figure 13 that no use is made of the saddle point at (-1.71, -2.38), nor of the infinite number of other saddle points. We also observe that the descent directions in the neighborhood of the closely spaced zeros and saddle points near the bottom of the figure is very detailed and complicated; however, none of that information is needed.

When t is increased to 5, figure 14 applies. Now the steepest descent contour out of the saddle point at $z_s = -i\beta = -i1.945$ heads into the zero at (-0.632, -2.485). How to connect from this latter point to $z = -\infty$ is not clear. Instead, we consider a horizontal descent contour out of $z_s = -i1.945$ until we get in the neighborhood of (-2, -2), and then we resume a steepest descent contour heading toward $y = -t = -5$. The major contribution to this descent contour is given by the neighborhood of the saddle point at $x = 0$; the descent and steepest descent contours are tangent at this saddle point.

The aid afforded by the steepest descent directions depicted in figures 13 and 14 is extremely worthwhile, since the exponential in (29) and (35) makes an analytical approach very difficult. The ability to discard or avoid certain regions of the z-plane in determining an appropriate descent contour is rather obvious from the figures when coupled with basic information about the integrand, like the asymptotic behavior of the steepest descent contour.

The above results yield exact values for the original integral, since we have simply determined contours equivalent to those originally specified. Now we will derive an asymptotic expansion for the probability density in (28) for large $t = v/\sigma_0$, i.e., at values v much larger than the standard deviation of the Poisson process.

To do this, we need

$$w''(z) = a\left(1 - e^{-a}\right)^{-2}\left[z^2 - 1 + e^{-a}\left(1 - z^2 - z^2 a\right)\right] , \tag{40}$$

where $a = a(z)$ is defined in (30). Then we find

$$w''(z_s) = t^2 + i\frac{t}{z_s}\left(1 - z_s^2 - z_s^2 a(z_s)\right) \tag{41}$$

and, in particular,

$$w''(-i\beta) = t^2 - \frac{t}{\beta}\left(1 + \beta^2 + \beta^2 A \exp(\beta^2/2)\right) , \tag{42}$$
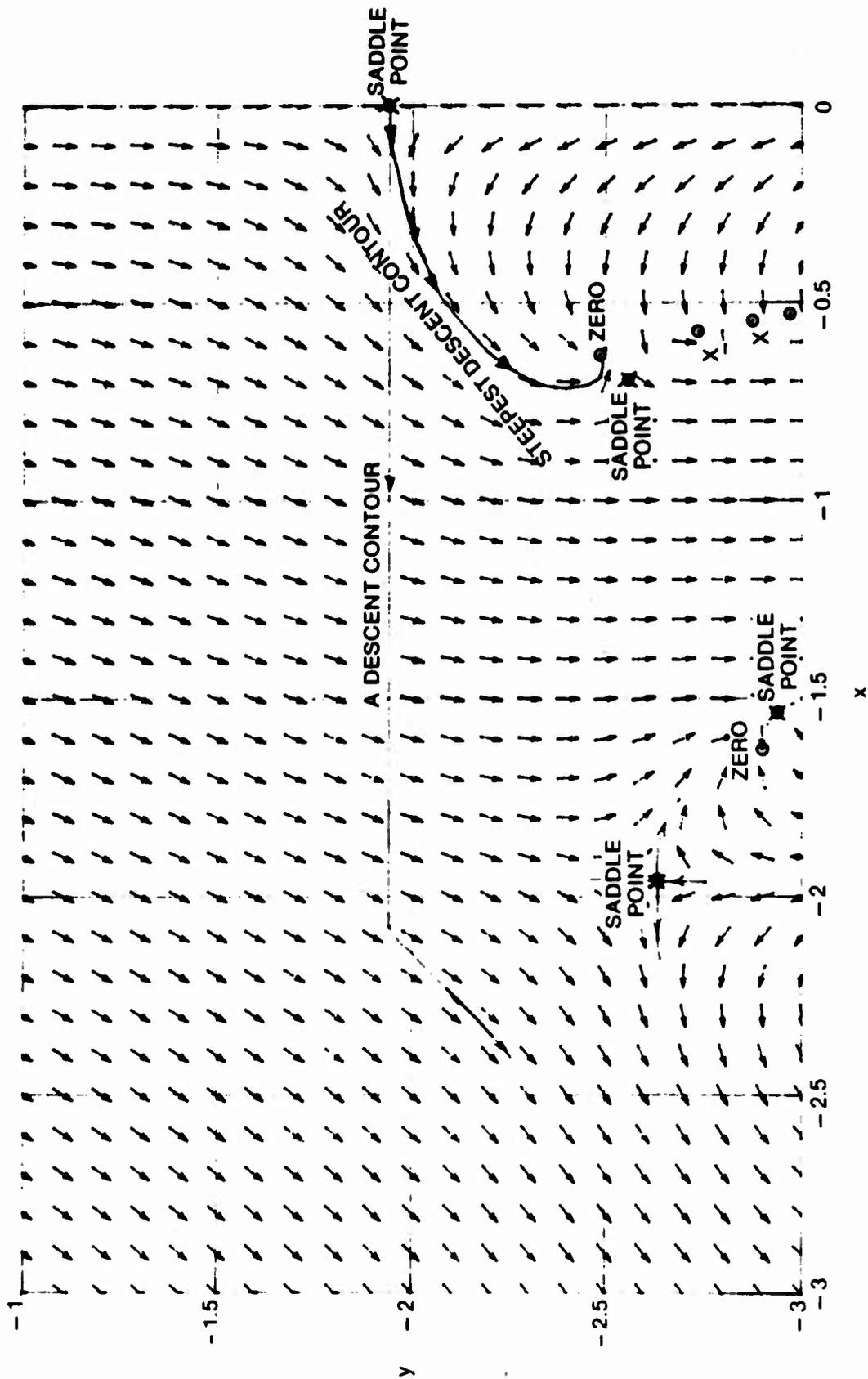
Figure 14. Descent Contours for A = 0.35, λ = 5

for the saddle point at $z_s = -i\beta$. Also $w(-i\beta)$ is given by (39). Then we have the approximation (reference 1, chapter 7)

$$p(v) \sim \frac{\exp(-A - \beta t)}{2\pi\sigma_0}\left[\exp\left(A \exp(\beta^2/2)\right) - 1\right]\left(\frac{2\pi}{|w'(-i\beta)|}\right)^{1/2} \text{ as } t \to +\infty \quad .$$

(43)

One drawback with this solution is that β depends on t through the solution of transcendental equation (37). For large t, we have, to first order,

$$\beta \approx [2 \ln(t/A)]^{1/2} \equiv L^{1/2} \quad .$$

(44)

For development of additional terms and the general philosophy of solution of these types of problems, see reference 6, pp. 11-16 and 83-84. We find, more generally,

$$\beta^2 \approx L - \frac{L}{L + 1} \ln(L) \quad .$$

(45)

Even so, substitution into (43) yields a very complicated expression for the probability density function unless t is excessively large. It can be seen that (43) decays slightly faster than an exponential for large t.

Summary

A technique for simple determination of the steepest descent direction at any point in the complex plane has been presented and illustrated with numerous examples. The movement of the original contour to an equivalent descent or steepest descent contour is an exact representation and can be deduced fairly easily from the descent information. At this point, two alternatives are available, either exact numerical evaluation of the integral or an approximation such as Laplace's method. Very difficult integrands can be handled very effectively via this approach.

Appendix A

Computer Programs

The major part of the calculations required for the six examples in the main text is that of w'(z). This is accomplished in the Subroutine Wderivative in the main program written in BASIC and listed at the end of this appendix. But, first, the six subroutines for the examples given are listed. They illustrate how little programming is actually needed to compute w'(z), provided that one has already written subroutines for the standard complex operations and functions like multiply, divide, exp(z), log(z), square root, arg(z), sin(z), cos(z), etc. These latter functions are listed for completeness as subroutines at the end of the enclosed program.

```
410     SUB Wderivative(X,Y,Rew1,Imw1)          ! Airy function
411     COM Ct,St                               ! cos(theta), sin(theta)
412     CALL Mul(X,Y,X,Y,A,B)
420     Rew1=Ct-A
430     Imw1=St-B
440     SUBEND
```

```
410     SUB Wderivative(X,Y,Rew1,Imw1)          ! Hankel function
411     Beta=.5
412     CALL Sin(X,Y,A,B)
420     Rew1=B
430     Imw1=Beta-A
440     SUBEND
```

```
410     SUB Wderivative(X,Y,Rew1,Imw1)          ! Klein-Gordon
411     Theta=.8
412     CALL Sqr(Y,-X+1,A,B)
413     CALL Sqr(Y,-X-1,C,D)
414     CALL Mul(A,B,C,D,E,F)
415     CALL Div(Theta*X,Theta*Y,E,F,G,H)
420     Rew1=G
430     Imw1=H-1
440     SUBEND
```

```
410     SUB Wderivative(X,Y,Rew1,Imw1)   ! Essential singularity
411     K=3
412     CALL Mul(X-1,Y,X-1,Y,A,B)
413     CALL Div(K,0,A,B,C,D)
414     CALL Div(1,0,X,Y,E,F)
420     Rew1=C+E
430     Imw1=D+F
440     SUBEND
```

```
410    SUB Wderivative(X,Y,Rew1,Imw1)          ! cubic function
411    COM Ca,Sa                               ! cos(alpha), sin(alpha)
412    CALL Mul(X,Y,X,Y,A,B)
413    CALL Mul(Ca,Sa,8*X-3*B,8*Y-5+3*A,C,D)
420    Rew1=-C
430    Imw1=-D
440    SUBEND
```

```
410    SUB Wderivative(X,Y,Rew1,Imw1)     ! Gaussian exponent
411    A=.35
412    T=2
413    CALL Mul(X,Y,X,Y,T1,T2)
414    CALL Exp(-.5*T1,-.5*T2,T3,T4)
415    Ar=A*T3
416    Ai=A*T4
417    CALL Exp(-Ar,-Ai,Er,Ei)
418    CALL Mul(X,Y,Ar,Ai,T1,T2)
419    CALL Div(T1,T2,1-Er,-Ei,T3,T4)
420    Rew1=-T3
430    Imw1=-T-T4
440    SUBEND
```

```
1 ! Steepest Descent via w'(z); use SUB Wderivative in line 410
10     X1=-3                              ! LEFT ABSCISSA
20     X2=3                               ! RIGHT ABSCISSA
30     Y1=-3                              ! BOTTOM ORDINATE
40     Y2=3                               ! TOP ORDINATE
50     Dx=.2                              ! X INCREMENT
60     Dy=.2                              ! Y INCREMENT
70     PLOTTER IS "GRAPHICS"
80     GRAPHICS
90     SCALE X1,X2,Y1,Y2
100    LINE TYPE 3
110    GRID 1,1                           ! GRID LINE SPACING
120    LINE TYPE 1
130    F=.2*SQR(Dx*Dx+Dy*Dy)       ! ARROW
140    Q1=1-COS(PI/12)*.8          !   INFOR-
150    Q2=SIN(PI/12)*.8            !     MATION
160    FOR X=X1 TO X2 STEP Dx
170    FOR Y=Y1 TO Y2 STEP Dy
180    CALL Wderivative(X,Y,Rewl,Imwl)     ! w'(z)
190    CALL Direction(Rewl,Imwl,Cos,Sin)  ! direction of steepest descent
200    IF ABS(Cos)+ABS(Sin)>0 THEN 230
210    OUTPUT 0;"SADDLE POINT AT ";X;Y
220    GOTO 350
230    T1=F*Cos
240    T2=F*Sin
250    Xa=X+T1*Q1
260    Xb=T2*Q2
270    Ya=Y+T2*Q1
280    Yb=T1*Q2
290    MOVE X-T1,Y-T2
300    DRAW X+T1,Y+T2
310    MOVE Xa+Xb,Ya-Yb
320    DRAW X+T1,Y+T2
330    DRAW Xa-Xb,Ya+Yb
340    PENUP
350    NEXT Y
360    NEXT X
370    PAUSE
380    DUMP GRAPHICS
390    END
400    !
410    SUB Wderivative(X,Y,Rewl,Imwl)       ! Airy function
420    Rewl=1-X*X+Y*Y                       ! w(z)=z-z^3/3
430    Imwl=-2*X*Y                          ! w'(z)=1-z^2
440    SUBEND
450    !
460    SUB Direction(Rewl,Imwl,Cos,Sin)     ! direction of
470    T=SQR(Rewl*Rewl+Imwl*Imwl)           ! steepest descent
480    IF T>0 THEN 510
490    Cos=Sin=0
500    GOTO 530
510    Cos=-Rewl/T
520    Sin=Imwl/T
530    SUBEND
540    !
```

```
550     SUB Mul(X1,Y1,X2,Y2,A,B)              ! Z1*Z2
560     A=X1*X2-Y1*Y2
570     B=X1*Y2+X2*Y1
580     SUBEND
590     !
600     SUB Div(X1,Y1,X2,Y2,A,B)             ! Z1/Z2
610     T=X2*X2+Y2*Y2
620     A=(X1*X2+Y1*Y2)/T
630     B=(Y1*X2-X1*Y2)/T
640     SUBEND
650     !
660     SUB Exp(X,Y,A,B)                      ! EXP(Z)
670     T=EXP(X)
680     A=T*COS(Y)
690     B=T*SIN(Y)
700     SUBEND
710     !
720     SUB Log(X,Y,A,B)           ! PRINCIPAL LOG(Z)
730     A=.5*LOG(X*X+Y*Y)
740     IF X<>0 THEN 770
750     B=.5*PI*SGN(Y)
760     GOTO 790
770     B=ATN(Y/X)
780     IF X<0 THEN B=B+PI*(1-2*(Y<0))
790     SUBEND
800     !
810     SUB Sqr(X,Y,A,B)           ! PRINCIPAL SQR(Z)
820     IF X<>0 THEN 860
830     A=B=SQR(.5*ABS(Y))
840     IF Y<0 THEN B=-B
850     GOTO 970
860     F=SQR(SQR(X*X+Y*Y))
870     T=.5*ATN(Y/X)
880     A=F*COS(T)
890     B=F*SIN(T)
900     IF X>0 THEN 970
910     T=A
920     A=-B
930     B=T
940     IF Y>=0 THEN 970
950     A=-A
960     B=-B
970     SUBEND
980     !
990     SUB Arg(X,Y,A)             ! PRINCIPAL ARG(Z)
1000    IF X=0 THEN A=.5*PI*SGN(Y)
1010    IF X<>0 THEN A=ATN(Y/X)
1020    IF X<0 THEN A=A+PI*(1-2*(Y<0))
1030    SUBEND
1040    !
1050    SUB Power(X,Y,R,A,B)  ! PRINCIPAL POWER Z^R
1060    F=EXP(.5*R*LOG(X*X+Y*Y))
1070    CALL Arg(X,Y,T)
1080    A=F*COS(R*T)
1090    B=F*SIN(R*T)
1100    SUBEND
1110    !
```

```
1120    SUB Sin(X,Y,A,B)                  ! SIN(Z)
1130    E=EXP(Y)
1140    A=.5*SIN(X)*(E+1/E)
1150    IF ABS(Y)<.1 THEN 1180
1160    S=.5*(E-1/E)
1170    GOTO 1200
1180    S=Y*Y
1190    S=Y*(120+S*(20+S))/120
1200    B=COS(X)*S
1210    SUBEND
1220    !
1230    SUB Cos(X,Y,A,B)                  ! COS(Z)
1240    E=EXP(Y)
1250    A=.5*COS(X)*(E+1/E)
1260    IF ABS(Y)<.1 THEN 1290
1270    S=.5*(E-1/E)
1280    GOTO 1310
1290    S=Y*Y
1300    S=Y*(120+S*(20+S))/120
1310    B=-SIN(X)*S
1320    SUBEND
1330    !
1340    SUB Sinh(X,Y,A,B)                 ! SINH(Z)
1350    E=EXP(X)
1360    B=.5*SIN(Y)*(E+1/E)
1370    IF ABS(X)<.1 THEN 1400
1380    S=.5*(E-1/E)
1390    GOTO 1420
1400    S=X*X
1410    S=X*(120+S*(20+S))/120
1420    A=COS(Y)*S
1430    SUBEND
1440    !
1450    SUB Cosh(X,Y,A,B)                 ! COSH(Z)
1460    E=EXP(X)
1470    A=.5*COS(Y)*(E+1/E)
1480    IF ABS(X)<.1 THEN 1510
1490    S=.5*(E-1/E)
1500    GOTO 1530
1510    S=X*X
1520    S=X*(120+S*(20+S))/120
1530    B=SIN(Y)*S
1540    SUBEND
```

Appendix B

Steepest Descent for General Analytic Function

Here we do not force the integrand to be of the form in (1), but consider the general integral $\int_C dz\, f(z)$. Let $f(z)$ be analytic in a region in the complex plane. The magnitude-squared value is

$$M = |f(z)|^2 \equiv f_r^2 + f_i^2 \quad . \tag{B-1}$$

The direction of steepest descent for M is opposite to the gradient of M, which is

$$\nabla M = \frac{\partial M}{\partial x}\, \vec{a}_x + \frac{\partial M}{\partial y}\, \vec{a}_y \quad . \tag{B-2}$$

But, from (B-1),

$$\frac{\partial M}{\partial x} = 2\left(f_r\, \frac{\partial f_r}{\partial x} + f_i\, \frac{\partial f_i}{\partial x}\right) \quad ,$$

$$\frac{\partial M}{\partial y} = 2\left(f_r\, \frac{\partial f_r}{\partial y} + f_i\, \frac{\partial f_i}{\partial y}\right) \quad . \tag{B-3}$$

Now if function f is analytic at z, then

$$f'(z) = \frac{\partial f_r}{\partial x} + i\frac{\partial f_i}{\partial x} = \frac{\partial f_i}{\partial y} - i\frac{\partial f_r}{\partial y} \quad . \tag{B-4}$$

So we can express

$$\frac{\partial M}{\partial x} = 2\, \mathrm{Re}\{f^*(z)\, f'(z)\} \quad ,$$

$$\frac{\partial M}{\partial y} = -2\, \mathrm{Im}\{f^*(z)\, f'(z)\} \quad . \tag{B-5}$$

Therefore, the steepest descent direction for M has components that are the negatives of (B-5) or, equivalently, are proportional to

$$-\mathrm{Re}\left\{\frac{f'(z)}{f(z)}\right\} \quad , \quad \mathrm{Im}\left\{\frac{f'(z)}{f(z)}\right\} \quad . \tag{B-6}$$

The basic calculation for determination of steepest descent directions is thus seen to be

$$f'(z)/f(z) = \frac{d}{dz}\ln f(z).$$

When $f'(z_s) = 0$, we have a saddle point of f at $z = z_s$. Near the saddle point,

$$f(z) \cong f(z_s) + \frac{1}{2}f''(z_s)(z - z_s)^2 \equiv f_0 + \frac{1}{2}f_2\Delta^2 \quad \text{for small } \Delta \quad . \tag{B-7}$$

Then

$$M \cong \left(f_0 + \tfrac{1}{2}f_2\Delta^2\right)\left(f_0^* + \tfrac{1}{2}f_2^*\Delta^{*2}\right)$$

$$\cong |f_0|^2 + \text{Re}\left\{f_0^* f_2 \Delta^2\right\} \quad \text{for small } \Delta \quad . \tag{B-8}$$

Now let

$$f_0^* f_2 = ae^{i\alpha} \,, \quad \Delta = re^{i\theta} \quad . \tag{B-9}$$

Then (B-8) yields

$$M \cong |f_0|^2 + ar^2 \cos(\alpha + 2\theta) \quad , \tag{B-10}$$

which has two peaks and two valleys versus θ in a 2π interval (for $a \neq 0$); this is characteristic of a saddle point. The directions of steepest descent, θ_d, at $z = z_s$ occur when

$$\alpha + 2\theta_d = \pi \text{ or } 3\pi \,, \quad \theta_d = \frac{\pi - \alpha}{2} \text{ or } \frac{\pi - \alpha}{2} + \pi \quad . \tag{B-11}$$

Notice, from (B-9), that

$$\alpha = \arg\left\{f_0^* f_2\right\} = \arg\left\{f^*(z_s)\, f''(z_s)\right\} = \arg\left\{f''(z_s)/f(z_s)\right\} \quad . \tag{B-12}$$

Now let us investigate the behavior of the complex function $f(z)$ near z_s, along the steepest descent contours. For θ_d as given by (B-11), Δ in (B-9) becomes

$$\Delta = \pm r \exp\left(i\frac{\pi - \alpha}{2}\right) \,, \quad \Delta^2 = -r^2 e^{-i\alpha} \quad ; \tag{B-13}$$

and there follows, from (B-7) and (B-9), for small r,

$$f(z) \cong f_0 - \tfrac{1}{2}r^2 f_2 e^{-i\alpha} = f_0\left(1 - \frac{r^2 a}{2|f_0|^2}\right)$$

$$\text{on steepest descent contour near } z_s \quad . \tag{B-14}$$

That is, since the term in parentheses is real and positive,

$$\arg f(z) = \arg f(z_s) \text{ on steepest descent contour near } z_s. \tag{B-15}$$

More generally, it can be shown that

$$\arg f(z) = \arg f(z_s)$$

$$\text{everywhere on steepest descent contour through } z_s. \tag{B-16}$$

Now let us apply these general results to the special case where $f(z) = \exp(w(z))$. Then $f'(z) = w'(z) \exp(w(z))$, and $f'(z) = 0$ when $w'(z) = 0$. Thus saddle point locations are as usually stated. Also, as needed in (B-6),

$$\frac{f'(z)}{f(z)} = w'(z) \quad , \tag{B-17}$$

and then (B-6) agrees with (9). Furthermore, since

$$f''(z) = [w''(z) + w'^2(z)] \exp(w(z)) \quad ,$$

$$f''(z_s) = w''(z_s) \exp(w(z_s)) \quad , \tag{B-18}$$

then (B-12) yields

$$\alpha = \arg w''(z_s) \quad . \tag{B-19}$$

When (B-19) is used in (B-11), the steepest descent directions corroborate reference 1, (7.1.8) and (7.1.19). Finally, (B-16) yields

$$\arg f(z) = \arg\{\exp(w(z))\} = \arg\{\exp(u + iv)\} = v \quad , \tag{B-20}$$

meaning that v is constant on steepest descent contours; this agrees with reference 1, Lemma 7.1.

What this all demonstrates is that, for a general given integrand f(z), we can let $w(z) = \ln f(z)$ and apply our usual techniques on w(z). This procedure was adopted in the Gaussian exponent example in the main text of this report.

References

1. N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals*, Holt, Rinehart, and Winston, NY, 1975.

2. G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of a Complex Variable*, McGraw-Hill Book Co., NY, 1966.

3. C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill Book Co., NY, 1978.

4. Personal communication with N. Bleistein, 21 January 1981.

5. A. D. Spaulding and D. Middleton, "Optimum Reception in an Impulsive Interference Environment - Part I: Coherent Detection," *IEEE Trans. on Communications*, vol. COM-25, no. 9, September 1977, pp. 910-923.

6. F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, NY, 1974.

On Characterization Of the Probability Distribution of Measured Data

A. H. Nuttall

ABSTRACT

A procedure for determining the probability density function or cumulative distribution function of measured data is considered, whereby the sample distribution is plotted on several types of transformed axes and compared with candidate distributions. The abscissa transformations considered are linear or logarithmic, while the ordinate transformations considered are Gaussian or logarithmic. The four different distributions which plot as straight lines on the appropriate combinations of transformations are displayed, as well as being cross-plotted on mismatched paper. Applications to random data are given.

INTRODUCTION

In order to characterize the probability density function or cumulative distribution function (CDF) of measured data, the sample CDF is often displayed on a candidate type of graph paper, as for example, a linear abscissa with a Gaussianly-transformed ordinate. If the underlying process is truly Gaussian, the resulting CDF plot will resemble a straight line, if the number of independent samples is large. However, other CDFs will plot as curves on this type of paper.

Two frequently-used abscissa transformations are linear and logarithmic (dB for example). And two popular ordinate transformations are Gaussian and logarithmic. Since there are four possible combinations of these two pairs of transformations, it would be worthwhile to have several common analytic CDFs plotted on all four types of paper, to serve as comparisons for future measured data. As comparisons, we select the four CDFs, each of which leads to a straight line plot on one of the four combinations. In addition, we consider a more general CDF with several parameters, the Q_M distribution, which occurs for narrowband processing, and illustrate a transition between two of the basic CDFs.

In order to guarantee that the CDF graphs lie in a standard region of plotting space, the abscissas are normalized, by subtracting the mean of the random variable under consideration, and dividing by the standard deviation. This normalization realizes, in some cases, a unique plot for the CDF, regardless of the values of the parameters of the CDF. In other cases, it reduces the dependence on the number of parameters. This convenient behavior minimizes the number of plots and comparisons that have to be considered. Also, for measured data, the two required statistics, namely the sample mean and sample variance, can be simply and easily computed.

LINEAR ABSCISSA

Let random variable (RV) r have cumulative distribution function (CDF)

$$P_r(u) = \text{Prob}(r < u),\qquad(1)$$

probability density function (PDF)

$$p_r(u) = P_r'(u),\qquad(2)$$

mean m_r, and variance σ_r^2.

For plotting purposes, define normalized abscissa

$$x = \frac{u - m_r}{\sigma_r}\qquad(3)$$

as a linear transformation. Thus x measures the number of standard deviations that the RV is from its mean. For the ordinate of the plot, define

$$y = g\{P_r(u)\},\qquad(4)$$

where transformation $g\{\ \}$ is as-yet unspecified. Substitution of (3) in (4) yields the characteristic of interest:

$$y = g\{P_r(m_r + \sigma_r x)\}.\qquad(5)$$

Thus, for a given CDF P_r, y can be plotted vs x, as a function of any parameters of P_r that still remain in (5), after m_r and σ_r have been evaluated and used in (5).

For the examples below, the following definitions will be useful (see Ref. 1, (26.2), (26.2.22), and (26.2.23)):

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

$$\Phi(x) = \int_{-\infty}^{x} ds \, \phi(s), \tag{6}$$

$$\tilde{\Phi}(\ ) = \text{inverse function to } \Phi(\ ).$$

Example 1. Gaussian

The PDF and CDF are

$$p_r(u) = \frac{1}{b} \phi\left(\frac{u-a}{b}\right) \quad \text{for all } u,$$

$$P_r(u) = \Phi\left(\frac{u-a}{b}\right) \quad \text{for all } u. \tag{7}$$

The mean and standard deviation of RV r are

$$m_r = a, \quad \sigma_r = b; \quad \text{thus } x = \frac{u-a}{b}. \tag{8}$$

Substitution of (7) and (8) in (5) yields characteristic

$$y = g\left\{\Phi(x)\right\} \quad \text{for all } x. \tag{9}$$

This result is independent of parameters a and b of the Gaussian PDF, regardless of what ordinate transformation $g\{\ \}$ is employed. This convenient behavior is a result of the selection of linear transformation (3). Since no parameters remain in (9), a unique plot will result for (9), regardless of the $g\{\ \}$ selected.

Example 2. Exponential

$$p_r(u) = b \exp(-b(u-a)) \text{ for } u \geq a,$$

$$P_r(u) = 1 - \exp(-b(u-a)) \text{ for } u \geq a. \tag{10}$$

There follows

$$m_r = a + \frac{1}{b} \, , \quad \sigma_r = \frac{1}{b} \, , \quad x = b(u-a) - 1. \tag{11}$$

Substitution of (10) and (11) in (5) yields

$$y = g\{1 - e^{-1-x}\} \text{ for } x \geq -1. \tag{12}$$

This result is also independent of parameters a and b.

Example 3. Log-Normal

Here, the natural logarithm of RV r is Gaussianly distributed. That is, the PDF and CDF of r are

$$p_r(u) = \frac{1}{bu} \phi\left(\frac{\ln u - a}{b}\right) \text{ for } u > 0,$$

$$P_r(u) = \Phi\left(\frac{\ln u - a}{b}\right) \text{ for } u > 0. \tag{13}$$

There follows

$$m_r = e^a\, e^{b^2/2}, \qquad \sigma_r = e^a \sqrt{e^{2b^2} - e^{b^2}},$$

$$x = \frac{u - e^a\, e^{b^2/2}}{e^a \sqrt{e^{2b^2} - e^{b^2}}}. \tag{14}$$

Substitution of (13) and (14) in (5) yields

$$y = g\left\{ \Phi\left(\frac{1}{b}\ln\left[x\sqrt{e^{2b^2} - e^{b^2}} + e^{b^2/2}\right]\right)\right\} \quad \text{for } x > -\frac{e^{b^2/2}}{\sqrt{e^{2b^2} - e^{b^2}}}. \tag{15}$$

This result is independent of a, but it does depend on parameter b of (13); thus (15) will yield a family of characteristics, regardless of the choice of ordinate transformation $g\{\ \}$.

Example 4.   Power-Law

$$p_r(u) = b\, a^b / u^{b+1} \quad \text{for } u \geq a,$$

$$P_r(u) = 1 - (a/u)^b \quad \text{for } u \geq a. \tag{16}$$

Then for paramter $b > 2$,

$$m_r = \frac{ab}{b-1}, \qquad \sigma_r = \frac{a}{b-1}\sqrt{\frac{b}{b-2}},$$

$$x = \frac{u - \frac{ab}{b-1}}{\frac{a}{b-1}\sqrt{\frac{b}{b-2}}}. \tag{17}$$

Substitution of (16) and (17) in (5) yields

$$y = g\left\{ 1 - \left(\frac{x}{b-1}\sqrt{\frac{b}{b-2}} + \frac{b}{b-1}\right)^{-b}\right\} \quad \text{for } x \geq -\sqrt{\frac{b-2}{b}}. \tag{18}$$

This result depends on b, but not on a.

LOGARITHMIC ABSCISSA

The other abscissa transformation we are interested in is a logarithmic one, as employed for example in dB plots. We now develop the basic relations and normalization for this case. When RV r above is confined to positive values, let transformed RV

$$t = \ln r \tag{19}$$

have CDF and PDF

$$P_t(v) = \text{Prob}(t < v), \qquad p_t(v) = P_t'(v), \tag{20}$$

respectively. And let the mean and standard deviation of RV t be denoted by m_t and σ_t.

We then have, by use of (20), (19), and (1), the CDF of t as

$$P_t(v) = \text{Prob}(t < v) = \text{Prob}(\ln r < v) = \text{Prob}(r < e^v) = P_r(e^v). \tag{21}$$

The PDF follows as

$$p_t(v) = P_t'(v) = e^v p_r(e^v). \tag{22}$$

Several alternative expressions are available for the n-moment of RV t:

$$\overline{t^n} = \overline{(\ln r)^n} = \int dv \, v^n p_t(v) = \int du \, (\ln u)^n p_r(u) = \int dv \, v^n e^v p_r(e^v). \tag{23}$$

In particular, we have

$$m_t = \overline{t} \quad , \quad \sigma_t^2 = \overline{t^2} - \overline{t}^2 . \tag{24}$$

The normalized abscissa we adopt for plotting purposes is a linear transformation on v:

$$x = \frac{v - m_t}{\sigma_t} . \tag{25}$$

Here, x measures the number of standard deviations that the RV is from its mean. The ordinate is obtained according to

$$y = g\left\{ P_t(v) \right\} = g\left\{ P_t\left(m_t + \sigma_t \, x \right) \right\} , \tag{26}$$

where transformation $g\{\ \}$ is as-yet unspecified. We can now plot y vs x, once a CDF, $P_t(\ )$, is specified.

Example 1. Gaussian

The CDF for the RV r is given in (7). Since RV r can go negative, we cannot apply transformation (19), and this example is not applicable.

Example 2. Exponential

We take the special case of a = 0 here; see (10). (The case of a > 0 can be worked, but involves exponential integrals.) We have

$$P_r(u) = 1 - \exp(-bu) \quad \text{for} \quad u \geq 0,$$

$$p_r(u) = b \exp(-bu) \quad \text{for} \quad u \geq 0. \tag{27}$$

By use of (21), we obtain the CLF of t as

$$P_t(v) = 1 - \exp(-b\, e^v) \quad \text{for all } v. \tag{28}$$

And from (23),

$$\overline{t^n} = \int_0^\infty du\, (\ln u)^n\, b\, e^{-bu} = \int_0^\infty ds\, (\ln s - \ln b)^n\, e^{-s}. \tag{29}$$

In particular,

$$\overline{t} = \int_0^\infty ds\, (\ln s - \ln b)\, e^{-s} = -\gamma - \ln b,$$

$$\overline{t^2} = \int_0^\infty ds\, (\ln s - \ln b)^2\, e^{-s} = \frac{\pi^2}{6} + \gamma^2 + 2\gamma \ln b + (\ln b)^2, \tag{30}$$

by use of Ref. 2, 4.331 1 and 4.335 1, respectively; γ is Euler's constant. There follows

$$m_t = -\gamma - \ln b, \quad \sigma_t = \pi/\sqrt{6},$$

$$x = \frac{v + \ln b + \gamma}{\pi/\sqrt{6}}. \tag{31}$$

Substitution of (28) and (31) in (26) yields

$$y = g\left\{1 - \exp\left(-e^{x\frac{\pi}{\sqrt{6}} - \gamma}\right)\right\} \quad \text{for all } x. \tag{32}$$

This result is independent of b; a was set to 0 here.

Example 3. Log-Normal

From (13) and (21), the two relevant CDFs are

$$P_r(u) = \Phi\left(\frac{\ln u - a}{b}\right) \quad \text{for } u > 0,$$

$$P_t(v) = \Phi\left(\frac{v - a}{b}\right) \quad \text{for all } v. \tag{33}$$

There follows immediately

$$m_t = a, \quad \sigma_t = b, \quad x = \frac{v - a}{b}. \tag{34}$$

Substitution of (33) and (34) in (26) yields

$$y = g\left\{\Phi(x)\right\} \quad \text{for all } x. \tag{35}$$

This result is independent of a and b; contrast this with the linear abscissa result in (15). Thus the CDF for a log-normal RV will plot as a unique curve on a logarithmic abscissa, regardless of the ordinate transformation g{ }.

Example 4. Power-Law

From (16) and (21),

$$P_r(u) = 1 - (a/u)^b \quad \text{for} \quad u \geq a,$$

$$P_t(v) = 1 - a^b e^{-bv} \quad \text{for} \quad v \geq \ln a. \tag{36}$$

Then

$$p_t(v) = a^b b \, e^{-bv} \quad \text{for} \quad v \geq \ln a, \tag{37}$$

and there follows for the n-th moment of t,

$$\overline{t^n} = \int_{\ln a}^{\infty} dv \, v^n a^b b \, e^{-bv} = \frac{1}{b^n} \int_0^{\infty} ds \, (s + b \ln a)^n e^{-s}. \tag{38}$$

In particular,

$$\overline{t} = \frac{1}{b} + \ln a, \qquad \overline{t^2} = \frac{2}{b^2} + \frac{2}{b} \ln a + (\ln a)^2, \tag{39}$$

and therefore

$$m_t = \frac{1}{b} + \ln a, \qquad \sigma_t = \frac{1}{b},$$

$$x = bv - b \ln a - 1. \tag{40}$$

Substitution of (36) and (40) in (26) yields

$$y = g \left\{ 1 - e^{-1-x} \right\} \quad \text{for} \quad x \geq -1. \tag{41}$$

This result is independent of a and b, in contrast with the linear abscissa result in (18).

ORDINATE TRANSFORMATIONS

Thus far, ordinate transformation $g\{\}$ in (5) or (26) has been arbitrary. The particular two transformations we are interested in are

$$g\{P\} = \tilde{\tilde{\Phi}}(P) \qquad \text{Gaussian ordinate,} \qquad (42)$$

and

$$g\{P\} = -\ln(1-P) \qquad \text{logarithmic ordinate.} \qquad (43)$$

The transformation in (42) is the inverse CDF for a normalized Gaussian RV, as introduced in (6).

We now consider the possible combinations of (42) and (43) with the results above for the linear abscissa and the logarithmic abscissa. The first case is

1. Linear Abscissa and Gaussian Ordinate

We employ (42) on (9), (12), (15) and (18), to get

$$
\begin{aligned}
&\text{1-1} & & y = x \quad \text{for all } x \\
&\text{1-2} & & y = \tilde{\tilde{\Phi}}\left(1 - e^{-1-x}\right) \quad \text{for } x > -1 \\
&\text{1-3} & & y = \frac{1}{b}\ln\left(x\sqrt{e^{2b^2}-e^{b^2}} + e^{b^2/2}\right) \quad \text{for } x > -\frac{e^{b^2/2}}{\sqrt{e^{2b^2}-e^{b^2}}} \\
&\text{1-4} & & y = \tilde{\tilde{\Phi}}\left(1 - \left(\frac{x}{b-1}\sqrt{\frac{b}{b-2}} + \frac{b}{b-1}\right)^{-b}\right) \quad \text{for } x > -\sqrt{\frac{b-2}{b}}
\end{aligned}
\tag{44}
$$

2. Linear Abscissa and Logarithmic Ordinate

In this case, we use ordinate tranformation (43) on (9), (12), (15) and (18), to obtain

$$2\text{-}1 \quad y = -\ln\left(1 - \overline{\Phi}(x)\right) \quad \text{for all } x$$

$$2\text{-}2 \quad y = 1 + x \quad \text{for } x \geq -1$$

$$2\text{-}3 \quad y = -\ln\left(1 - \overline{\Phi}\left(\frac{1}{b}\ln\left(x\sqrt{e^{2b^2} - e^{b^2}} + e^{b^2/2}\right)\right)\right)$$

$$\text{for } x > -\frac{e^{b^2/2}}{\sqrt{e^{2b^2} - e^{b^2}}} \tag{45}$$

$$2\text{-}4 \quad y = b\ln\left(\frac{x}{b-1}\sqrt{\frac{b}{b-2}} + \frac{b}{b-1}\right) \quad \text{for } x > -\sqrt{\frac{b-2}{b}}$$

3. Logarithmic Abscissa and Gaussian Ordinate

Here we apply (42) to (32), (35), and (41); the Gaussian example is not applicable.

$$3\text{-}1 \quad \text{Not applicable}$$

$$3\text{-}2 \quad y = \tilde{\Phi}\left(1 - \exp\left(-e^{x\frac{\pi}{\sqrt{6}} - \gamma}\right)\right) \quad \text{for all } x$$

$$\tag{46}$$

$$3\text{-}3 \quad y = x \quad \text{for all } x$$

$$3\text{-}4 \quad y = \tilde{\Phi}\left(1 - e^{-1-x}\right) \quad \text{for } x > -1$$

16

4. Logarithmic Abscissa and Logarithmic Ordinate

In this case, we employ (43) on (32), (35), and (41).

4-1   Not applicable

4-2 $y = \exp\left(x \frac{\pi}{\sqrt{6}} - \gamma\right)$ for all x (47)

4-3 $y = -\ln\left(1 - \Phi(x)\right)$ for all x

4-4 $y = 1 + x$ for $x \geq -1$

GRAPHICAL RESULTS FOR THE FOUR PDF EXAMPLES

In figure 1 are collected the results for a linear abscissa and Gaussian ordinate. In particular, figure 1A contains the plots for RVs with a Gaussian PDF (case 1-1 of (44)) and an exponential PDF (case 1-2 of (44)). The abscissa limits correspond to ± 3 standard deviations from the mean, while the ordinate limits are for probability values .001 and .999. This is the type of paper on which a Gaussian CDF plots as the unique straight line indicated.

Figure 1B gives the situation for a log-normal PDF (case 1-3 of (44)), for various values of parameter b. The curve labelled $b = 0^+$ is the limit of 1-3 in (44) as $b \rightarrow 0^+$, namely

$$y \rightarrow x \quad \text{for all } x \quad \text{as } b \rightarrow 0+. \tag{48}$$

Thus as the spread of the log-normal RV (see (14)) tends to zero, the plot tends to the straight line for a Gaussian RV; see figure 1A.

Figure 1C contains the power-law PDF (case 1-4 of (44)) for various values of b. The curve labelled $b = {}^+\infty$ is the limit of 1-4 in (44) as $b \rightarrow {}^+\infty$, namely

$$y \rightarrow \tilde{\tilde{\Phi}}\left(1 - e^{-1-x}\right) \quad \text{for } x > -1 \quad \text{as } b \rightarrow +\infty. \tag{49}$$

This is the exponential PDF plotted in figure 1A.
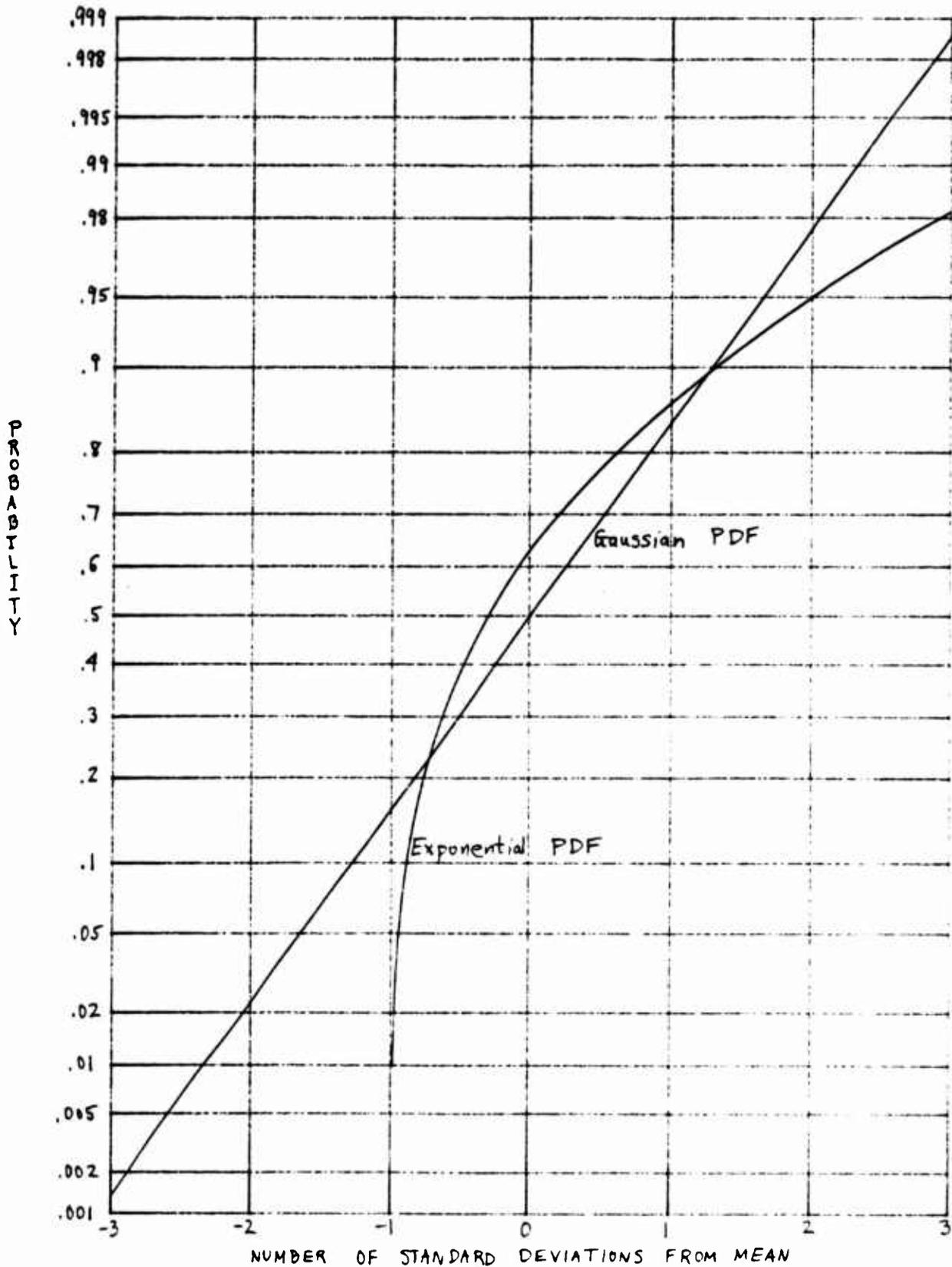
Figure 1A. Gaussian and Exponential PDFs

Figure 1. Plots for Linear Abscissa and Gaussian Ordinate
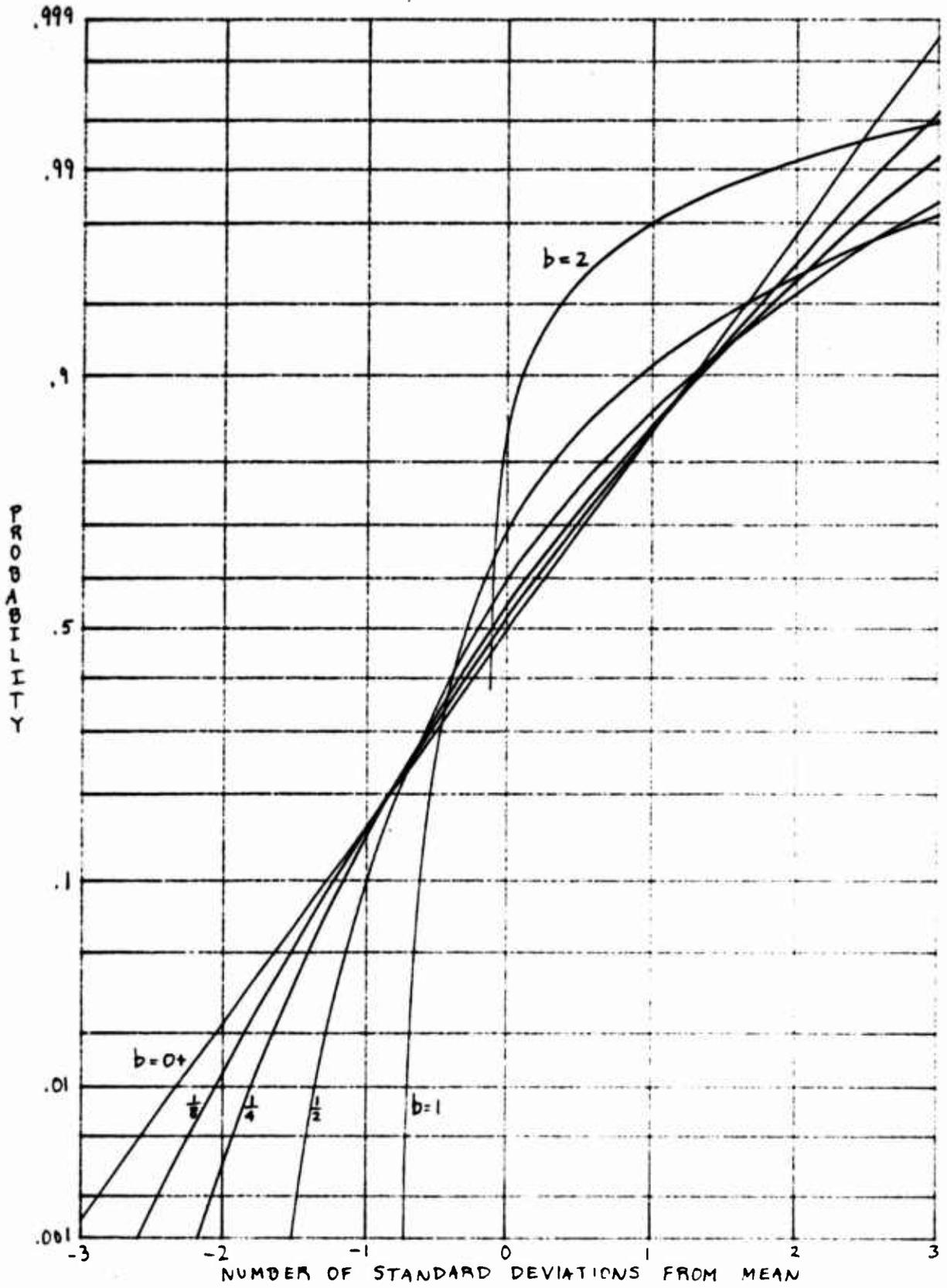
Figure 1 B. Log-Normal PDF

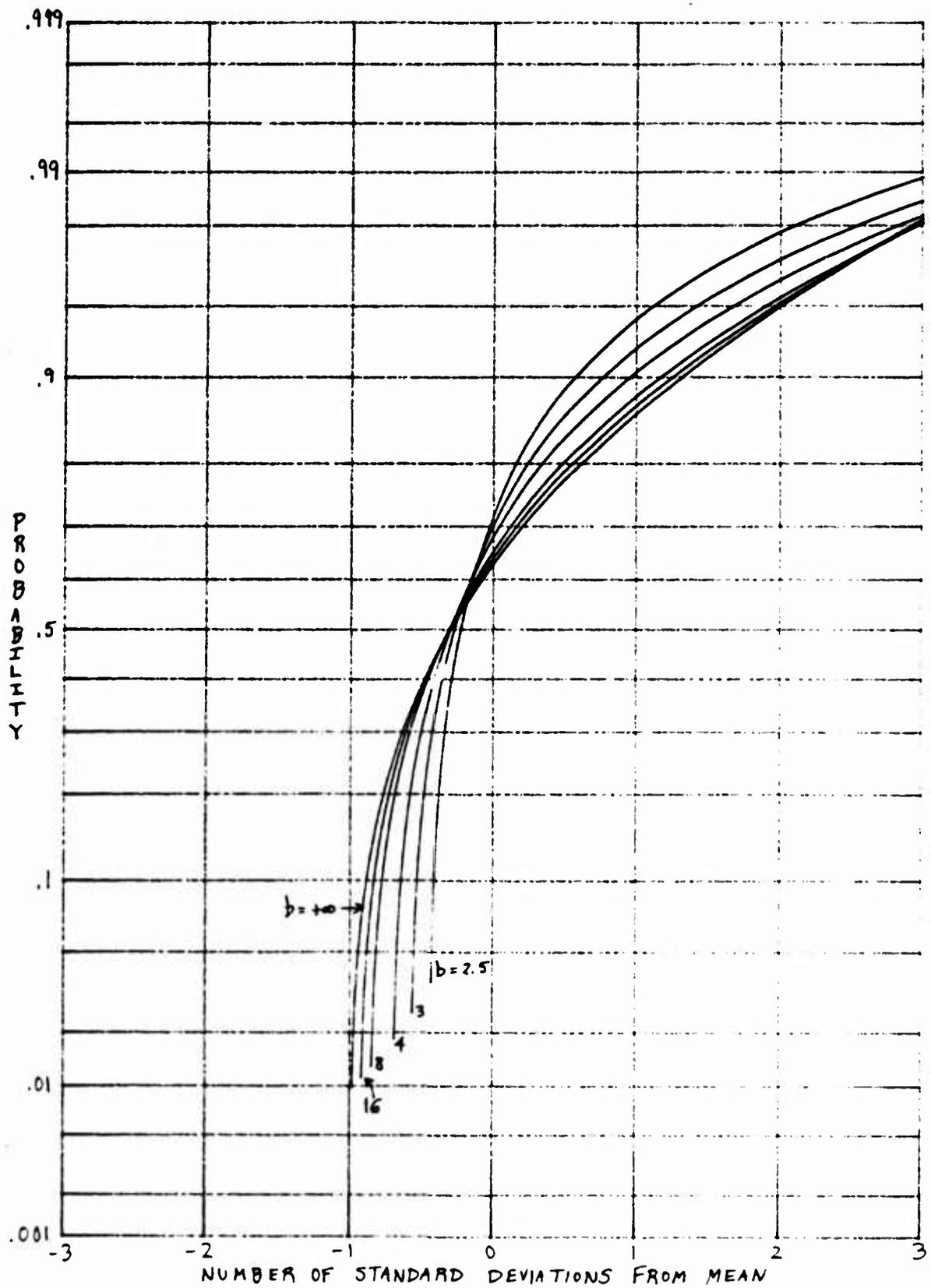Figure 1. Plots for Linear Abscissa and Gaussian Ordinate

Figure 1C. Power-Law PDF

Figure 1. Plots for Linear Abscissa and Gaussian Ordinate

The results for a linear abscissa and a logarithmic ordinate are given in parts A, B, and C of figure 2. Now the exponential PDF plots as a straight line; see figure 2A. The log-normal PDF is displayed in figure 2B. The curve labelled b = 0⁺ is the limit of 2-3 in (45):

$$y \rightarrow -\ln\left(1 - \Phi(x)\right) \text{ for all } x \qquad \text{as } b \rightarrow 0+. \tag{50}$$

This is also the Gaussian plot in figure 2A. The power-law PDF is given in figure 2C. The curve labelled b = +∞ is the limit of 2-4 in (45):

$$y \rightarrow 1+x \text{ for } x \geq -1 \qquad \text{as } b \rightarrow +\infty. \tag{51}$$

This is also the exponential plot in figure 2A.

For a logarithmic abscissa and Gaussian ordinate, the equations in (46) are plotted on figure 3. Here, the log-normal PDF yields a straight line. There are no parameters that have to be investigated on this type of plotting paper, if the PDF is exponential, log-normal, or power-law; see (46). Other PDFs could, of course, still involve parameters that are not suppressed.

The analogous results for a logarithmic abscissa and logarithmic ordinate are displayed in figure 4. This time, the power-law PDF yields the straight line. Again, no parameters of these particular PDFs are involved, as they have disappeared in (47).
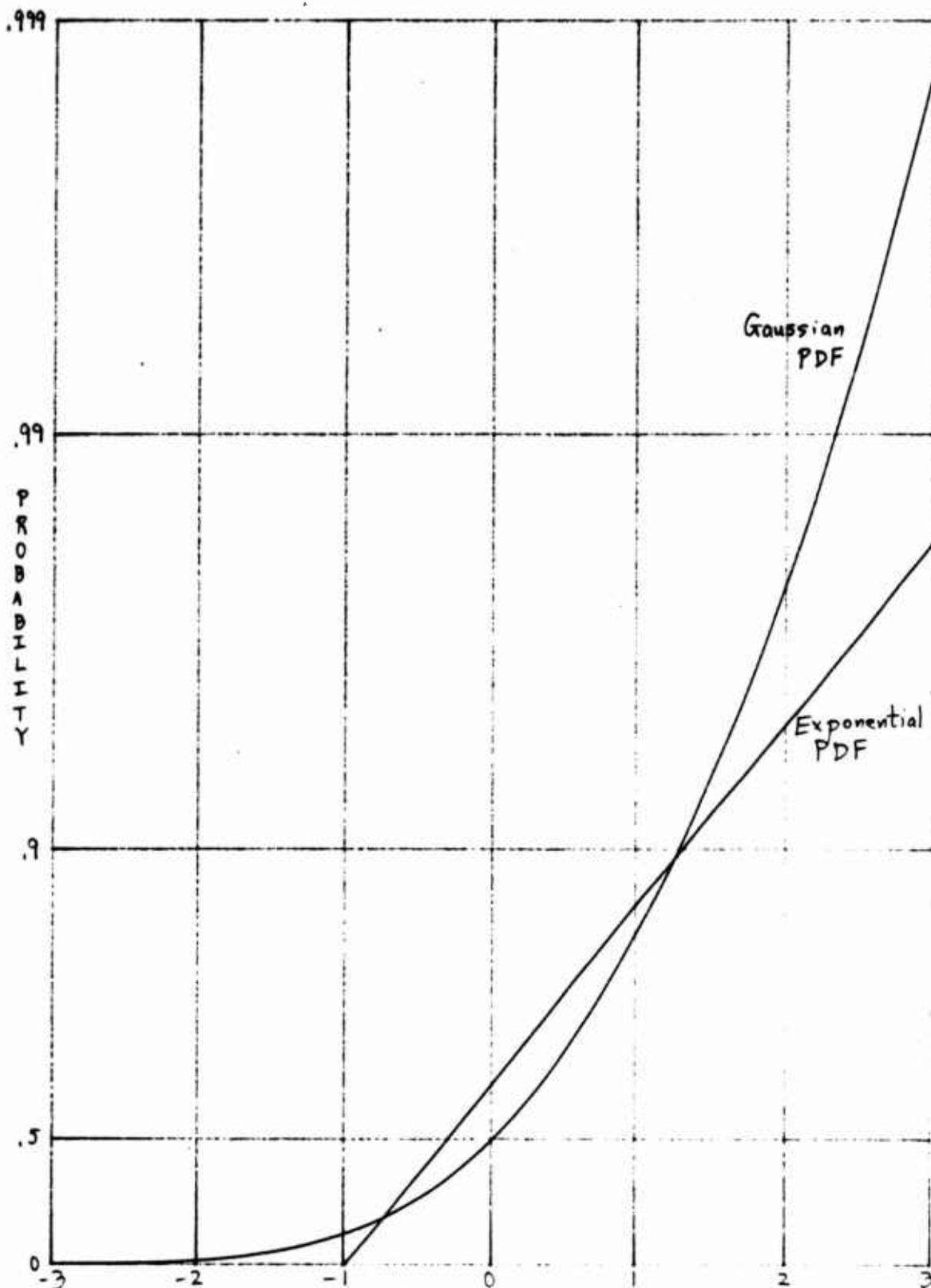
NUMBER OF STANDARD DEVIATIONS FROM MEAN
Figure 2A. Gaussian and Exponential PDFs
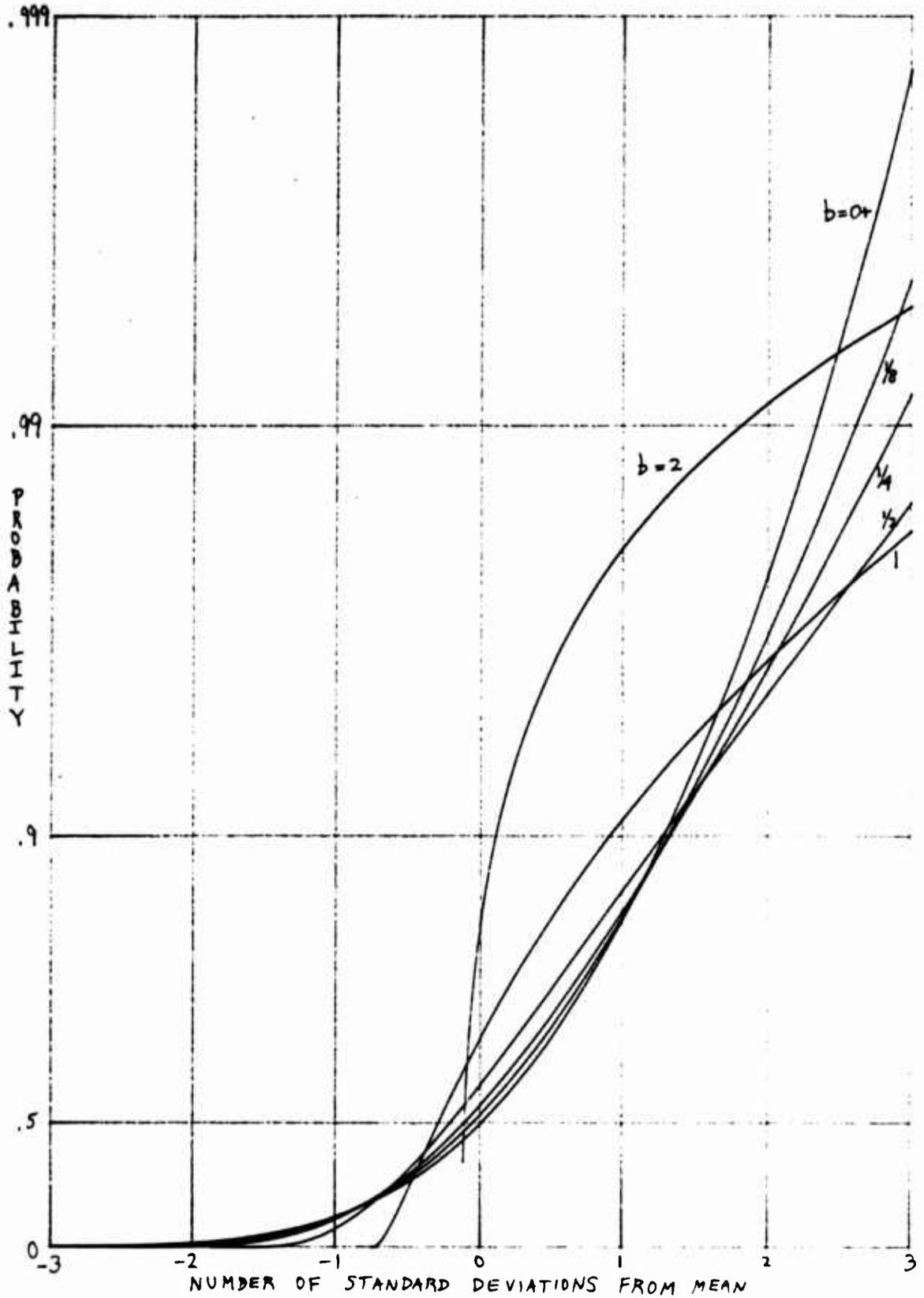Figure 2. Plots for Linear Abscissa and Logarithmic Ordinate

Figure 2b. Log-Normal PDF

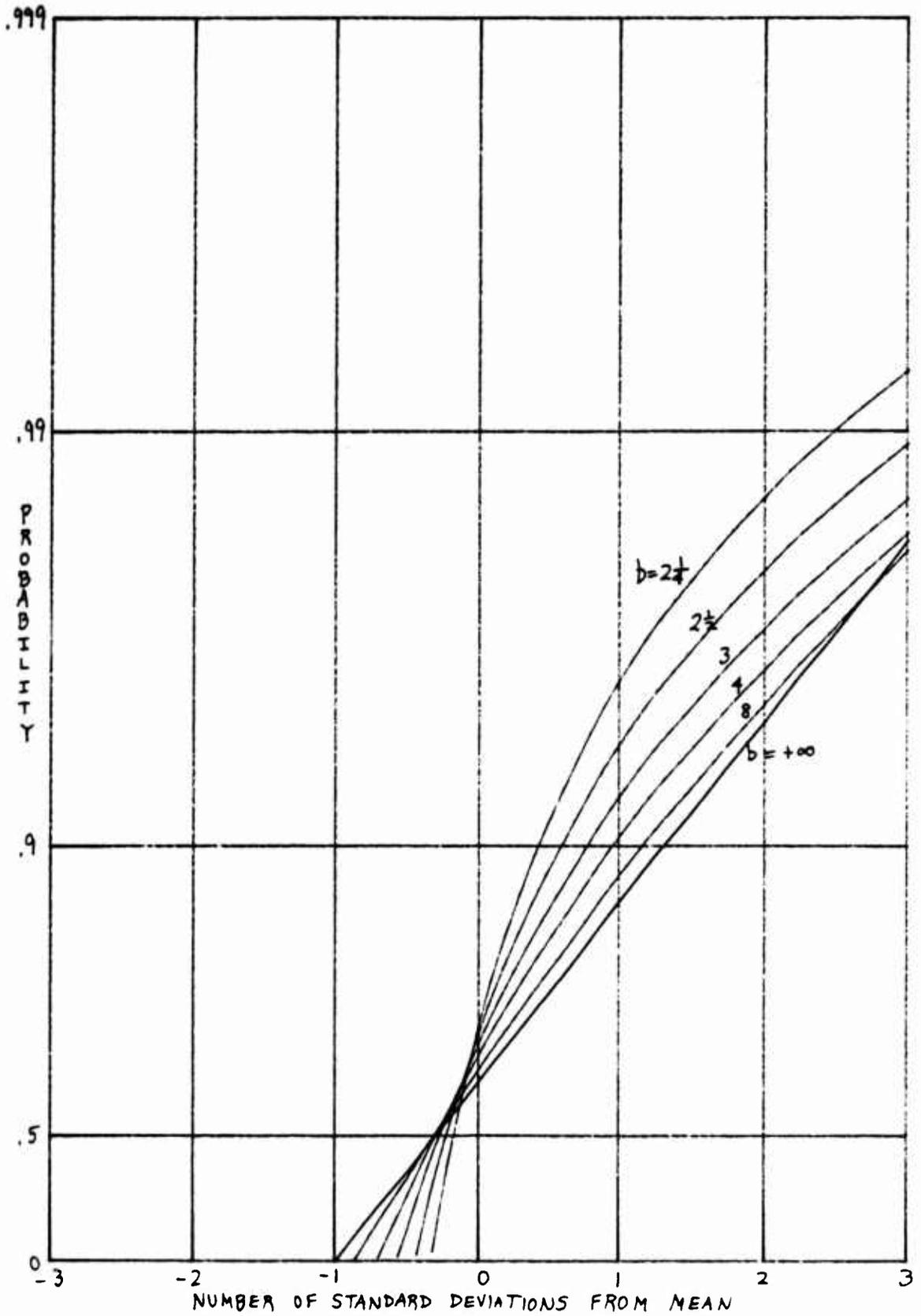Figure 2. Plots for Linear Abscissa and Logarithmic Ordinate

24

Figure 2C. Power-Law PDF

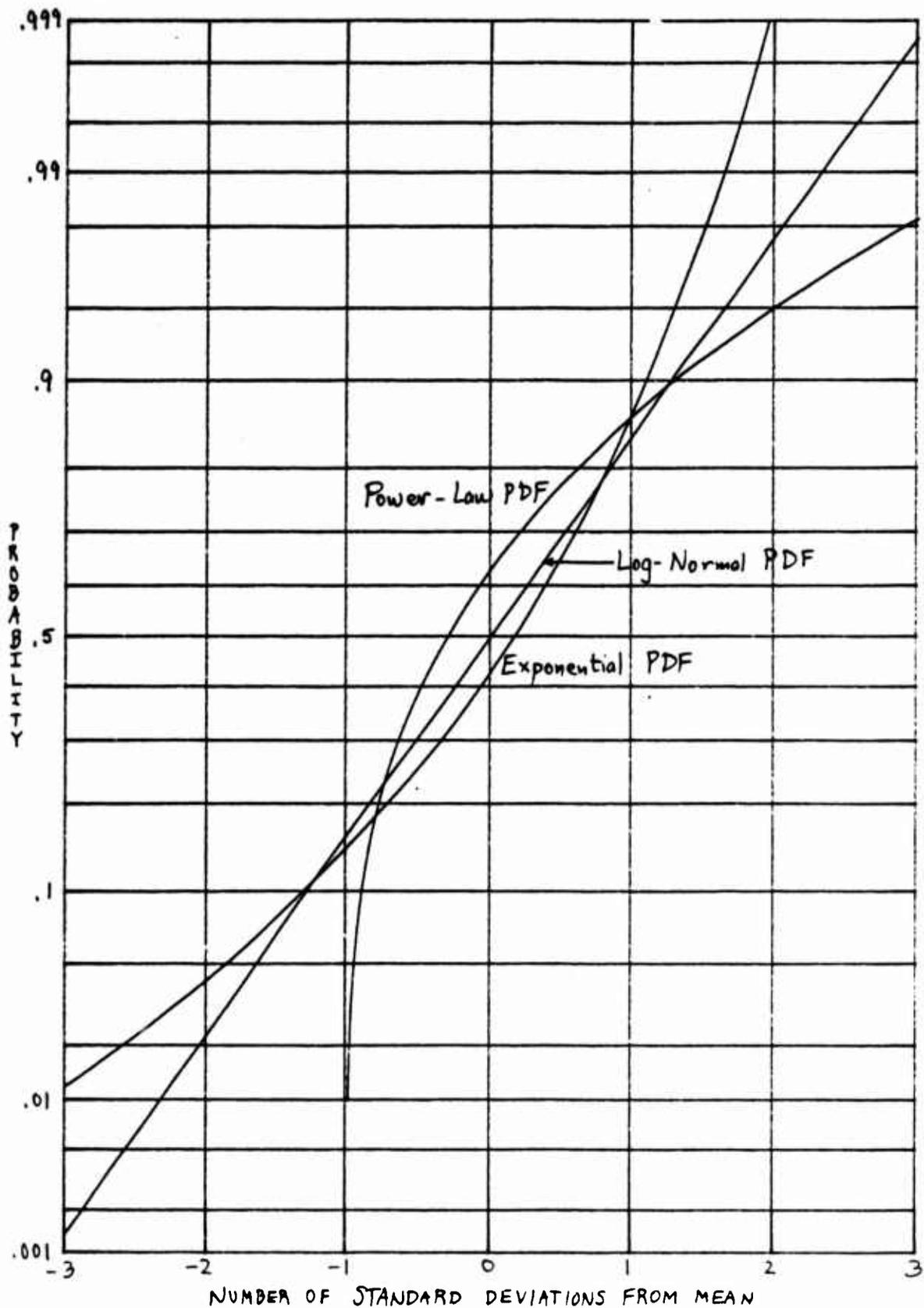Figure 2. Plots for Linear Abscissa and Logarithmic Ordinate

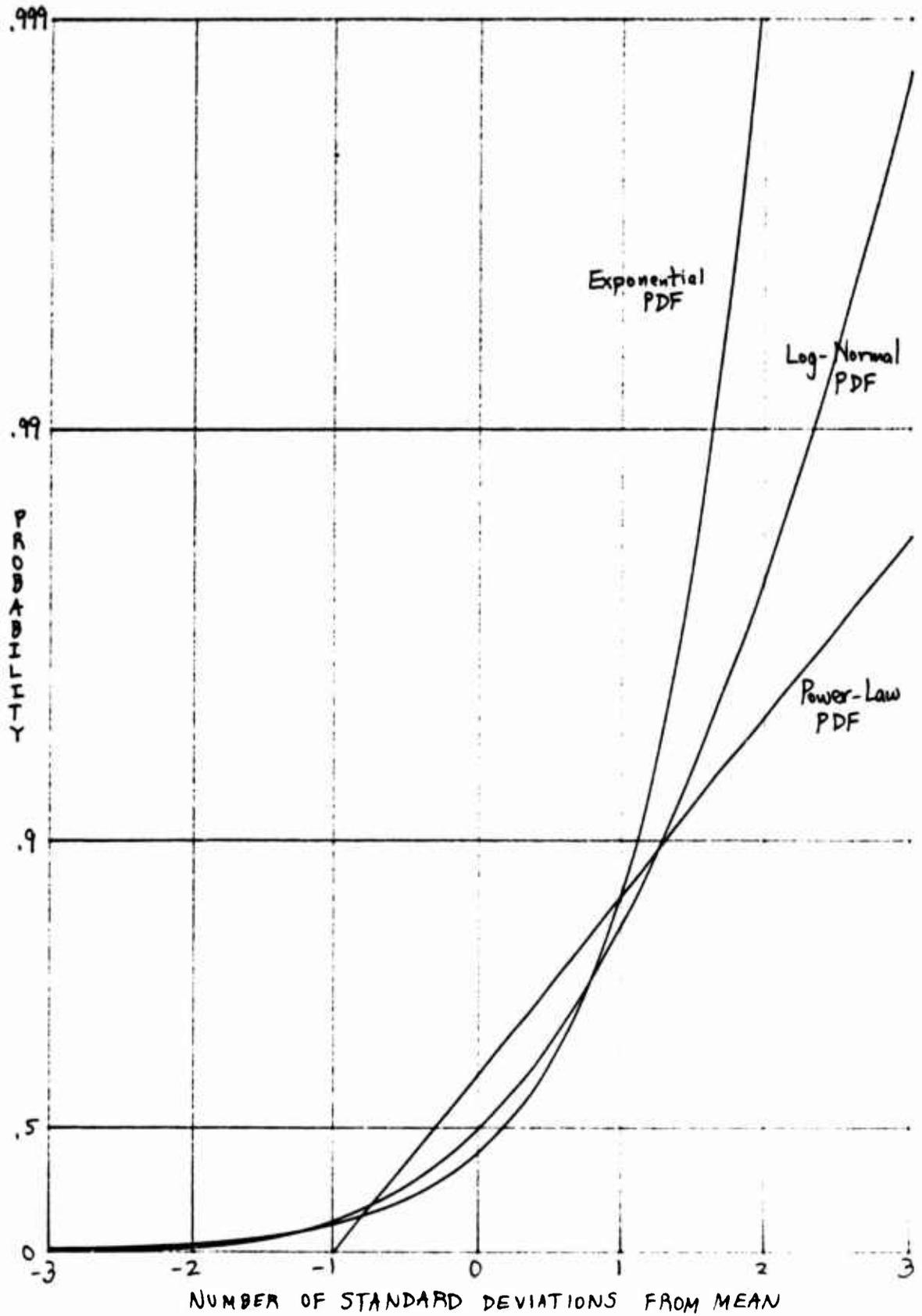Figure 3. Plots for Logarithmic Abscissa and Gaussian Ordinate

Figure 4. Plots for Logarithmic Abscissa and Logarithmic Ordinate

PLOTS FOR Q_M DISTRIBUTION

The last analytical example we consider is the Q_M distribution with a CDF given by (Ref. 3)

$$P_r(u) = 1 - Q_M\left(d, \sqrt{2au}\right) \quad \text{for } u \geq 0, \tag{52}$$

and PDF

$$p_r(u) = a\left(\frac{au}{d^2/2}\right)^{\frac{M-1}{2}} \exp\left(-\frac{d^2}{2} - au\right) I_{M-1}\left(d\sqrt{2au}\right) \quad \text{for } u \geq 0. \tag{53}$$

For M = 1, (53) is closely related to the Rice PDF. On the other hand, for d = 0, it reduces to

$$\frac{a(au)^{M-1} \exp(-au)}{(M-1)!} \quad \text{for } u \geq 0, \tag{54}$$

which is the chi-squared PDF with 2M degrees of freedom. Thus, (53) is a general form which subsumes several common cases.

The characteristic function of RV r is (Ref. 2, 6.631 4)

$$f_r(\zeta) = \int du \, e^{i\zeta u} p_r(u) = \left(1 - i\zeta/a\right)^{-M} \exp\left(\frac{d^2}{2} \frac{i\zeta/a}{1 - i\zeta/a}\right). \tag{55}$$

Therefore

$$\ln f_r(\zeta) = -M \ln\left(1 - \frac{i\zeta}{a}\right) + \frac{d^2}{2} \frac{i\zeta/a}{1 - i\zeta/a}$$

$$= M \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{i\zeta}{a}\right)^n + \frac{d^2}{2} \sum_{n=1}^{\infty}\left(\frac{i\zeta}{a}\right)^n. \tag{56}$$

The cumulants of RV r are immediately available:

$$\lambda_r(n) = \frac{(n-1)!}{a^n}\left(M + n\frac{d^2}{2}\right) \quad \text{for} \quad n \geq 1. \tag{57}$$

In particular,

$$m_r = \lambda_r(1) = \frac{1}{a}\left(M + \frac{d^2}{2}\right),$$

$$\sigma_r^2 = \lambda_r(2) = \frac{1}{a^2}\left(M + d^2\right). \tag{58}$$

Then from (3) (on a linear abscissa),

$$X = \frac{au - M - \frac{d^2}{2}}{\sqrt{M + d^2}}, \tag{59}$$

while from (5) and (52),

$$y = g\left\{1 - Q_M\left(d, \left[2M + d^2 + 2\sqrt{M+d^2}\, X\right]^{1/2}\right)\right\} \quad \text{for} \quad X \geq -\frac{2M + d^2}{2\sqrt{M+d^2}}. \tag{60}$$

This result is independent of scaling parameter a in (52) and (53); however, it depends on both M and d as indicated.

Equation (60) is plotted on a linear abscissa and Gaussian ordinate for M = 1, 2, 4, 10 in the four parts of figure 5, as parameter d takes on the values 0, 1, 2, 4, 8. The curve labelled d = $+\infty$ is the same as that for a Gaussian PDF. As d varies from 0 to $+\infty$, the CDF sweeps out the region between the chi-squared PDF (given in (54)) and the Gaussian PDF. Also, for a given value of d, larger M values result in a PDF curve that is more nearly Gaussian.
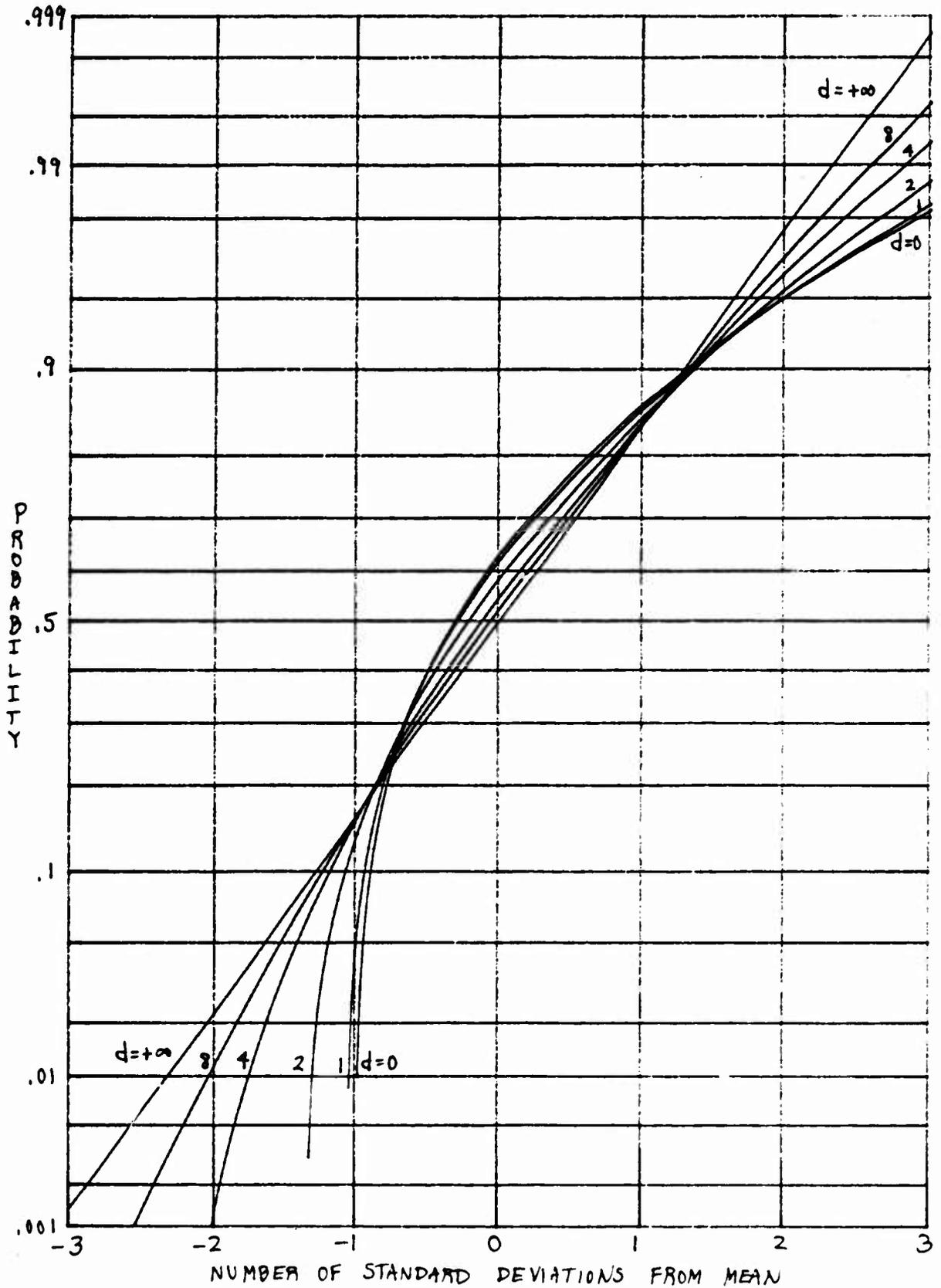
Figure 5A. M = 1

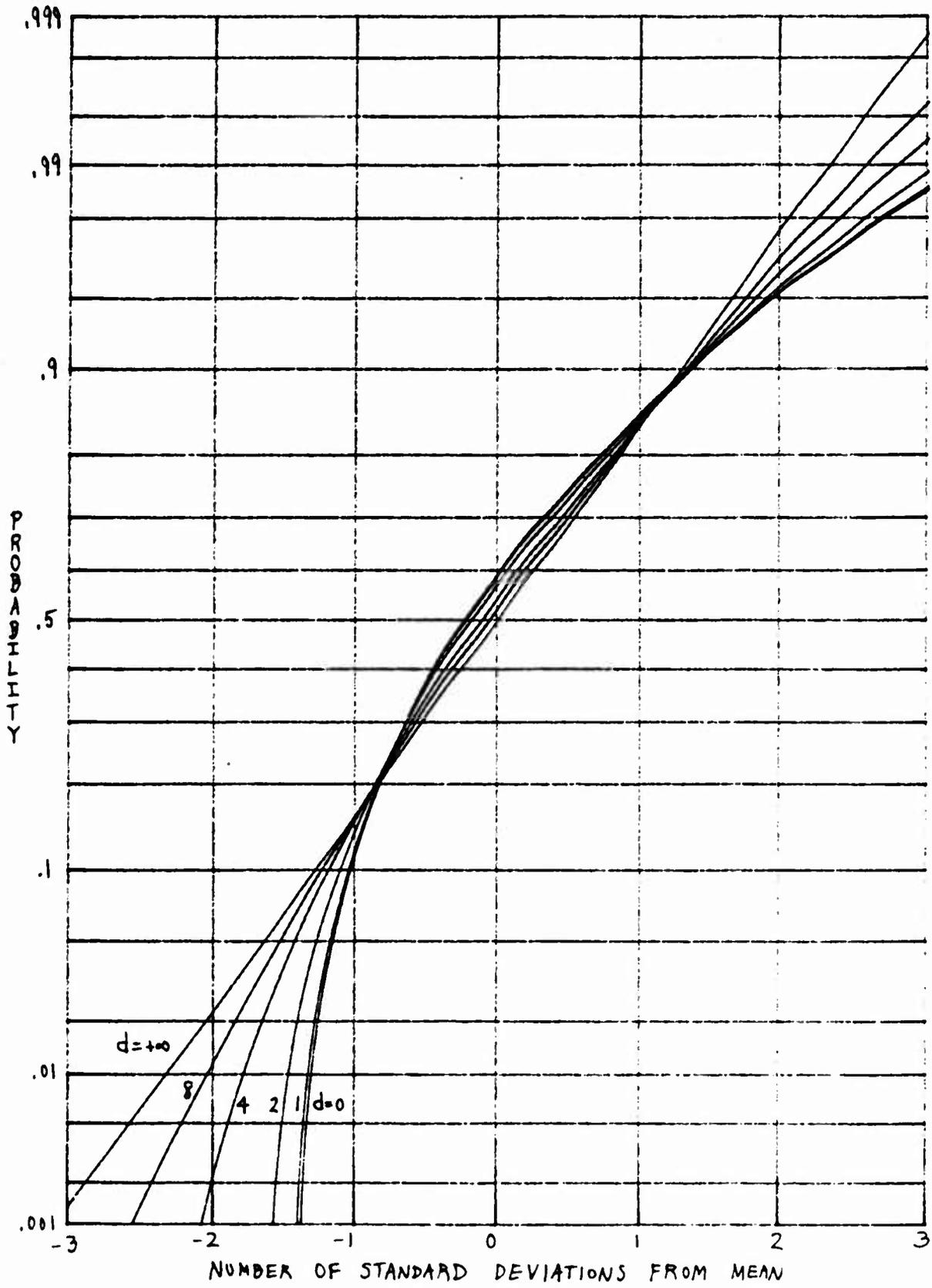Figure 5. Q_M Distribution on Linear Abscissa and Gaussian Ordinate

Figure 5B. M = 2

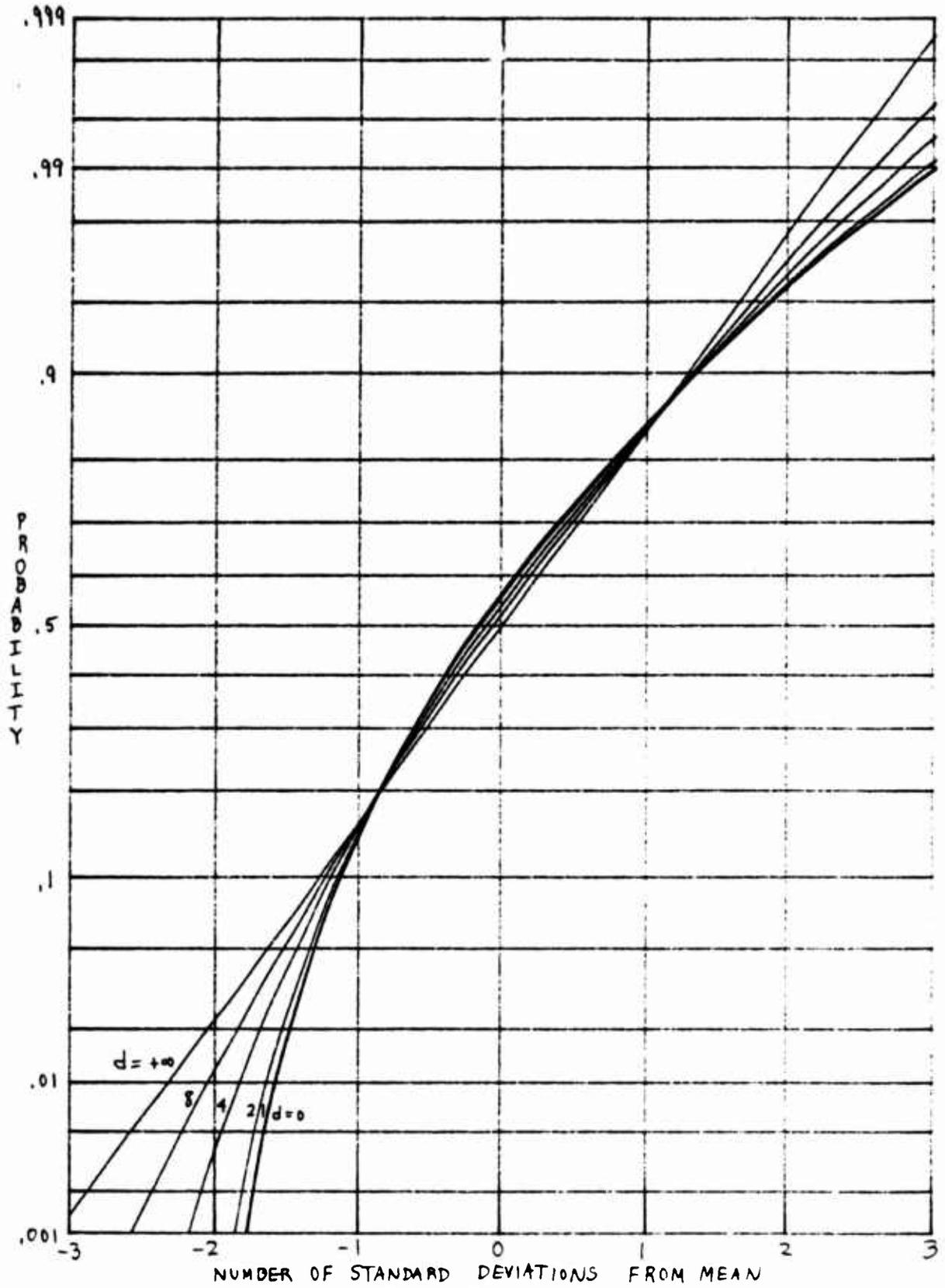Figure 5. Q_M Distribution on Linear Abscissa and Gaussian Ordinate

Figure 5C. M = 4

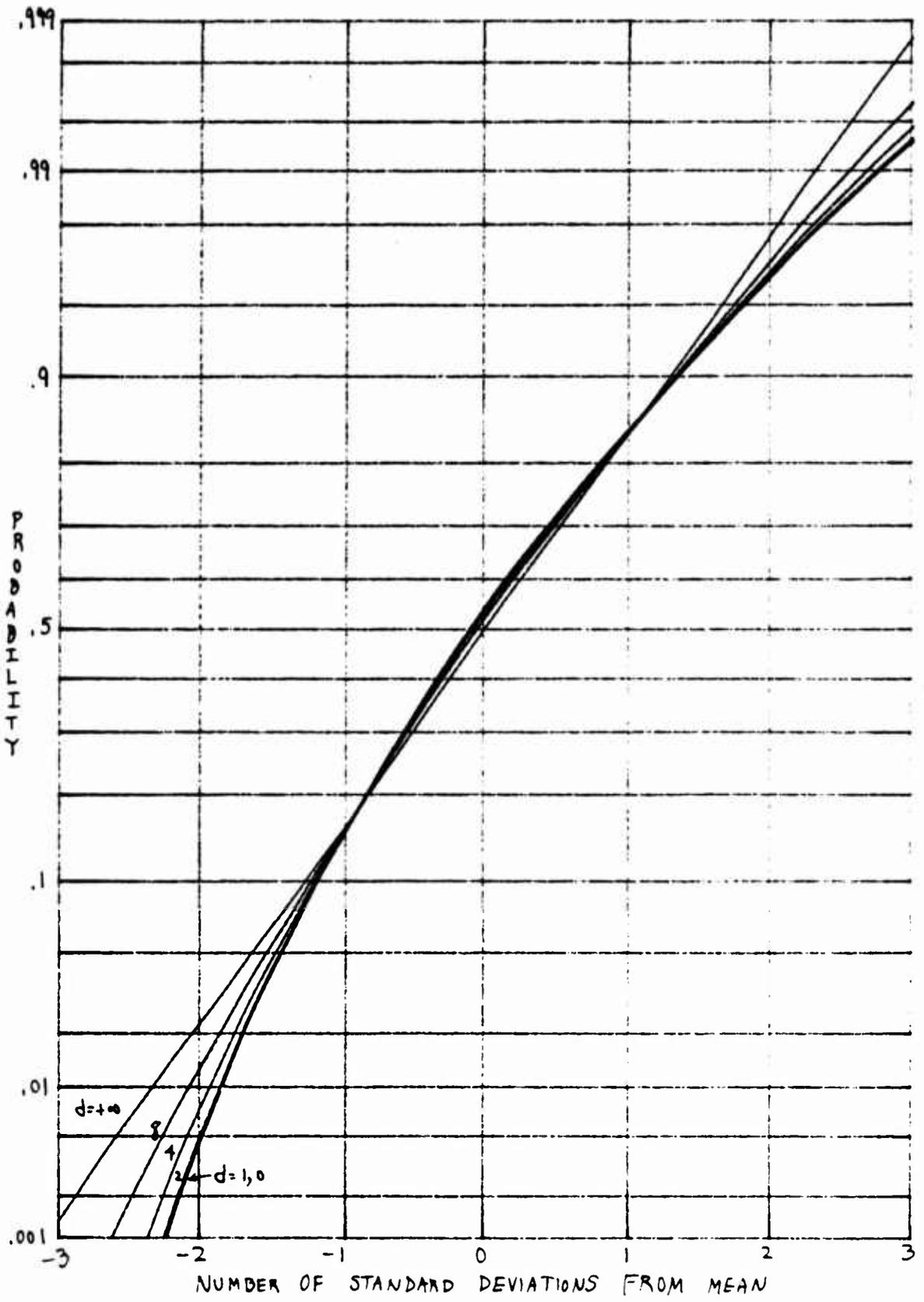Figure 5. Q_M Distribution on Linear Abscissa and Gaussian Ordinate

Figure 5D. M = 10

Figure 5. Q_M Distribution on Linear Abscissa and Gaussian Ordinate

RAYLEIGH PDF

The PDF and CDF for a Rayleigh RV r are

$$p_r(u) = \frac{u}{b^2} \exp\left(-\frac{u^2}{2b^2}\right) \quad \text{for} \quad u \geq 0,$$

$$P_r(v) = 1 - \exp\left(-\frac{u^2}{2b^2}\right) \quad \text{for} \quad u \geq 0. \tag{61}$$

The n-th moment of r is

$$\overline{r^n} = b^n \, 2^{\frac{n}{2}} \, \Gamma\left(\frac{n}{2}+1\right), \tag{62}$$

leading to (for a linear abscissa)

$$m_r = b\sqrt{\frac{\pi}{2}} \;, \quad \sigma_r = b\sqrt{2-\frac{\pi}{2}} \;, \quad x = \frac{u - b\sqrt{\frac{\pi}{2}}}{b\sqrt{2-\frac{\pi}{2}}} \;. \tag{63}$$

Then (61) and (5) yield

$$y = g\left\{1 - \exp\left(-\frac{\pi}{4}\left(1+\sqrt{\frac{4}{\pi}-1}\,x\right)^2\right)\right\} \quad \text{for} \quad x \geq -\sqrt{\frac{\pi}{4-\pi}} \;. \tag{64}$$

This result is independent of scaling parameter b in (61).

If $g\{\}$ is taken as the logarithmic transformation given by (43), then (64) takes the special form

$$y = \frac{\pi}{4}\left(1+\sqrt{\frac{4}{\pi}-1}\,x\right)^2 \quad \text{for} \quad x \geq -\sqrt{\frac{\pi}{4-\pi}} \;. \tag{65}$$

That is, the Rayleigh PDF plots as a parabola on a linear abscissa and logarithmic ordinate.

For a logarithmic abscissa, (19), (21), and (61) yield the CDF for RV t as

$$P_t(v) = 1 - \exp\left(-\frac{1}{2b^2} e^{2v}\right) \quad \text{for all } v. \tag{66}$$

The n-th moment of RV t is, according to (23) and (61),

$$\overline{t^n} = \int_0^\infty du \, (\ln u)^n \, \frac{u}{b^2} \exp\left(-\frac{u^2}{2b^2}\right)$$

$$= \int_0^\infty ds \, e^{-s} \left[\ln\left(\sqrt{2}\,b\right) + \frac{1}{2}\ln s\right]^n. \tag{67}$$

There follows (in a fashion similar to (30))

$$m_t = \ln\left(\sqrt{2}\,b\right) - \frac{1}{2}\gamma, \quad \sigma_t = \frac{\pi}{2\sqrt{6}}, \quad x = \frac{v - \ln\left(\sqrt{2}\,b\right) + \frac{1}{2}\gamma}{\frac{\pi}{2\sqrt{6}}}. \tag{68}$$

Then (26) and (66) yield

$$y = g\left\{1 - \exp\left(-e^{x\frac{\pi}{\sqrt{6}} - \gamma}\right)\right\} \quad \text{for all } x, \tag{69}$$

which is independent of b.

But this result is identical to (32) for an exponential PDF on a logarithmic abscissa. The reason for this is that the Rayleigh RV r here is related to the exponential RV $\tilde{r}$ in (27) according to (using parameter $\tilde{b}$ for (27))

$$r = \beta\sqrt{\tilde{r}} \quad \text{where} \quad \beta = b\sqrt{2\tilde{b}}. \tag{70}$$

Thus

$$t = \ln r = \ln \beta + \tfrac{1}{2} \ln \tilde{r} = \ln \beta + \tfrac{1}{2} \tilde{t} . \qquad (71)$$

That is, logarithmically-transformed variates t and $\tilde{t}$ are linearly related to each other, and the normalization in (25) will treat these two quantities equally.

More generally, for a RV $\tilde{r}$ with an arbitrary PDF, consider the non-linearly related RV

$$r = \alpha \, \tilde{r}^{\nu} . \qquad (72)$$

Then the logarithmically-transformed RVs are related according to

$$t = \ln r = \ln \alpha + \nu \ln \tilde{r} = \ln \alpha + \nu \tilde{t} . \qquad (73)$$

But transformation (25) will result in exactly the same plot for CDF $P_t(\ )$ as for $P_{\tilde{t}}(\ )$. This holds for a logarithmic abscissa, regardless of what ordinate transformation $g\{\ \}$ is employed. (The analoguous situation for a linear abscissa is for the RVs to be related according to $r = \alpha + \beta \tilde{r}$).

SIMULATION RESULTS

N independent Gaussian RVs $\left\{r_k\right\}_1^N$ with an arbitrary mean and standard deviation were generated (Ref. 4). The sample mean $\hat{m}$ and sample standard deviation $\hat{\sigma}$ of the set was computed, and each RV was modified according to

$$\hat{r}_k = \frac{r_k - \hat{m}}{\hat{\sigma}} \quad \text{for} \quad 1 \leq k \leq N, \tag{74}$$

giving a new set with zero sample mean and unit sample standard deviation. Then the new set $\left\{\hat{r}_k\right\}_1^N$ was ordered according to size, from most-negative to most-positive, giving set $\left\{\overset{\vee}{r}_k\right\}_1^N$. The sample CDF was then realized by setting it to value $(k - \frac{1}{2})/N$ at abscissa $\overset{\vee}{r}_k$, for $1 \leq k \leq N$, and drawing straight lines between these points. The results for a linear abscissa and Gaussian ordinate are given as the jagged curves in figure 6, with N = 100 in figure 6A, and N = 1000 in figure 6B. The straight line overlaid on the plots is the Gaussian PDF result.

When scaled and shifted exponential RVs are generated and subjected to the same procedure, the results in figure 7 are obtained. The smooth curve overlaid on the jagged sample CDF plots is the exponential PDF as given in figure 1A. The results in figures 6 and 7 allow for ready confirmation of the character of the sample CDF. A sample program for the procedure is given in the appendix.
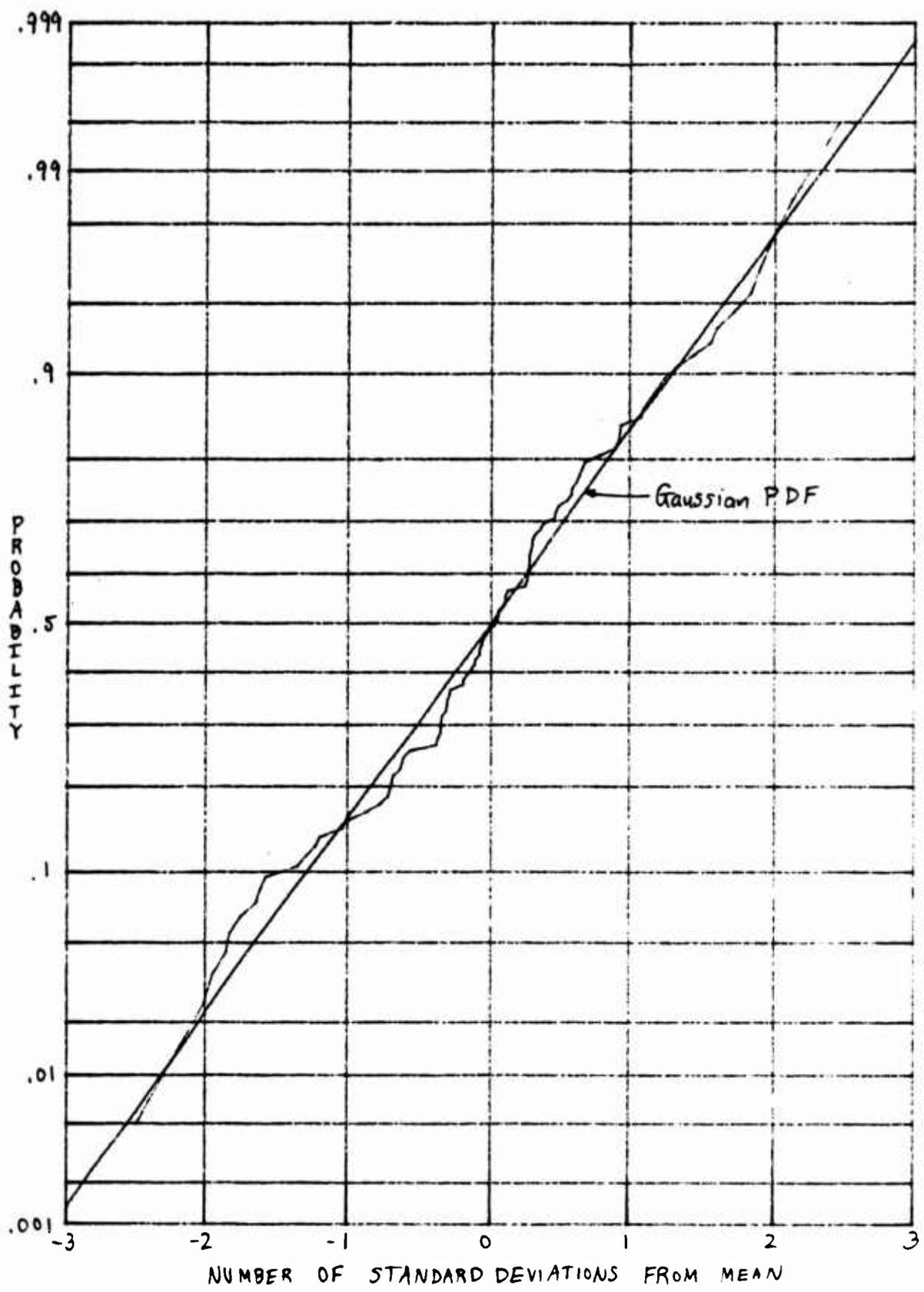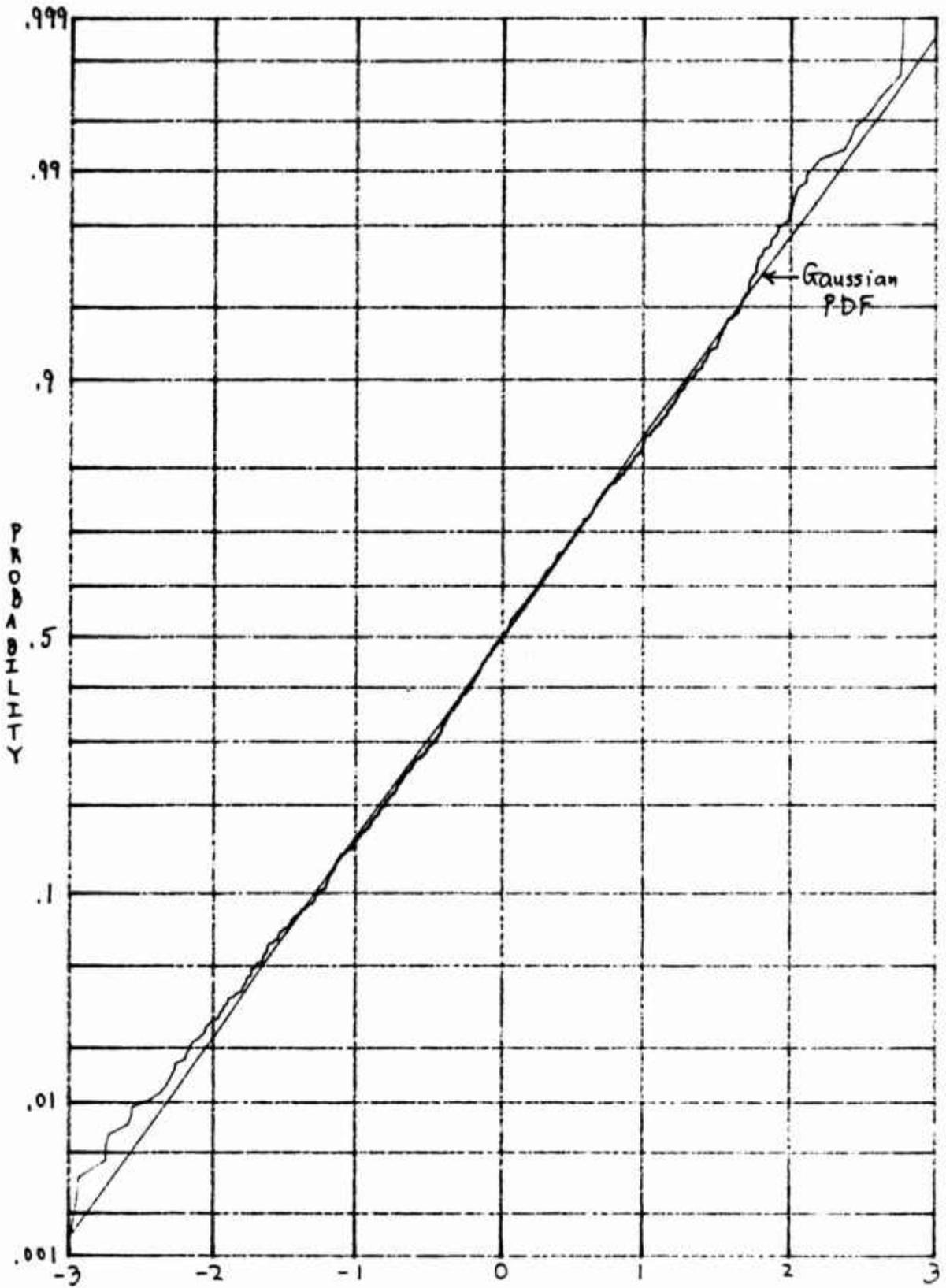
Figure 6A. N = 100

Figure 6. Simulation Results for a Gaussian Random Variable

Figure 6B. N = 1000

Figure 6. Simulation Results for a Gaussian Random Variable
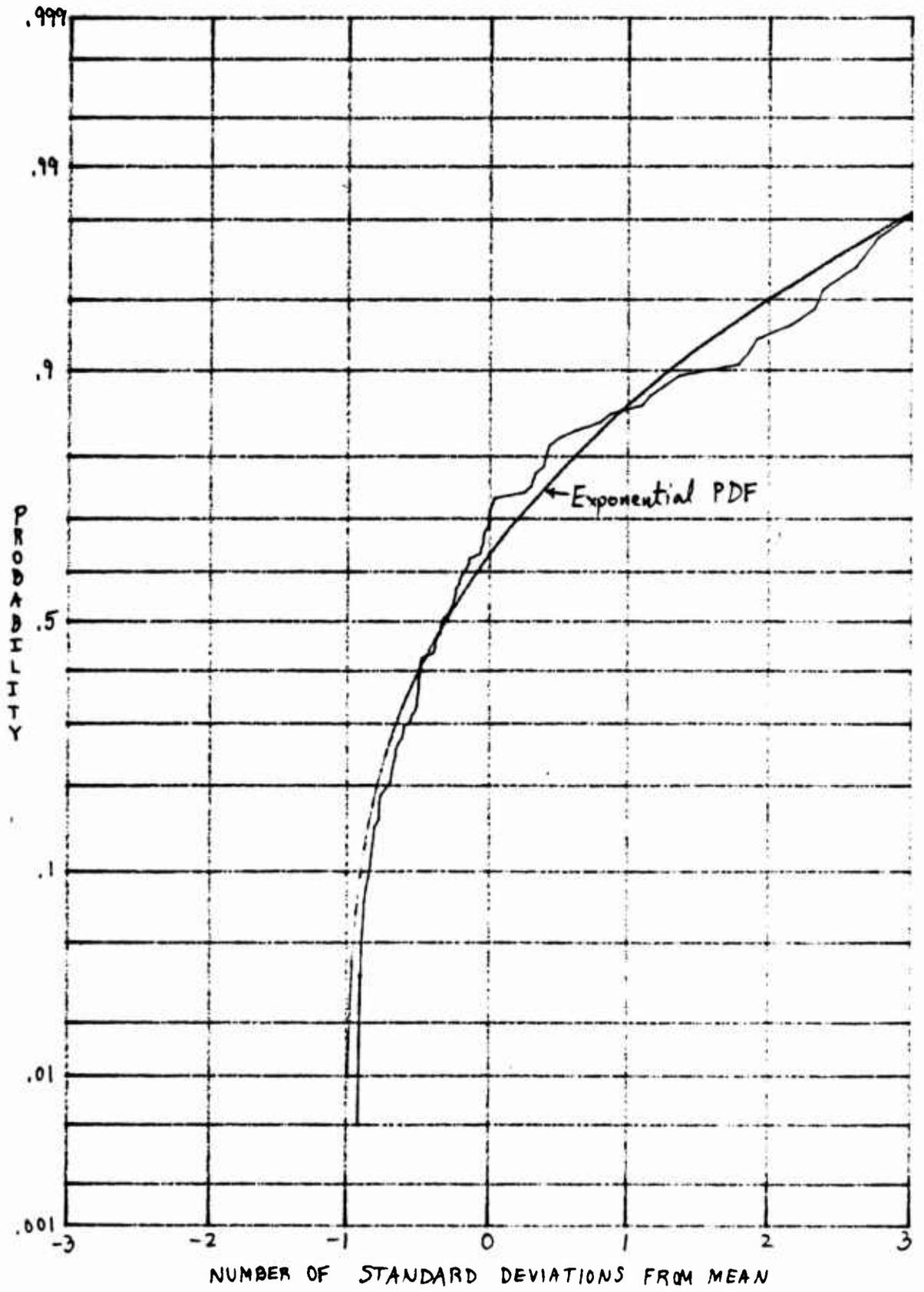
NUMBER OF STANDARD DEVIATIONS FROM MEAN

Figure 7A. N = 100

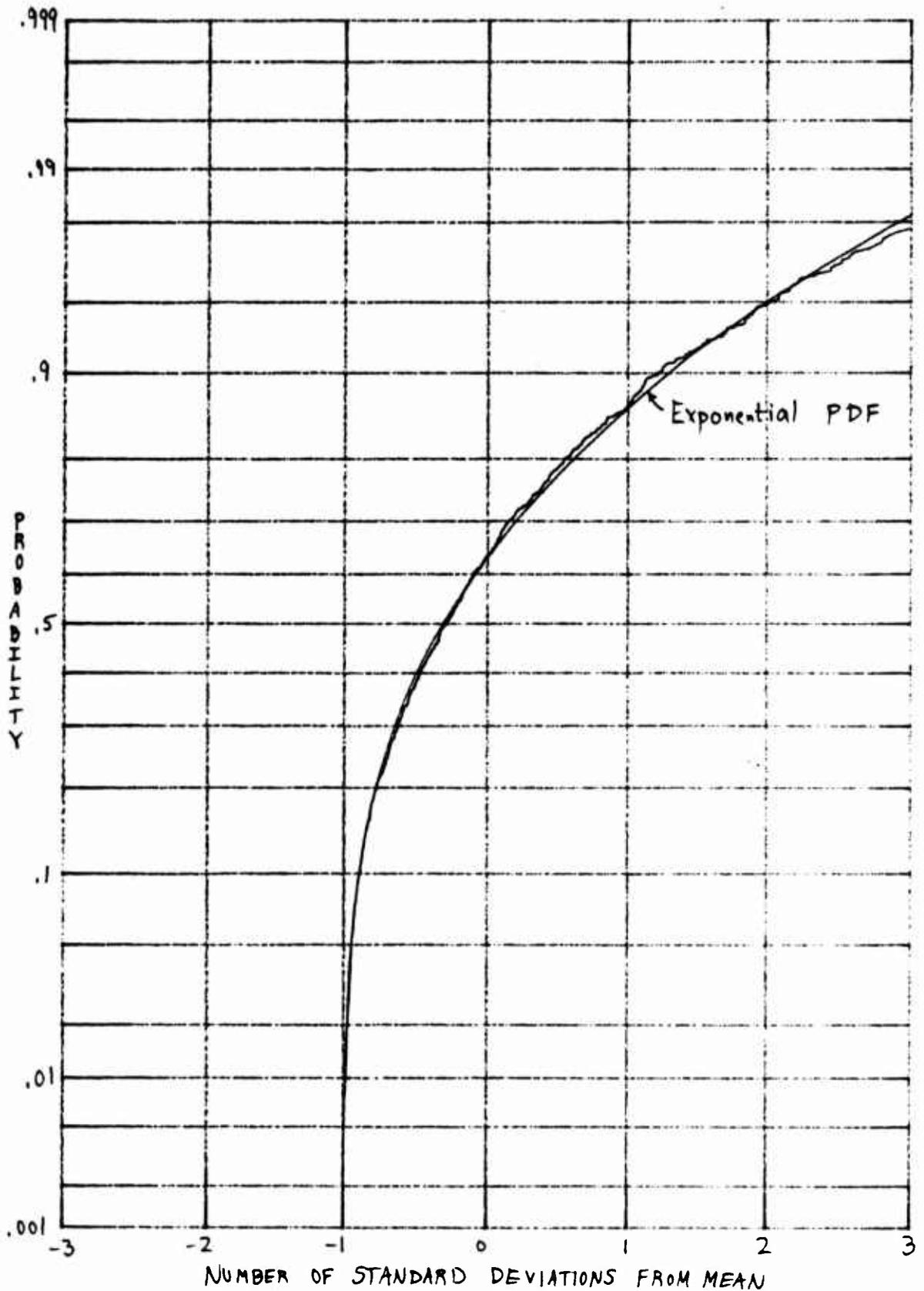Figure 7. Simulation Results for an Exponential Random Variable

Figure 7B. N = 1000

Figure 7. Simulation Results for an Exponential Random Variable

DISCUSSION

Additional candidate PDFs or CDFs can be added to those given herein, thereby building up a catalog of comparison cases. Also, other ordinate transformations, like the arc tanh ($\sqrt{P}$) function utilized in coherence transformation (Ref. 5), can be included. In this way, an experimental sample CDF can be readily compared with several ideal analytical forms, in an attempt to easily find a reasonable characterization for the statistics.

APPENDIX

PROGRAM FOR SIMULATION RESULTS

```
10      RANDOMIZE SQR(.6)              ! RANDOM NUMBER SEED
20      N=100                         ! NUMBER OF SAMPLES
30      DIM Y(1:1000),B(1:1000)
40      REDIM Y(1:N),B(1:N)
50      FOR I=1 TO N
60      Y(I)=-LOG(RND)*3.1+.57        ! EXPONENTIAL PDF,
70      NEXT I                        ! SCALED AND SHIFTED
80      FOR I=1 TO N
90      S1=S1+Y(I)
100     S2=S2+Y(I)^2
110     NEXT I
120     Mean=S1/N
130     Var=(S2-S1*S1/N)/(N-1)
140     Stddev=SQR(Var)
150     FOR I=1 TO N
160     Y(I)=(Y(I)-Mean)/Stddev       ! NORMALIZE
170     NEXT I
180     FOR K=1 TO N                  ! SORT INTO B ARRAY
190     A=1E30
200     J=0
210     FOR I=1 TO N
220     IF Y(I)>A THEN 250
230     A=Y(I)
240     J=I
250     NEXT I
260     B(K)=A
270     Y(J)=9E99
280     NEXT K
290     PLOTTER IS "GRAPHICS"
300     GRAPHICS
310     X1=-3
320     X2=3
330     Y1=FNInvphi(.001)             ! GAUSSIAN ORDINATE
340     Y2=FNInvphi(.999)
350     SCALE X1,X2,Y1,Y2
360     FOR X=X1 TO X2
370     MOVE X,Y1
380     DRAW X,Y2
390     NEXT X
400     PENUP
```

43

```
410    DATA .001,.002,.005,.01,.02,.05,.1,.2,.3,.4,.5
420    DATA .6,.7,.8,.9,.95,.98,.99,.995,.998,.999
430    DIM O(1:21)
440    READ O(*)
450    FOR I=1 TO 21
460    T=FNInvphi(O(I))
470    MOVE X1,T
480    DRAW X2,T
490    NEXT I
500    PENUP
510    FOR X=-.99 TO X2+.03 STEP .03
520    Y=FNInvphi(1-EXP(-1-X))    ! EXPONENTIAL CDF
530    PLOT X,Y
540    NEXT X
550    PENUP
560    FOR I=1 TO N
570    Y=FNInvphi((I-.5)/N)        ! SAMPLE CDF ORDINATE
580    PLOT B(I),Y
590    NEXT I
600    PENUP
610    END
620    !
630    DEF FNInvphi(X)             ! INVERSE PHI FUNCTION
640    IF (X>=0) AND (X<=1) THEN 670
650    PRINT "ARGUMENT ";X;"IS DISALLOWED"
660    STOP
670    IF (X>0) AND (X<1) THEN 700
680    P=9.99999999999E99*(2*X-1)
690    GOTO 760
700    P=X
710    IF X>.5 THEN P=.5-(X-.5)
720    P=SQR(-2*LOG(P))
730    T=1+P*(1.432788+P*(.189269+P*.001308))
740    P=P-(2.515517+P*(.802853+P*.010328))/T
750    IF X<.5 THEN P=-P
760    RETURN P
770    FNEND
```

REFERENCES

1. Handbook of Mathematical Functions, U.S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series No. 55, U.S. Government Printing Office, June 1964.

2. I. S. Gradshteyn and I. W. Ryzhik, <u>Table of Integrals, Series, and Products</u>, Academic Press, New York, 1965.

3. A. H. Nuttall, "Some Integrals Involving the Q_M-Function", NUSC Technical Report 4755, 15 May 1974.

4. A. H. Nuttall, "Some Techniques for Generation of Random Numbers with Specified Distributions or Densities", NUSC Technical Memorandum 811135, 1 October 1981.

5. A. H. Nuttall, "Statistics of Nonlinearly Transformed Coherence Estimates", NUSC Technical Report 6445, 12 March 1981.

On the Distribution
Of a Chi-Squared
Variate Raised
To a Power

A. H. Nuttall

ABSTRACT

The probability distribution of a chi-squared variate tends toward normality as the number of degrees of freedom increases. However, some powers of the chi-squared variate tend to normality much faster. For example, the 1/3 power of a chi-squared variate of just 4 degrees of freedom is virtually normal over the whole range (.001, .999) of probabilities. Inspection of the cumulants reveals that the 1/3 power is best for minimizing the third and fifth cumulants, and nearly optimum for the fourth cumulant.

INTRODUCTION

The approach of a chi-squared variate, z, to normality as the number of degrees of freedom, K, increases is well known. However, it has been observed that the square-root of z tends to normality faster; see [1, p. 79, par. 4], [2, pp. 251 and 420], and [3, pp. 371-2]. Furthermore, the cube root of z tends even faster to normality, based on consideration of the low order cumulants; see [3, pp. 371-4] and [4]. Here we will look at the cumulative distribution function and cumulants of z^ν for various values of power ν and degrees of freedom K, and furnish a quantitative measure of the discrepancy from the Gaussian distribution.

CHI-SQUARED VARIATE OF K DEGREES OF FREEDOM

Let x_k, for $1 \leq k \leq K$, be K independent Gaussian random variables with zero mean and unit variance. Then

$$z = \chi^2 = x_1^2 + x_2^2 + \ldots + x_K^2 \qquad (1)$$

is called a chi-squared variate of K degrees of freedom. The characteristic function of z is

$$f_z(\xi) = \overline{\exp(i\xi z)} = (1 - i2\xi)^{-K/2} \quad ; \qquad (2)$$

the probability density function of z is*

$$p_z(u) = \frac{u^{\frac{K}{2} - 1} \exp(-u/2)}{2^{K/2} \, \Gamma(K/2)} \quad \text{for } u > 0 \quad ; \qquad (3)$$

and the cumulative distribution function of z is [7, eq. 6.5.5 and 26.4.1]

$$P_z(u) = \text{Prob}\{z < u\} = \int_0^u dt \; \frac{t^{\frac{K}{2} - 1} \exp(-t/2)}{2^{K/2} \, \Gamma(K/2)} = P(u|K) \quad \text{for} \quad u > 0. \qquad (4)$$

The cumulative distribution function is also expressable in terms of the incomplete gamma function and confluent hypergeometric function (for non-integer K) as [7, eq. 6.5.2]

$$P_z(u) = \gamma\left(\frac{K}{2}, \frac{u}{2}\right) \Big/ \Gamma\left(\frac{K}{2}\right)$$

$$= \left(\frac{u}{2}\right)^{\frac{K}{2}} \exp\left(-\frac{u}{2}\right) {}_1F_1\left(1; \; 1 + \frac{K}{2}; \; \frac{u}{2}\right) \Big/ \Gamma\left(\frac{K}{2} + 1\right). \qquad (5)$$

---

* The probability density functions and cumulative distribution functions encountered here are zero for negative arguments, except for Φ in (8).

The general ν-th moment of chi-squared variate z is

$$\overline{z^\nu} = 2^\nu \frac{\Gamma\left(\frac{K}{2} + \nu\right)}{\Gamma\left(\frac{K}{2}\right)} \quad \text{for} \quad \nu > -\frac{K}{2} \quad , \tag{6}$$

and the cumulants are

$$\lambda_z(n) = K \, 2^{n-1}(n-1)! \quad \text{for} \quad n = 1, 2, \ldots \tag{7}$$

Special cases of (4) are

$$P(u|1) = 2\Phi(\sqrt{u}) - 1, \quad \Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} dx \, \exp(-x^2/2),$$

$$P(u|2) = 1 - \exp(-u/2), \tag{8}$$

and more generally, we have recursion

$$P(u|K) = P(u|K-2) - \left(\frac{u}{2}\right)^{\frac{K}{2} - 1} \exp(-u/2)/\Gamma(K/2) \quad \text{for} \quad K \geq 3. \tag{9}$$

The function Φ in (8) is the cumulative distribution function for a zero-mean unit-variance Gaussian random variable.

ν-TH POWER OF A CHI-SQUARED VARIATE

Let random variable v be the ν-th power of the chi-squared variate:

$$v = z^\nu = \left(\chi^2\right)^\nu \quad , \tag{10}$$

where $\nu > 0$. The cumulative distribution function of v is, upon use of (4),

$$P_v(u) = \text{Prob}\{v < u\} = \text{Prob}\{z^\nu < u\} = \text{Prob}\{z < u^{1/\nu}\}$$

$$= P_z\left(u^{1/\nu}\right) = P\left(u^{1/\nu}\big|K\right) \quad , \tag{11}$$

in terms of the cumulative distribution function of a chi-squared variate.

The moments of v are, using (10) and (6),

$$\mu_n \equiv \overline{v^n} = \overline{z^{n\nu}} = 2^{n\nu}\,\frac{\Gamma\left(\frac{K}{2} + n\nu\right)}{\Gamma\left(\frac{K}{2}\right)} \quad . \tag{12}$$

In particular, the mean and standard deviation of v are

$$m_v \equiv \overline{v} = \mu_1, \quad \sigma_v = \left(\overline{v^2} - \overline{v}^2\right)^{1/2} = \left(\mu_2 - \mu_1^2\right)^{1/2} . \tag{13}$$

On a normalized abscissa x, we therefore have to consider the distribution [5, eq. 5]

$$P_v(m_v + \sigma_v x) = P\left((m_v + \sigma_v x)^{1/\nu}\big|K\right); \tag{14}$$

here we used (11).

Plots of cumulative distribution functions

For $\nu = 1$, v in (10) is simply the chi-squared variate of K degrees of freedom; its cumulative distribution function (14) is plotted in figure 1.
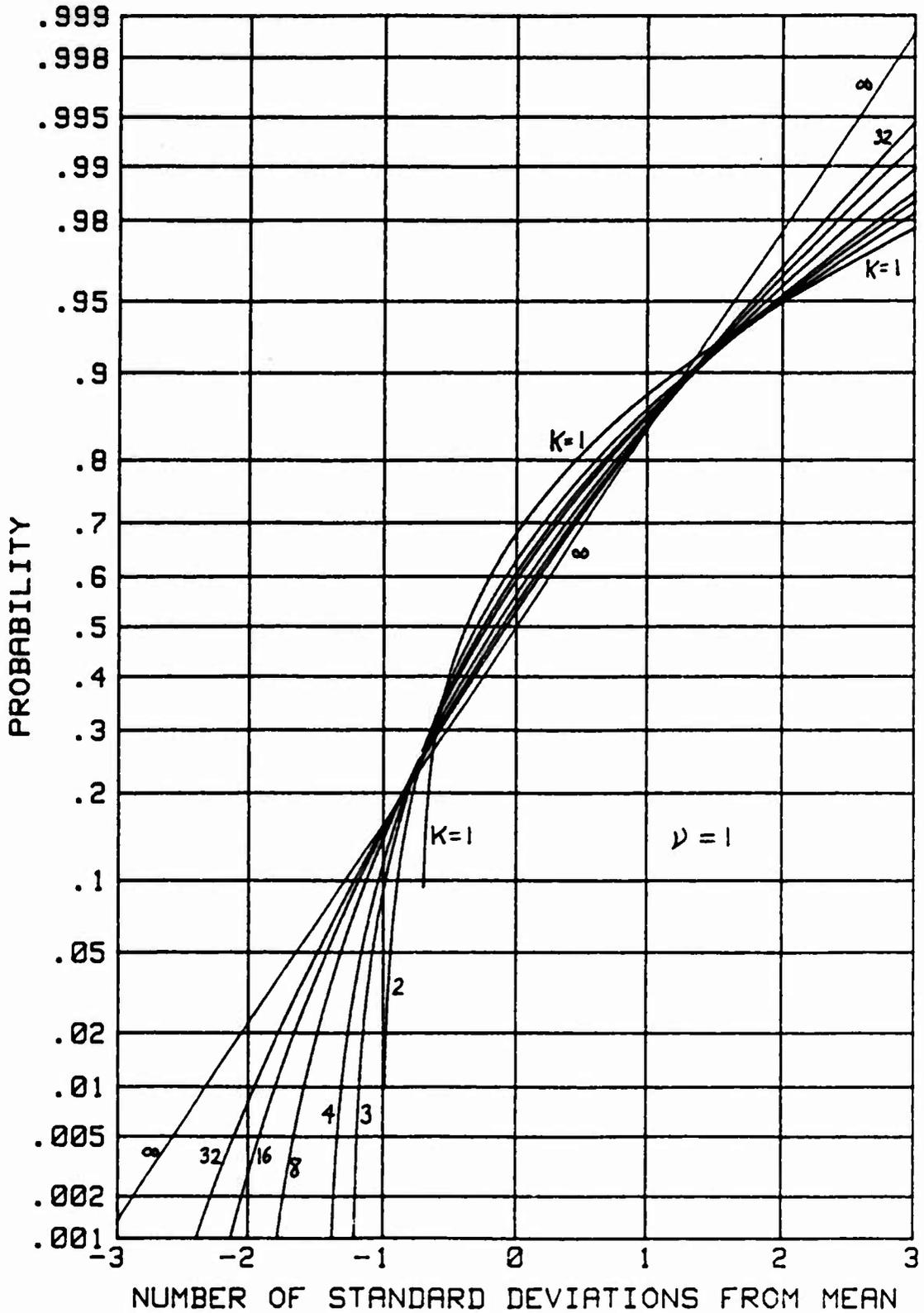
Figure 1. Distribution of Chi-Squared Variate Z

The curve labelled K=∞ is the Gaussian cumulative distribution function and is a straight line on this paper with a normal ordinate. The approach to this Gaussian curve progresses rather slowly with K, especially on the tails.

Plots of the cumulative distribution function of v for ν = 1/2, 1/3, 1/4, 1/5, as given by (14), are presented in figures 2-5, respectively. The improvement in approach to normality for ν = 1/2 in figure 2 over that for ν = 1 in figure 1 is significant. However, that for ν = 1/3 in figure 3 is remarkably good; in fact, the cumulative distribution functions for K $\geq$ 8 in figure 3 are virtually on top of the Gaussian curve in the entire range of probabilities (.001, .999).

When we continue with the trend toward smaller ν values in figures 4 and 5, the cumulative distribution functions begin to deviate further from the Gaussian curve, at both ends of the probability scale. For example, for ν = 1/5 in figure 5, the cumulative distribution functions are all above the Gaussian result, whereas for ν = 1/3, they were all below. All the cumulative distribution functions approach the Gaussian curve as K increases, but the approach is not monotonic with K.

Finally, for a ν value larger than 1, namely ν = 5/4, the plot in figure 6 reveals even greater discrepancy from the Gaussian curve than for the chi-squared variate itself, i.e., ν = 1 in figure 1.

Cumulants

Another measure of the non-Gaussian character of a random variable are the normalized cumulants. The cumulants of v [3, p. 70] are, in terms of (12),

$$\lambda_v(2) = u_2 - u_1^2 = \overline{(v-\overline{v})^2}$$

$$\lambda_v(3) = \overline{(v-\overline{v})^3} = u_3 - 3u_2 u_1 + 2u_1^3$$

$$\lambda_v(4) = \overline{(v-\overline{v})^4} - 3\lambda_v^2(2) = u_4 - 4u_3 u_1 - 3u_2^2 + 12u_2 u_1^2 - 6u_1^4$$

$$\lambda_v(5) = u_5 - 5u_4 u_1 - 10u_3 u_2 + 20u_3 u_1^2 + 30u_2^2 u_1 - 60u_2 u_1^3 + 24u_1^5 \quad . \tag{15}$$
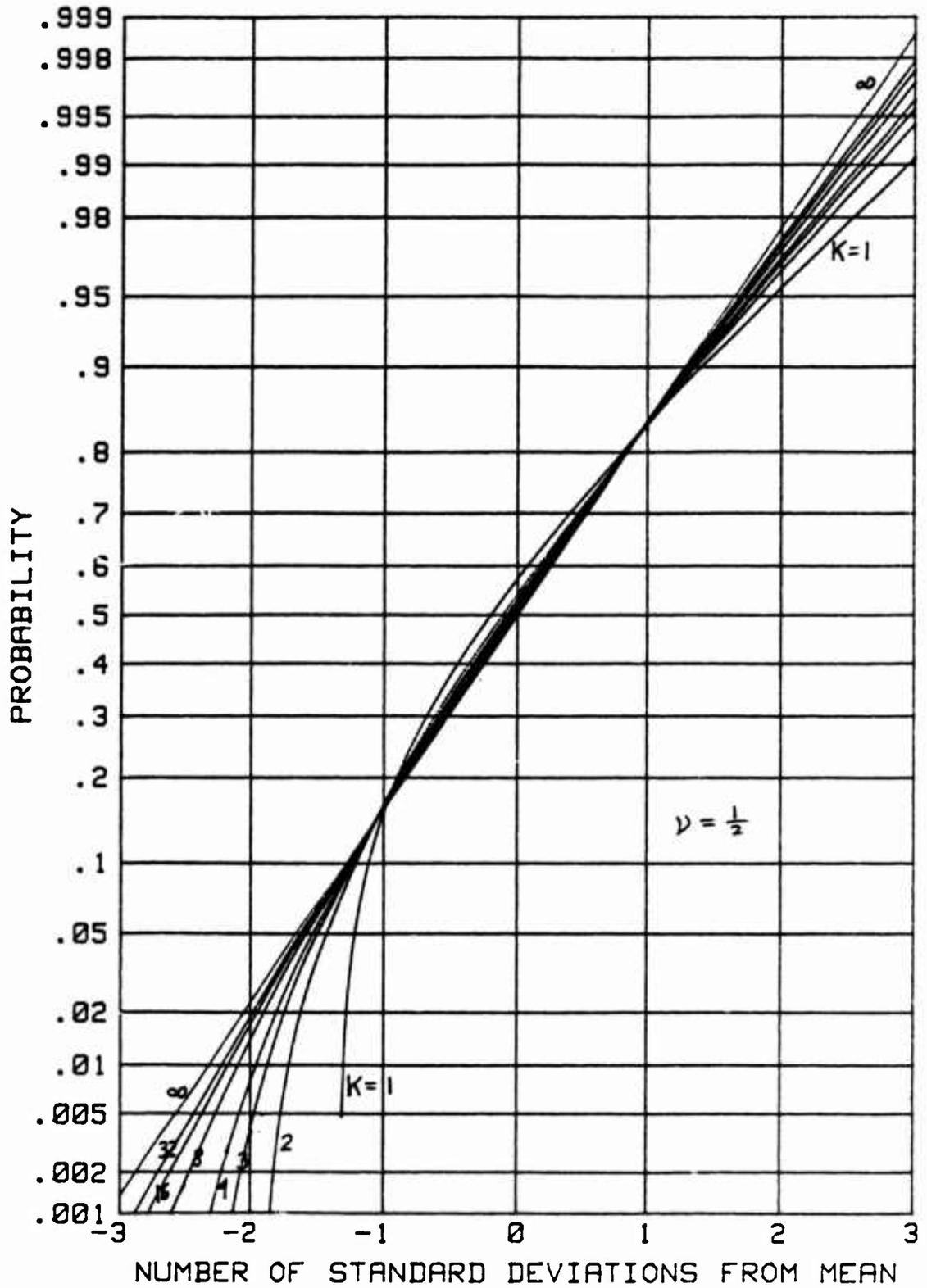
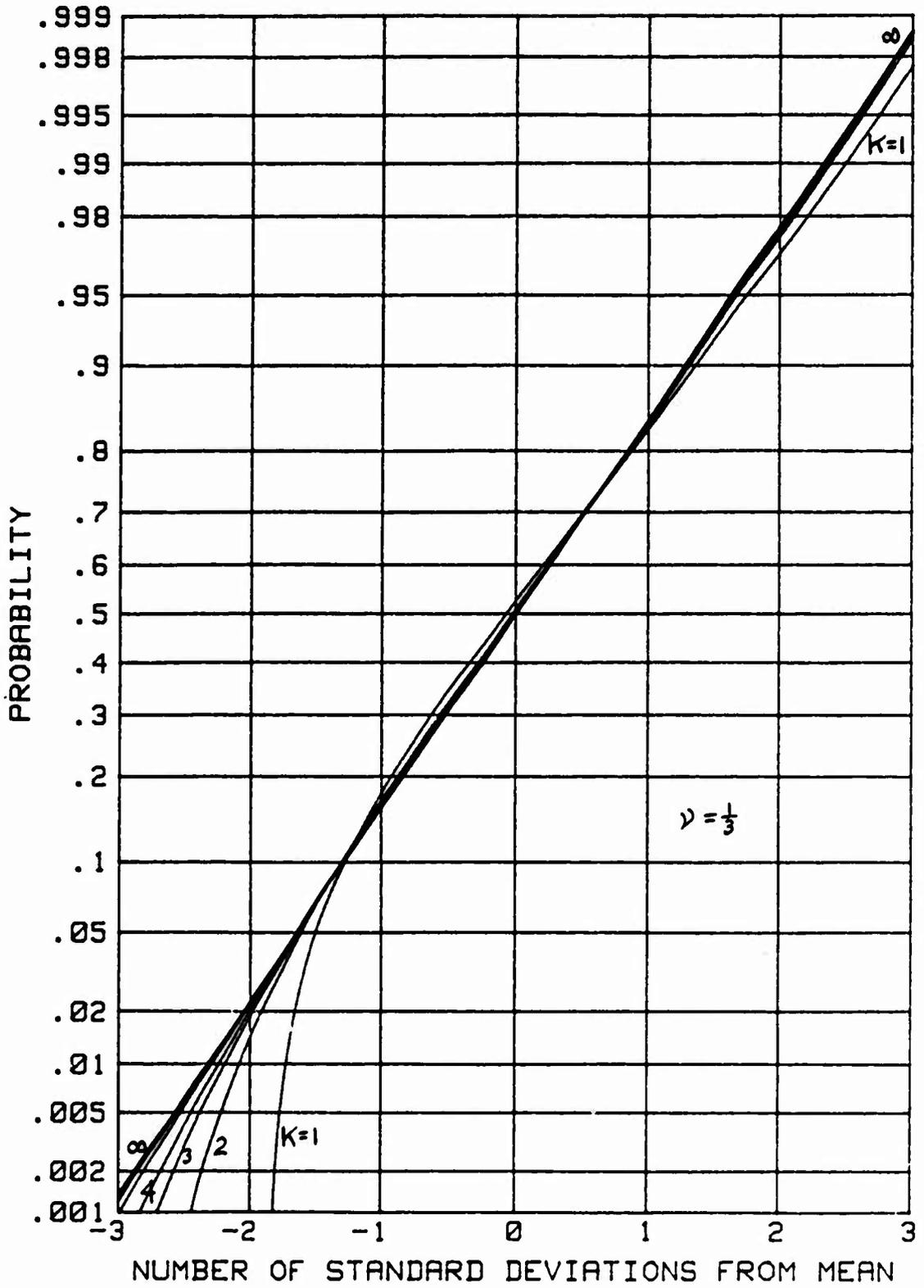Figure 2. Distribution of $v = z^{\nu}$ for $\nu = \frac{1}{2}$

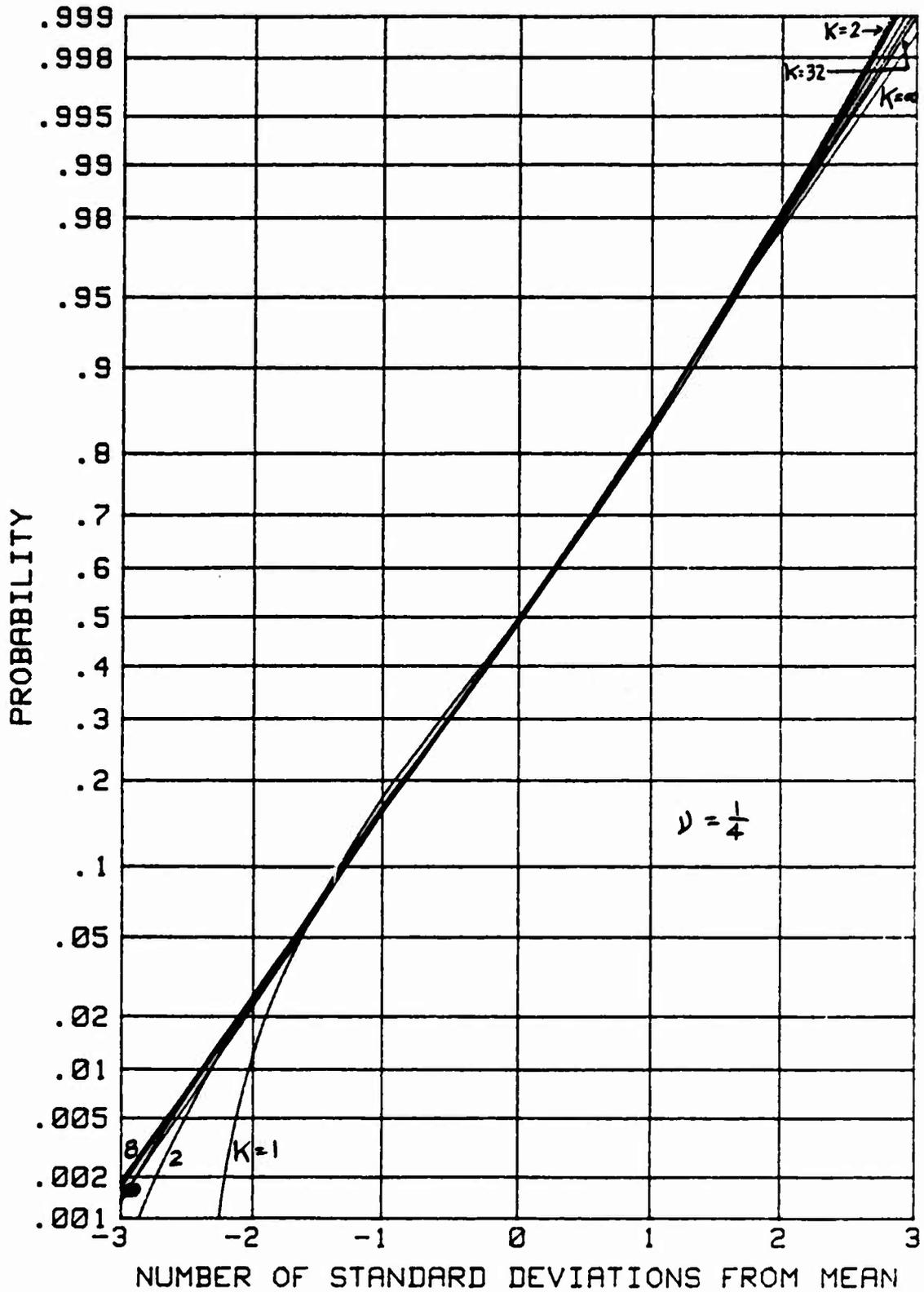Figure 3. Distribution of $v = z^\nu$ for $\nu = \frac{1}{3}$

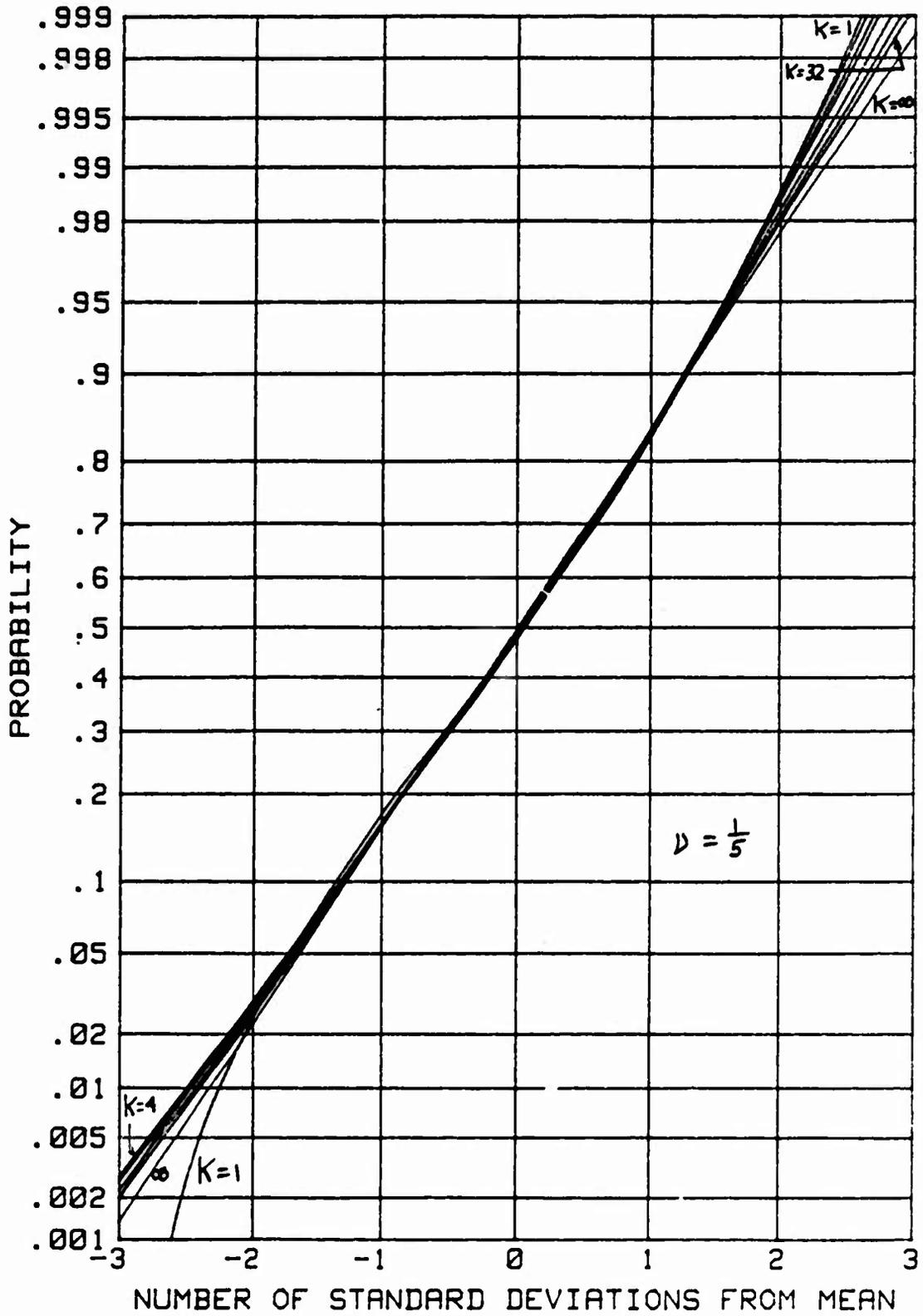Figure 4. Distribution of $V = z^\nu$ for $\nu = \frac{1}{4}$

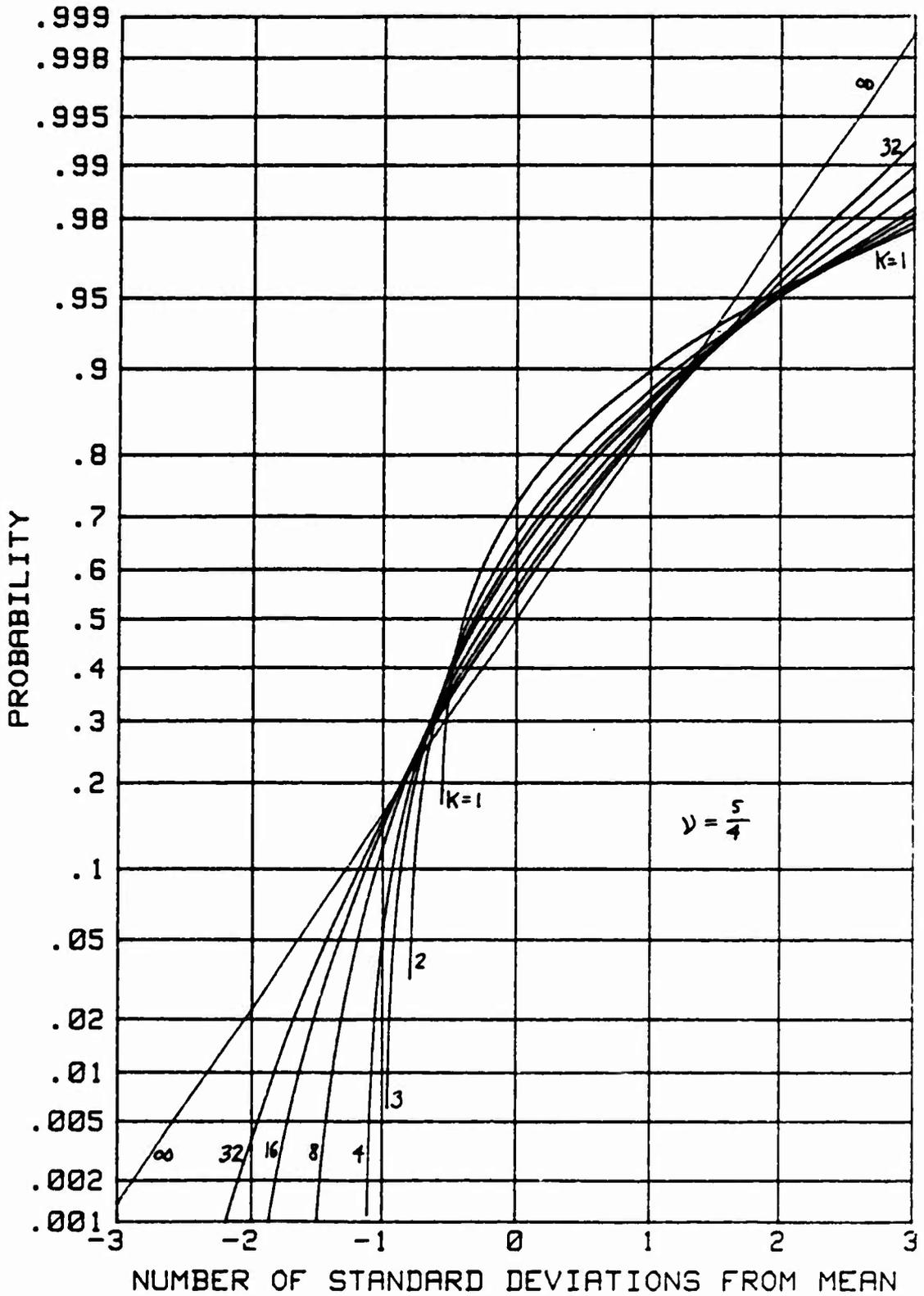Figure 5. Distribution of $v = z^\nu$ for $\nu = \frac{1}{5}$

Figure 6. Distribution of $v = z^\nu$ for $\nu = \frac{5}{4}$

The normalized n-th cumulant of v is

$$\frac{\lambda_v(n)}{\left[\lambda_v(2)\right]^{n/2}} \equiv r_n \quad . \tag{16}$$

This quantity is plotted in figures 7, 8, 9 for n = 3, 4, 5 respectively. The normalized third cumulant in figure 7 is near 0 for $v = 1/3$ when K is large. Thus this particular measure of non-Gaussianness is minimized by choosing $v \approx 1/3$.

The normalized fourth cumulant in figure 8 is smallest in the range (1/6, 1/2) for v, when K is large. It appears to be approaching zero for $v = 1/2$ and $v = 1/4$; however, it is still quite small for $v=1/3$.

The normalized fifth cumulant in figure 9 approaches zero (as K increases) at $v = 1/6, 1/3, 7/12$, and is nearly zero over an extended range of v.

Approximation to Cumulative Distribution Functions of v and z

For $v = 1/3$, since random variable v is nearly normal, we have as an approximation to its cumulative distribution function,

$$P_v(u) \cong \Phi\left(\frac{u-m_v}{\sigma_v}\right) \quad , \tag{17}$$

where Φ is the Gaussian cumulative distribution function defined in (8). And from (12)-(13), we have

$$m_v = \mu_1 = 2^{1/3}\frac{\Gamma\left(\frac{K}{2} + \frac{1}{3}\right)}{\Gamma\left(\frac{K}{2}\right)}$$

$$\mu_2 = 2^{2/3}\frac{\Gamma\left(\frac{K}{2} + \frac{2}{3}\right)}{\Gamma\left(\frac{K}{2}\right)}$$

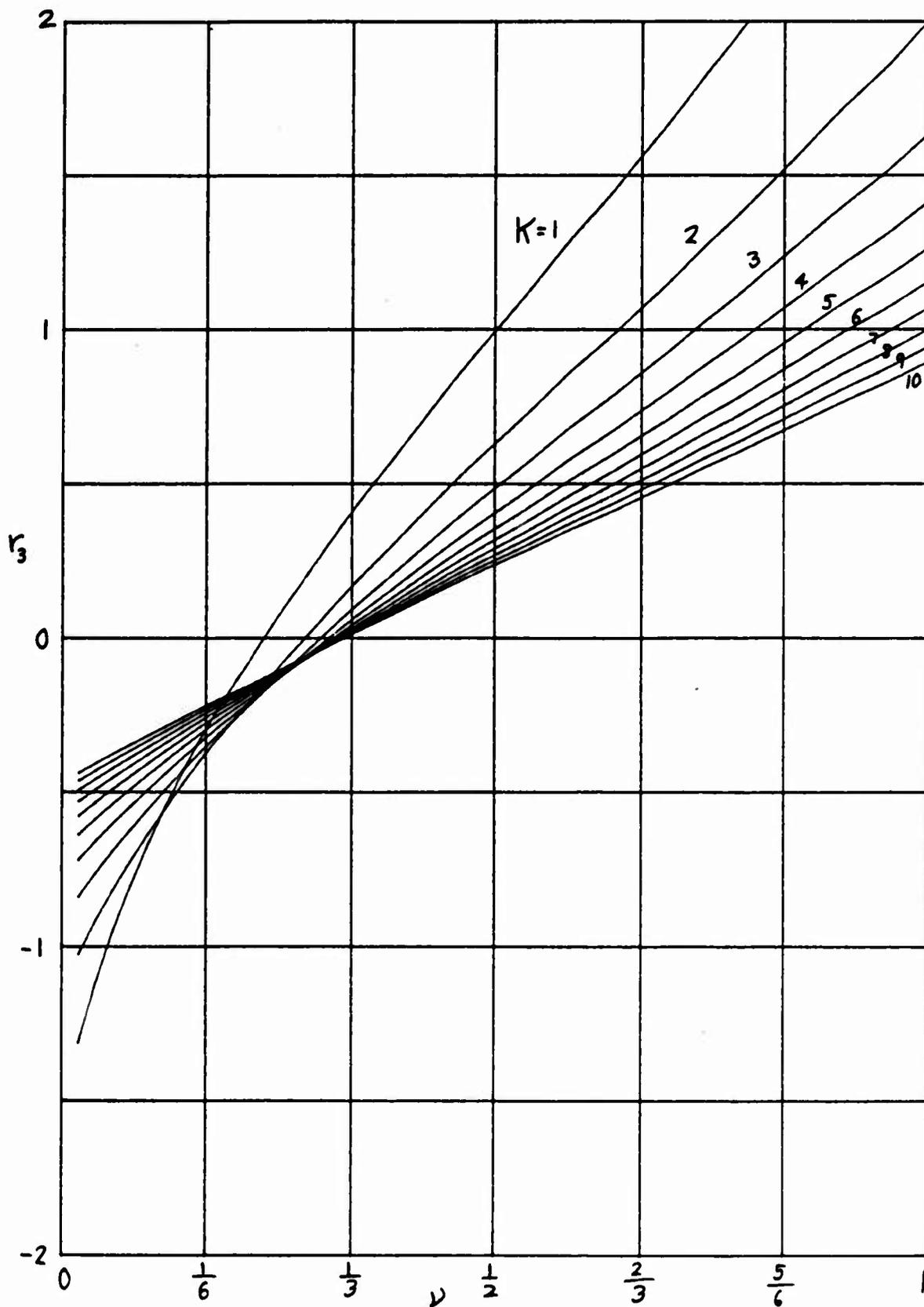$$\sigma_v = \left(\mu_2 - \mu_1^2\right)^{1/2} \quad . \tag{18}$$
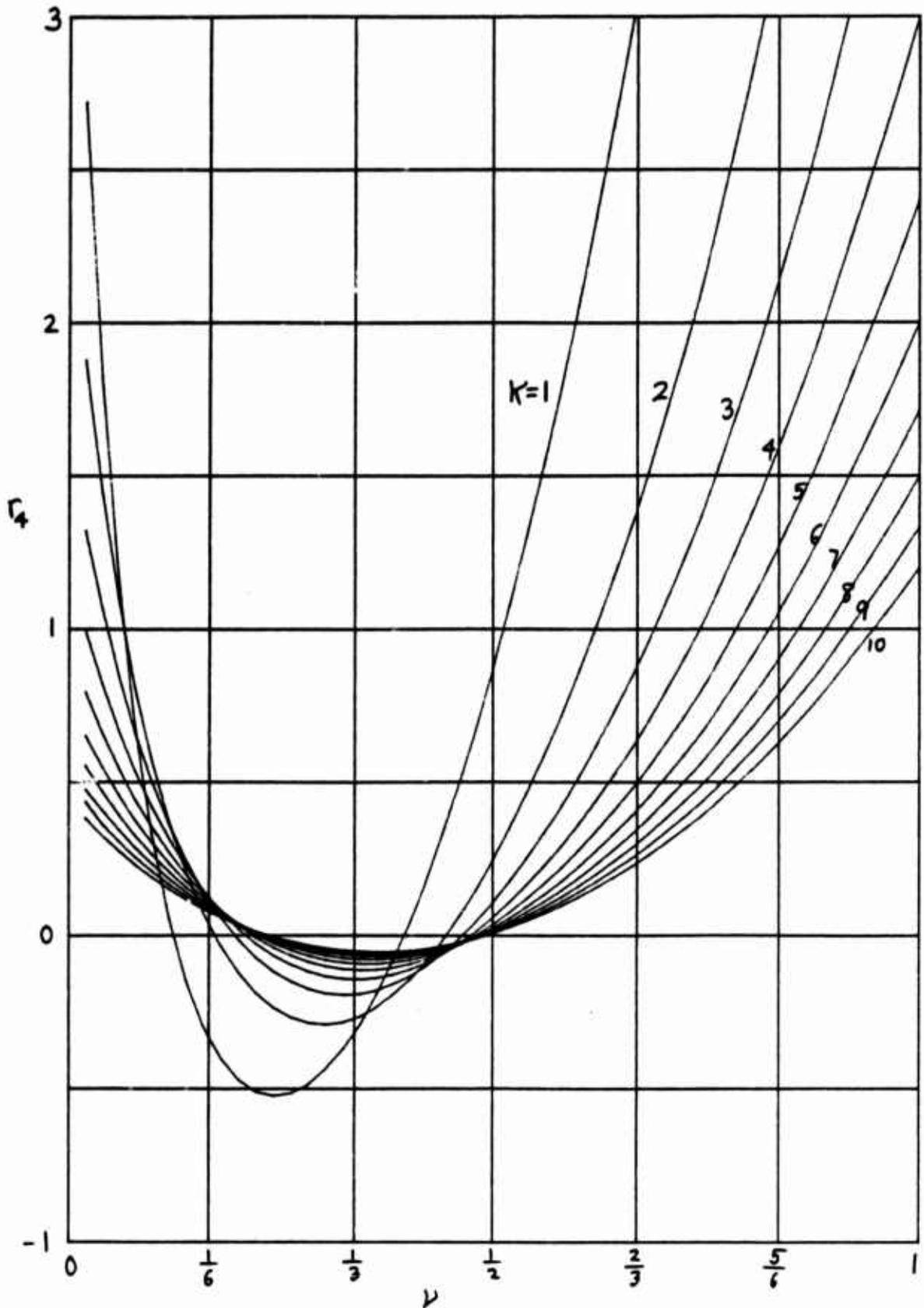
Figure 7. Normalized Third Cumulant of $v = z^\nu$
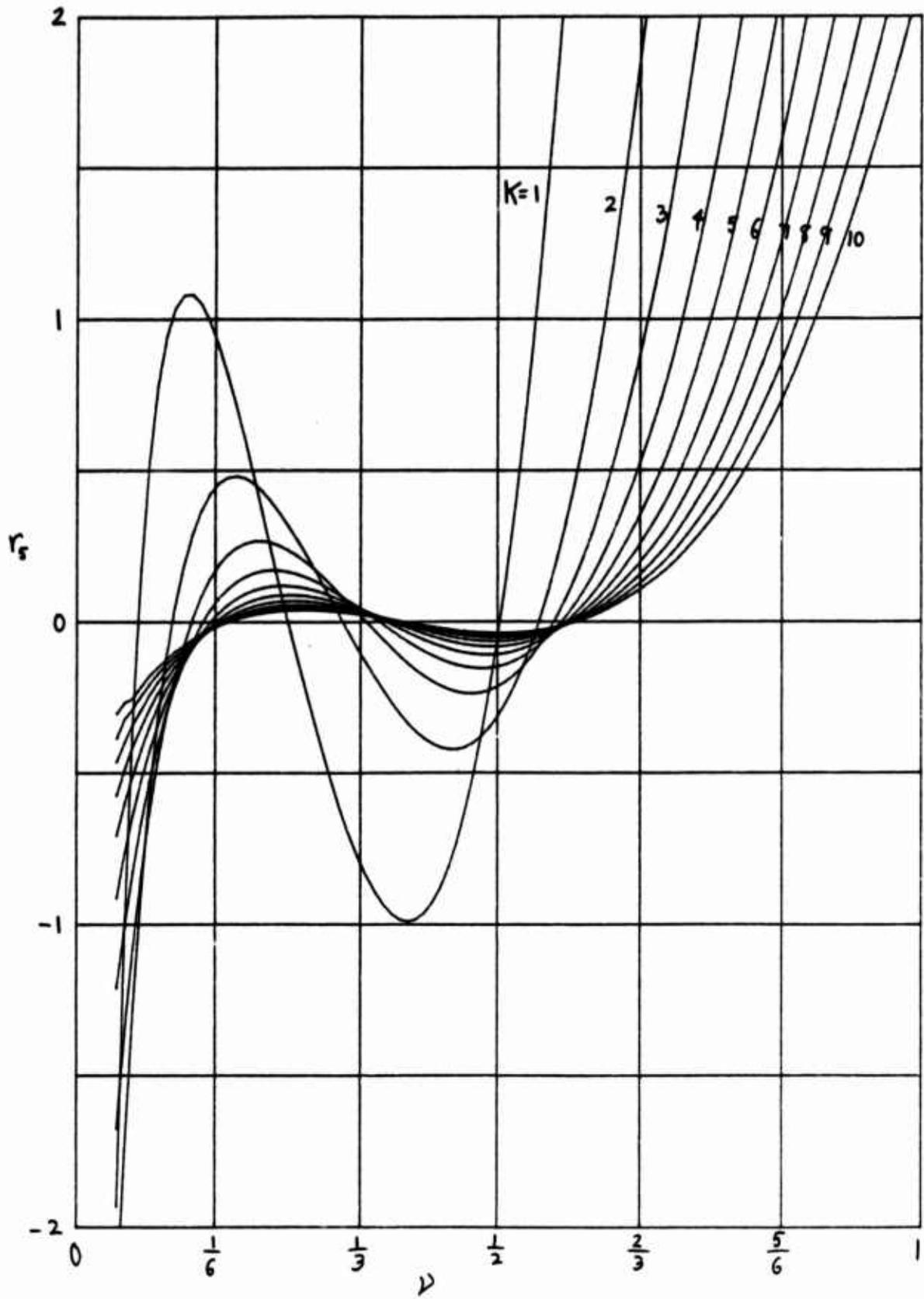
Figure 8. Normalized Fourth Cumulant of $v = z^\nu$

Figure 9. Normalized Fifth Cumulant of $v = z^\nu$

Then the cumulative distribution function of chi-squared variate z is, upon use of (10) (with $v = 1/3$) and (17),

$$P_z(u) = \text{Prob}\left\{z < u\right\} = \text{Prob}\left\{v^3 < u\right\} = \text{Prob}\left\{v < u^{1/3}\right\}$$

$$= P_v\left(u^{1/3}\right) \cong \Phi\left(\frac{u^{1/3} - m_v}{\sigma_v}\right) \quad . \tag{19}$$

This is similiar to the procedure in [6, p. 597, 19.3-2].

Asymptotic Behavior of Moments of v

The n-th moment of $v = z^v = (\chi^2)^v$ was given in (12). By use of [7, eq. 6.1.47], identifying $z \to K/2$, $a \to nv$, $b \to 0$, we have, as K tends to $+\infty$ in all the following asymptotic relations,

$$\mu_n = K^a\left[1 + \frac{a(a-1)}{K} + \frac{a(a-1)(a-2)(3a-1)}{6K^2} + O(K^{-3})\right]$$

or

$$\mu_n = (K-1+a)^a \left[1+O(K^{-2})\right]; \tag{20}$$

here a is to be interpreted as nv. See also [3, pp. 371-4].

Particular cases of (20) are first moment

$$\mu_1 = K^v\left[1 + \frac{v(v-1)}{K} + \frac{v(v-1)(v-2)(3v-1)}{6K^2} + O(K^{-3})\right]$$

or

$$\mu_1 = (K-1+v)^v \left[1+O(K^{-2})\right] \quad , \tag{21}$$

and second moment

$$\mu_2 = K^{2v}\left[1 + \frac{2v(2v-1)}{K} + \frac{2v(2v-1)(2v-2)(6v-1)}{6K^2} + O(K^{-3})\right]$$

or

$$\mu_2 = (K-1+2\nu)^{2\nu} [1+0(K^{-2})] \quad . \tag{22}$$

There follows for the variance of $v = z^\nu$,

$$\mu_2 - \mu_1^2 = K^{2\nu-1} \ 2\nu^2 \left[1 + \frac{(\nu-1)(3\nu-1)}{K} + 0(K^{-2})\right]$$

or

$$\mu_2 - \mu_1^2 = 2\nu^2 \ K^\nu (K-1+3\nu)^{\nu-1} [1+0(K^{-2})] \quad . \tag{23}$$

Particular cases of (21) and (23) are:

$$\left.\begin{array}{l} \mu_1 = \left(K - \dfrac{1}{2}\right)^{1/2} [1+0(K^{-2})] \\[2em] \mu_2 - \mu_1^2 = \dfrac{1}{2} \left(\dfrac{K}{K + \frac{1}{2}}\right)^{1/2} [1+0(K^{-2})] \end{array}\right\} \text{for } \nu = \frac{1}{2} \quad ; \tag{24}$$

these generalize [1, p. 79, par. 4], where the variable treated is $\sqrt{2\nu}$ rather than $\sqrt{\nu}$. Also

$$\left.\begin{array}{l} \mu_1 = \left(K - \dfrac{2}{3}\right)^{1/3} [1+0(K^{-2})] \\[2em] \mu_2 - \mu_1^2 = \dfrac{2}{9} \ K^{-1/3} [1+0(K^{-2})] \end{array}\right\} \text{for } \nu = \frac{1}{3} \quad . \tag{25}$$

The approximation

$$\mu_1 \cong \left(K - \frac{2}{3}\right)^{1/3} \qquad \text{for } \nu = 1/3 \tag{26}$$

is within 1% for $K \geq 3$, and the approximation

$$\mu_2 - \mu_1^2 \cong \frac{2}{9} K^{-1/3} \qquad \text{for } \nu = 1/3 \tag{27}$$

is within 1% for $K \geq 5$. The same errors hold for $\nu = 1/2$ also; i.e.

$$\left.\begin{array}{l} \mu_1 \cong \left(K - \frac{1}{2}\right)^{1/2} \\[1em] \mu_2 - \mu_1^2 \cong \frac{1}{2}\left(\frac{K}{K + \frac{1}{2}}\right)^{1/2} \end{array}\right\} \text{ for } \nu = 1/2 \quad . \tag{28}$$

STATISTICS OF THE SAMPLE GEOMETRIC MEAN OF A SET
OF MULTIPLES OF CHI-SQUARED VARIATES

This section will deal with the product of a number of independent chi-squared variates, instead of the power of a chi-squared variate. This operation (of extracting the sample geometric mean of a set of random variables) is encountered in maximum likelihood estimation of unknown parameters of a candidate cumulative distribution function of chi-squared form.

The probability density function of a multiple of a chi-squared variate, w, of K degrees of freedom, is from (3),

$$p_W(u) = \frac{u^{\frac{K}{2} - 1} \exp(-u/\beta)}{\beta^{K/2} \, \Gamma(K/2)} \quad \text{for} \quad u > 0 \quad . \tag{29}$$

The cumulative distribution function is, upon use of (4),

$$P_W(u) = P\left(\frac{2u}{\beta} \bigg| K\right) \quad . \tag{30}$$

Now let w_n, for $1 \leq n \leq N$, be a set of independent chi-squared variates, each with the probability density function in (29). The sample geometric mean of the set $\{w_n\}$ is

$$\hat{g} = (w_1 \, w_2 \, \cdots \, w_N)^{1/N} = \exp\left(\frac{\ln w_1 + \ldots + \ln w_N}{N}\right) \quad . \tag{31}$$

We are interested in the cumulative distribution function of $\hat{g}$. Define

$$\ell = \ln \hat{g} = \frac{1}{N} \sum_{n=1}^{N} y_n \quad , \quad (-\infty < \ell < +\infty) \tag{32}$$

where

$$y_n \equiv \ln w_n \quad \text{for} \quad 1 \leq n \leq N \quad . \tag{33}$$

The characteristic function of any one of the $\{y_n\}$ is, from (33) and (29),

$$f_y(\xi) = \overline{\exp(i\xi y)} = \overline{\exp(i\xi \ln w)} = \overline{w^{i\xi}}$$

$$= \int_0^\infty du \; \frac{u^{i\xi + \frac{K}{2} - 1} \exp(-u/\beta)}{\beta^{K/2} \, \Gamma(K/2)} = \beta^{i\xi} \frac{\Gamma\!\left(\frac{K}{2} + i\xi\right)}{\Gamma\!\left(\frac{K}{2}\right)} \quad . \qquad (34)$$

Therefore the characteristic function of ℓ in (32) is

$$f_\ell(\xi) = \overline{\exp(i\xi \ell)} = [f_y(\xi/N)]^N = \beta^{i\xi} \left[\frac{\Gamma\!\left(\frac{K}{2} + i\frac{\xi}{N}\right)}{\Gamma\!\left(\frac{K}{2}\right)} \right]^N \quad . \qquad (35)$$

Finally, the cumulative distribution function of the sample geometric mean $\hat{g}$ is, by use of (32),

$$P_{\hat{g}}(u) = \text{Prob}\left\{\hat{g} < u\right\} = \text{Prob}\left\{\ell < \ln u\right\}$$

$$= P_\ell(\ln u) \quad \text{for} \quad u > 0 \quad . \qquad (36)$$

The only numerical step required to find the cumulative distribution function of $\hat{g}$ is that from characteristic function f_ℓ in (35) to its cumulative distribution function P_ℓ required in (36); the techniques in [8, 9] are useful in this regard. For $N = 2$, a simpler approach for the cumulative distribution function of $\hat{g}$ is available in [10, p. 17].

SUMMARY

The near-Gaussian character of the one-third root of a chi-squared variate, with 4 degrees of freedom or more, leads to the simple approximation for the cumulative distribution function of a chi-squared variate as given by (19). The required parameters are given by (18) or approximations (25)-(27).

REFERENCES

1. R. A. Fisher, Statistical Methods for Research Workers, G. E. Stechert Co., N.Y., Eighth Edition, 1941.

2. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, N.J., 1961.

3. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 1, Hafner Publishing Co., N.Y., 1969.

4. E. B. Wilson and M. M. Hilferty, "The Distribution of Chi-Square", Proc. Nat. Acad. Sci., Vol. 17, pp. 684-688, 1931.

5. A. H. Nuttall, "On Characterization of the Probability Distribution of Measured Data," NUSC Technical Memorandum No. 811184, 10 December 1981.

6. G. A. Korn and T. M. Korn, Mathematical Handbook for Scientists and Engineers, McGraw-Hill Book Co., N.Y., 1961.

7. Handbook of Mathematical Functions, U.S. Dept. of Comm., Nat. Bur. of Std., Appl. Math. Series 55, U.S. Govt. Printing Office, June 1964.

8. A. H. Nuttall, "Numerical Evaluation of Cumulative Distribution Functions directly from Characteristic Functions," NUSL Report No. 1032, 11 August 1969; also Proc. IEEE, Vol. 57, No. 11, pp. 2071-2, November 1969.

9. A. H. Nuttall, "Alternative Forms and Computational Considerations for Numerical Evaluation of Cumulative Distribution Functions directly from Characteristic Functions," NUSC Report No. NL-3012, 12 August 1970; also Proc. IEEE, Vol. 58, No. 11, pp. 1872-3, November 1970.

10. A. H. Nuttall, "On Generation of Random Numbers with Specified Distributions or Densitites," NUSC Technical Report 6843, 1 December 1982.

Threshold Signal-to-Noise Ratio for Time Delay Estimation Via Cross-Correlation

A. H. Nuttall

ABSTRACT

An approximation for the threshold signal-to-noise ratio, at which the correlator performance estimate deviates from the Cramer-Rao lower bound, is derived as a function of the time-bandwidth product, TB, and the probability of an anomaly, P. The dependence on P is weak, and the dominant behavior of the threshold signal-to-noise ratio is according to $(TB)^{-1/2}$.

INTRODUCTION

Estimation of the time delay between two signals received at two sites in uncorrelated additive noises is often accomplished by filtering and cross-correlating the two received waveforms and locating the peak. Extensive results on the performance of this technique are given in [1-10]. It is often observed that the correlator performance deviates rather abruptly from the Cramér-Rao lower bound, in terms of the variance of the time delay estimate, when the input signal-to-noise ratio is decreased below a threshold value. Here we derive an approximation to this threshold signal-to-noise ratio, which is valid over a wide range of TB and P, where T is the observation time, B is the signal bandwidth, and P is the probability of an anomaly [4].

TECHNICAL CONSIDERATIONS

The starting point is the variance of the time delay estimate as given in [4]:

$$\sigma^2 = P\, T_0^2/3 + (1-P)\, \sigma_{CR}^2 = \sigma_{CR}^2 + P(T_0^2/3 - \sigma_{CR}^2), \tag{1}$$

where

$$P = 1 - \int dx\, \phi(x-\alpha)\, [\Phi(\beta x)]^{M-1} \quad . \tag{2}$$

Here P is the probability of an anomalous estimate, T_0 is the segment length* used in the correlation processing technique, and σ_{CR}^2 is the Cramér-Rao lower bound on the variance of the time delay estimate. The functions in (2) are

$$\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2) \quad ,$$

$$\Phi(x) = \int_{-\infty}^{x} dt\, \phi(t) \quad , \tag{3}$$

and we have parameters

$$M = 4T_0 B, \quad \alpha = \sqrt{2TB}\, \frac{R}{\sqrt{1+R^2}}, \quad \beta = \sqrt{1+R^2} \quad , \tag{4}$$

where

$$R = \frac{S}{S+N} = \frac{S/N}{1+S/N} \quad . \tag{5}$$

S and N are the input signal and noise powers at the correlator inputs. Thus S/N is the input signal-to-noise ratio on one element.

Now if the probability of an anomaly, P, is zero, the variance of the time delay estimate in (1) is simply σ_{CR}^2. However, as P increases slightly from zero, the $T_0^2/3$ term in (1) rapidly takes over and σ^2 increases

*More accurately, the search interval on the correlation estimate.

significantly. Since this effect is controlled by P, it is of interest to determine at what input signal-to-noise ratio, S/N, the probability P begins to deviate from zero. Mathematically, given a small P > 0, we must solve (2) for S/N, where the parameters are listed in (4).

From [10, figures 1-4], it is seen that for large TB, in order to realize a small value of P, the required input signal-to-noise ratio is much less than 1. In fact,

$$\frac{S}{N} \sim \frac{8}{\sqrt{2TB}} \quad \text{for} \quad TB \gg 1 \tag{6}$$

is a ballpark figure, which we want to improve on. Then

$$\alpha \sim 8, \ \beta \sim 1 \quad \text{for} \quad TB \gg 1 \quad . \tag{7}$$

Since $\phi(x-\alpha)$ peaks at $x=\alpha$, $\Phi(\beta x)$ in (2) takes its relevant argument values at $\beta x \sim \beta\alpha \sim 8$.

Now for $\alpha\beta$ large, we have

$$[\Phi(\beta x)]^{M-1} = [1 - \bar{\Phi}(-\beta x)]^{M-1} \simeq 1 - (M-1)\bar{\Phi}(-\beta x) \quad . \tag{8}$$

Then from (2),

$$P \simeq 1 - \int dx \ \phi(x-\alpha)[1-(M-1)\bar{\Phi}(-\beta x)] = (M-1)\int dx \ \phi(x-\alpha) \ \bar{\Phi}(-\beta x)$$

$$= (M-1) \int dt \ \phi(t) \ \bar{\Phi}(-\beta t-\alpha\beta) = (M-1)\bar{\Phi}\left(\frac{-\alpha\beta}{\sqrt{1+\beta^2}}\right) \quad . \tag{9}$$

Here we used the integral result

$$\int dx \ \exp\left(-\frac{1}{2}p^2 x^2 + qx\right) \Phi(ax+b) = \frac{\sqrt{2\pi}}{|p|} \exp\left(\frac{q^2}{2p^2}\right) \Phi\left(\frac{p^2 b + aq}{p\sqrt{p^2+a^2}}\right) \tag{10}$$

for $p \neq 0$.

The approximation in (9) is actually an upper bound; this follows from (8) and the observation that

$$(1-u)^{M-1} \geq 1-(M-1)u \quad \text{for} \quad 0 \leq u \leq 1 \quad , \tag{11}$$

which is simply a statement that the left-hand side of (11) is greater than the tangent at $u = 0$. Thus

$$P \leq (M-1) \, \tilde{\Phi}\left(\frac{-\alpha\beta}{\sqrt{1+\beta^2}}\right) \tag{12}$$

with no approximations.

By now employing the parameter definitions in (4), (9) becomes

$$P \approx (4T_0 B-1) \, \tilde{\Phi}\left(-\sqrt{TB}\,\frac{R}{\sqrt{1+R^2/2}}\right) \quad . \tag{13}$$

If we let $\tilde{\tilde{\Phi}}$ be the inverse function to $\tilde{\Phi}$, and define

$$y = -\frac{1}{\sqrt{TB}} \, \tilde{\tilde{\Phi}}\left(\frac{P}{4T_0 B-1}\right) \quad , \tag{14}$$

then (13) can be solved for R according to

$$R = \frac{y}{\sqrt{1-y^2/2}} \quad . \tag{15}$$

Then (5) yields the input signal-to-noise ratio required as

$$\frac{S}{N} = \frac{R}{1-R} = \frac{y}{\sqrt{1-y^2/2} - y} \quad , \tag{16}$$

in terms of the quantity defined in (14). For large TB, y is small, and (16) and (14) yield approximation

$$\frac{S}{N} \approx -\frac{1}{\sqrt{TB}} \, \tilde{\tilde{\Phi}}\left(\frac{P}{4T_0 B-1}\right) \quad . \tag{17}$$

This result is reminiscent of [9, eqs. 2 and 3], where we must note that SNR there is proportional to the square of S/N here; see [9, eq. 1]. Also those results are independent of the probability of anomaly, by virtue of the choice of threshold variance as being twice the Cramér-Rao lower bound.

The inverse function $\tilde{\Phi}$ in (14) and (17) is a relatively weak function of its argument; thus the dominant behavior of the required threshold signal-to-noise ratio is according to $(TB)^{-1/2}$, i.e., -1.5 dB per doubling of the TB product.

A short table of the required S/N as calculated from (14) and (16) is given in table 1 for $B = 100$ Hz, $T_o = 1/8$ sec, and $P = 10^{-6}$; similar results for $P = 10^{-5}$ are listed in table 2. Comparison of these tabular results with [10, figure 2] reveals that probability of anomaly $P = 10^{-6}$ marks the breakpoint between the Cramér-Rao lower bound and the correlator performance estimate.

T (sec)	S/N (dB)
2	-1.70
8	-6.13
32	-9.68
128	-12.92
512	-16.05

Table 1. Required Input Signal-to-Noise Ratio for $P = 10^{-6}$

T (sec)	S/N (dB)
2	-2.31
8	-6.57
32	-10.06
128	-13.29
512	-16.40

Table 2. Required Input Signal-to-Noise Ratio for $P = 10^{-5}$

9

APPENDIX. EVALUATION OF P IN (2)

We develop (2) as follows:

$$P = 1 - \int dx\ \phi(x-\alpha)\ [\Phi(\beta x)]^{M-1} \tag{A-1}$$

$$= \int dx\ \phi(x-\alpha)\left\{1-[\Phi(\beta x)]^{M-1}\right\} = \int_{-\infty}^{0} + \int_{0}^{+\infty}$$

$$= \int_{-\infty}^{0} dx\ \phi(x-\alpha)\left\{1-[\Phi(\beta x)]^{M-1}\right\} + \int_{0}^{+\infty} dx\ \phi(x-\alpha)\left\{1-[1-\Phi(-\beta x)]^{M-1}\right\}\ . \tag{A-2}$$

Form (A-1) is difficult to use, because it requires a final differencing from 1. For small P (the case of interest), the integral in (A-1) requires many digits of significance, in order to yield accurate P.

In (A-2), observe that since $x \leq 0$ in the first integral, the Φ function values there will be $\leq 1/2$. Also, since $x \geq 0$ in the second integral, those Φ values will also be $\leq 1/2$. Now consider the two functions

$$\left.\begin{array}{l} g\{u\} \equiv 1 - u^{M-1} \\[2em] h\{u\} \equiv 1 - (1-u)^{M-1} \end{array}\right\} \text{for } 0 \leq u \leq 1\ . \tag{A-3}$$

Then

$$P = \int_{-\infty}^{0} dx\ \phi(x-\alpha)\ g\{\Phi(\beta x)\} + \int_{0}^{+\infty} dx\ \phi(x-\alpha)\ h\{\Phi(-\beta x)\}\ . \tag{A-4}$$

The arguments of $g\{\ \}$ and $h\{\ \}$ are always $\leq 1/2$. For evaluation of g, we simply use the definition (A-3) directly, and get good accuracy. For evaluation of h, we must use a power series expansion about $u = 0$ when u is small. We have then

$$h\{u\} = 1-(1-u)^{M-1} = \sum_{k=1}^{M-1} \binom{M-1}{k}(-1)^{k-1}u^k\ . \tag{A-5}$$

For larger u $\left(\text{say } u > \frac{1}{M-1}\right)$, we use the definition (A-3) directly. So h can be accurately evaluated very easily. Now P is a sum of positive quantities, where all the quantities can be accurately evaluated.

The direct definition of h yields the alternative

$$h\{u\} = 1-(1-u)^{M-1} = 1- \exp[(M-1)\ln(1-u)]$$

$$= 1- \exp\left[-(M-1) \sum_{n=1}^{\infty} u^n/n\right] \quad . \tag{A-6}$$

If the $\ln$ is not sufficiently accurate for small u (like 1/(M-1)), it may be necessary to resort to the power series expansion illustrated. For $u > \frac{1}{M-1}$, accuracy is still retained in the difference from 1.

A program for the evaluation of P and then σ^2 via (1) is presented below.

```
1 ! TECH. MEMO. 831065, APRIL 83
10      Db=-16           ! 10 LGT(S/N)
20      Bc=100           ! BANDWIDTH B
30      Tc=512           ! OBSERVATION TIME T
40      To=1/8           ! SEGMENT LENGTH To
50      Mc=4*Bc*To       ! M
60      Sn=10^(.1*Db)    ! S/N
70      Bt=Bc*Tc         ! BT
80      S2cr=.375*(1+2*Sn)/(PI*PI*Sn*Sn*Bt*Bc*Bc)
90      OUTPUT 0;"dB = ";Db," B =";Bc," T =";Tc,"To =";To
100     OUTPUT 0;"LGT(SigmaCR) = ";LGT(SQR(S2cr))," M =";Mc
110     COM A1,Be,M1
120     Ra=Sn/(1+Sn)
130     Be=SQR(1+Ra*Ra)
140     A1=SQR(2*Bt)*Ra/Be
150     M1=Mc-1
160     L1=A1+FNInvphi(1E-12)
170     L1=MIN(-5,L1)
180     L2=FNPhi(-A1*Be)
190     L2=FNH(L2,M1)
200     IF L2>1E-12 THEN 230
210     L2=5
220     GOTO 240
230     L2=A1-FNInvphi(1E-12/L2)
240     A=L1
250     B=0
260     S=(FNS1(A)+FNS1(B))*.5
270     N=2
280     H=(B-A)*.5
290     F=(B-A)/3
300     V1=9E99
310     T=0
320     FOR K=1 TO N-1 STEP 2
330     T=T+FNS1(A+H*K)
340     NEXT K
350     S=S+T
360     Vo=V1
370     V1=(S+T)*F
380     PRINT USING "M.12DE,7D";V1,N
390     IF ABS(V1-Vo)<=1E-8 THEN 440
400     N=N*2
410     H=H*.5
420     F=F*.5
430     GOTO 310
440     PRINT
450     A=0
460     B=L2
470     S=(FNS2(A)+FNS2(B))*.5
480     N=2
490     H=(B-A)*.5
500     F=(B-A)/3
```

```
510     V2=9E99
520     T=0
530     FOR K=1 TO 4-1 STEP 2
540     T=T+FNS2(A+H*K)
550     NEXT K
560     S=S+T
570     Vo=V2
580     V2=(S+T)*F
590     PRINT USING "M.12DE,7D";V2,N
600     IF ABS(V2-Vo)<=1E-8 THEN 650
610     N=N*2
620     H=H*.5
630     F=F*.5
640     GOTO 520
650     P=V1+V2
660     S2cpe=P*To*To/3+(1-P)*S2cr
670     OUTPUT 0;"LGT(SigmaCPE) = ";LGT(SQR(S2cpe));" P =";P
680     END
690     !
700     DEF FNS1(X)
710     COM A1,Be,M1
720     T1=.398942280401*EXP(-.5*(X-A1)^2)
730     T2=FNPhi(Be*X)
740     T2=1-T2^M1
750     RETURN T1*T2
760     FNEND
770     !
780     DEF FNS2(X)
790     COM A1,Be,M1
800     T1=.398942280401*EXP(-.5*(X-A1)^2)
810     T2=FNPhi(-Be*X)
820     T2=FNH(T2,M1)
830     RETURN T1*T2
840     FNEND
850
860     DEF FNH(U,N)
870     IF U<1/N THEN 890
880     RETURN 1-(1-U)^N
890     N1=N+1
900     S=T=N*U
910     FOR K=2 TO N
920     T=-T*(N1-K)+U/K
930     S=S+T
940     IF ABS(T)<=1E-12*ABS(S) THEN 960
950     NEXT K
960     RETURN S
970     FNEND
980     !
```

```
990    DEF FNPhi(X)
1000   INTEGER J
1010   IF ABS(X)>5.14 THEN 1270
1020   A=.282842712475*X
1030   C=COS(A)
1040   S=SIN(A)
1050   B=C+C
1060   A=B*C-1
1070   C=A*(1.2536751E-18+B*7.10005E-20+A*7.4517E-21)
1080   C=A*(1.533423425E-16+B*1.01649277E-17+C)
1090   C=A*(1.36760444757E-14+B*1.0601364636E-15+C)
1100   C=A*(8.89786526722E-13+B*9.06068838945E-14+C)
1110   C=A*(4.22616144318E-11+B*4.46968229245E-12+C)
1120   C=A*(1.46660614234E-9+B*1.80848587810E-10+C)
1130   C=A*(3.72252349369E-8+B*5.34275027603E-9+C)
1140   C=A*(6.91927520325E-7+B*1.15330990944E-7+C)
1150   C=A*(9.43281169838E-6+B*1.82066316364E-6+C)
1160   C=A*(9.44909268810E-5+B*2.10404583073E-5+C)
1170   C=A*(6.97183792408E-4+B*1.78223016255E-4+C)
1180   C=A*(3.80150767985E-3+B*1.10860645342E-3+C)
1190   C=A*(.0.?3985726157+B*.005079069612+C)
1200   C=A*(.0467755234325+B*.0172439625887+C)
1210   C=A*(.108630245023+B*.0439773381941+C)
1220   C=A*(.201339747265+B*.0869894549959+C)
1230   C=A*(.330501521917+B*.144227226362+C)
1240   C=.70322500?744+B*.247255163140+C
1250   Phi=.5+.0450158158079*X+.5*S*C
1260   RETURN Phi
1270   IF X>7 THEN RETURN 1
1280   N=MAX(6,INT(69/ABS(X)),INT(525/(X*X)))+1
1290   A=1
1300   S=1
1310   B=1/X
1320   C=B*B
1330   FOR J=1 TO N
1340   A=(1-J-J)*A*C
1350   S=S+A
1360   NEXT J
1370   Phi=.398942280401*EXP(-.5*X*X)*ABS(B)*S
1380   IF X>0 THEN Phi=1-Phi
1390   RETURN Phi
1400   FNEND
1410   !
```

A-5

```
1420   DEF FNInvphi(Z)          !   invphi(z) via Strecok
1430   X=2*Z-1
1440   DIM T(0:20),A(0:20)
1450   INTEGER N
1460   DATA .992885376619,.120467516143,.0160781993421,.00263670443716
1470   DATA .49963473024E-3,.988982186E-4,.203910127E-4,.432727162E-5
1480   DATA .93803141E-6,.20673472E-6,.461597E-7,.1041666E-7,.23715E-8
1490   DATA .54393E-9,.12555E-9,.2914E-10,.679E-11,.159E-11,.37E-12
1500   DATA .912158803418,-.0162662818677,.433555472095E-3,.21443857007E-3
1510   DATA .262575108E-5,-.302109105E-5,-.1040606E-7,.6240661E-7,-.54012E-9
1520   DATA -.142321E-8,.3438E-10,.3358E-10,-.146E-11,-.81E-12,.5E-13,.2E-13
1530   DATA .956679709020,-.0231070043091,-.00437423689751,-.57650342265E-3
1540   DATA -.1096102231E-4,.2510854702E-4,.105612336E-4,.275441233E-5
1550   DATA .43248450E-6,-.2053034E-7,-.4389154E-7,-.17FS401E-7,-.396129E-8
1560   DATA -.18693E-9,.27292E-9,.13282E-9,.3183E-10,.167E-11,-.204E-11
1570   DATA -.965E-12,-.22E-12
1580   B=ABS(X)
1590   IF ABS(X)>=.8 THEN B=SQR(-LOG(4*Z*(1-Z)))
1600   IF ABS(X)<.8 THEN 1700
1610   IF ABS(X)<.9975 THEN 1660
1620   Nmax=20
1630   RESTORE 1530
1640   Y=-.559457631330*B+2.28791571626
1650   GOTO 1730
1660   Nmax=15
1670   RESTORE 1500
1680   Y=-1.54981304237*B+2.56549012315
1690   GOTO 1730
1700   Nmax=18
1710   RESTORE 1460
1720   Y=X*X*3.125-1
1730   REDIM A(0:Nmax)
1740   READ A(*)
1750   Y2=Y*2
1760   T(0)=1
1770   T(1)=Y
1780   FOR N=2 TO Nmax
1790   T(N)=Y2*T(N-1)-T(N-2)
1800   NEXT N
1810   R=0
1820   FOR N=Nmax TO 0 STEP -1
1830   R=R+A(N)*T(N)
1840   NEXT N
1850   Invphi=SGN(X)*B+R*1.41421356237
1860   RETURN Invphi
1870   FNEND
```

REFERENCES

1. C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Processing, vol. 24, pp. 320-327, August 1976.

2. W. R. Hahn, "Optimum signal processing for passive sonar range and bearing estimation," J. Acoust. Soc. Am., vol. 58, no. 1, pp. 201-207, July 1975.

3. R. L. Kirlin and J. N. Bradley, "Delay estimation simulations and a normalized comparison of published results," IEEE Trans. Acoust., Speech, Signal Processing, vol. 30, no. 3, pp. 580-611, June 1982.

4. J. P. Ianniello, "Time Delay Estimation via Cross-Correlation in the Presence of Large Estimation Errors", IEEE Trans. Acoust., Speech, Signal Proc., vol. ASSP-30, pp. 998-1003, 1982.

5. J. P. Ianniello, "Threshold Effects in Time Delay Estimation Using Narrowband Signals", ICASSP 82 Conference Proc., vol. 1, pp. 375-378.

6. S-K. Chow and P. M. Schultheiss, "Delay Estimation Using Narrowband Processors", IEEE Trans. Acoust., Speech, Signal Proc., vol. ASSP-29, pp. 478-484, 1981.

7. A. Weiss and E. Weinstein, "Fundamental Limitations in Passive Time Delay Estimation - Part I: Narrowband Systems", to appear in IEEE Trans. Acoust., Speech, Signal Proc., 1982.

8. E. Weinstein, "Performance Analysis of Time Delay Estimators", WHOI Technical Report, August 1982.

9. J. P. Ianniello, E. Weinstein, and A. Weiss, "Comparison of the Ziv-Zakai Lower Bound on Time Delay Estimation with Correlator Performance", IEEE ICASSP, Boston, MA, pp. 875-8, April 1983.

10. K. Scarbrough, R. Tremblay, and G. C. Carter," Performance Predictions for Coherent and Incoherent Processing Techniques of Time Delay Estimation", to appear in IEEE Trans. ASSP, October 1983.

Subject Matter Index