MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

REMARKS ON NONGAUSSIAN LINEAR PROCESSES
WITH ADDITIVE GAUSSIAN NOISE*

K. S. Lii
University of California, Riverside
Riverside, California 92502
and
M. Rosenblatt
University of California San Diego
La Jolla, California 92093

<u>The basic model</u>. In a number of papers [1,2,4,5,6] nonGaussian
linear processes are considered as the basic model. Questions
relating to the estimation of coefficients and deconvolution
were dealt with. We give the assumptions here. Let $\{v_t\}$ be
a sequence of independent, identically distributed random var-
iables with $E\, v_t \equiv 0$, $E\, v_t^2 \equiv 1$ and some higher order cumu-
lant $\gamma_s \neq 0$ $(s > 2)$. The real coefficients $\{\alpha_j\}$ are in
ℓ^2

$$(1) \qquad \qquad \sum \alpha_j^2 < \infty .$$

It's assumed that one observes the linear process

$$(2) \qquad \qquad X_t = \sum_j \alpha_j v_{t-j}$$

without knowledge of the α_j's or the v_t's. The object was to
determine procedures for the estimation of the coefficients
α_j *whether or not the system is minimum phase* and *effect de-*
convolution to estimate the v_t's.

We shall still look at aspects of the problem just described.
But our main concern will be with the modified problem in
which we observe *only*

$$(3) \qquad \qquad Y_t = X_t + \eta_t$$

86 2 18 137

where X_t has the structure given in (2) and $\{n_t\}$ is a Gaussian noise process independent of the process $\{X_t\}$.

In many cases the spectral range of the X and η processes will be disjoint. In such a case direct linear filtering will let us get the process $\{X_t\}$. This happens, for example, when the spectrum of the η's is in the high frequency range and that of the X process in the low frequency range. Suppose the interval [-a,a] contains the X spectral mass and that of the η process is outside. Further, let us assume that ±a are continuity points of the spectral distribution function F of Y. The indicator function of [-a,a] has the Fourier representation

$$\sum_j \frac{1}{\pi} \frac{\sin ja}{j} e^{-ij\lambda}$$

and this implies that

$$X_t = \sum_j \frac{1}{\pi} \frac{\sin ja}{j} Y_{t-j} .$$

Dwyer [3] has been concerned with techniques aimed at gauging the Gaussian or nonGaussian character of additive components of the process Y corresponding to different spectral ranges. He tries to assess the third and fourth order moment properties of such components.

We shall be interested in seeing what happens when the spectral ranges of the X and η processes overlap. If the η process is small compared to the X process, one can still try to deconvolve approximately even by proceeding naively as if η weren't there. An example of such a naive convolution is given below. All this is independent of whether the system is minimum phase.

In Figures 1 and 2 the process Y_t is generated by a Monte Carlo simulation where

$$X_t = v_t - 3.5 v_{t-1} + 1.5 v_{t-2}$$

and v_t is a sequence of exponential independent random variables of variance one. Here the roots of the polynomial $1 - 3.5z + 1.5z^2$ are 2 and 1/3. The additive noise n_t consists of independent Gaussian variables of mean zero and variance $\sigma^2 = 1$ in the case of Figure 1 and $\sigma^2 = 2$ for Figure 2. The first line of the figure graphs the sequence

IMOLD - 2   ROOT SET - 3   SIGMA - -1.0
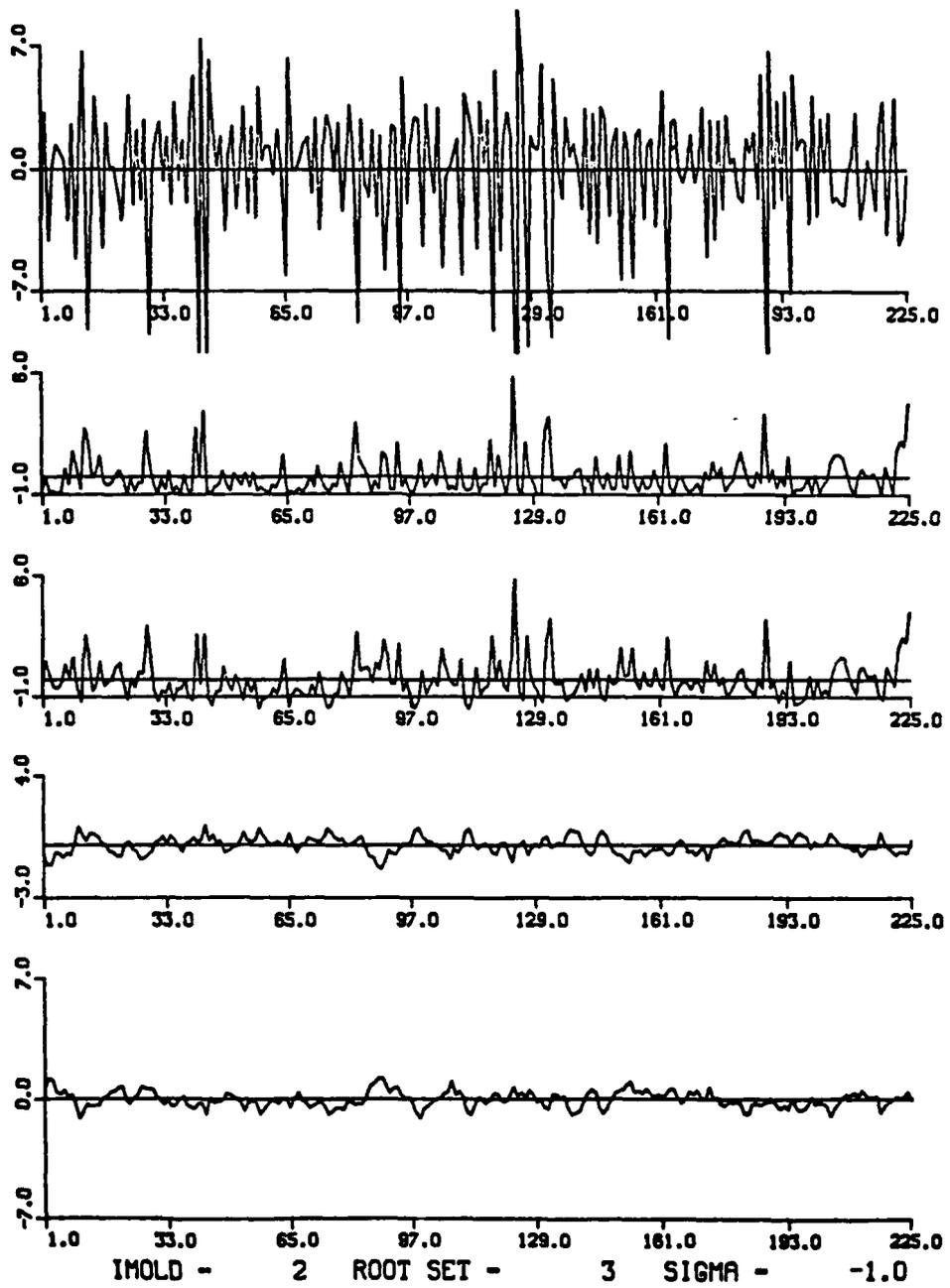
Figure 1.

By ....
Dist ibution/

Availability Codes

Dist	Avail and/or Special
A-1	

IMOLD — 3   ROOT SET — 3   SIGMA — -2.0
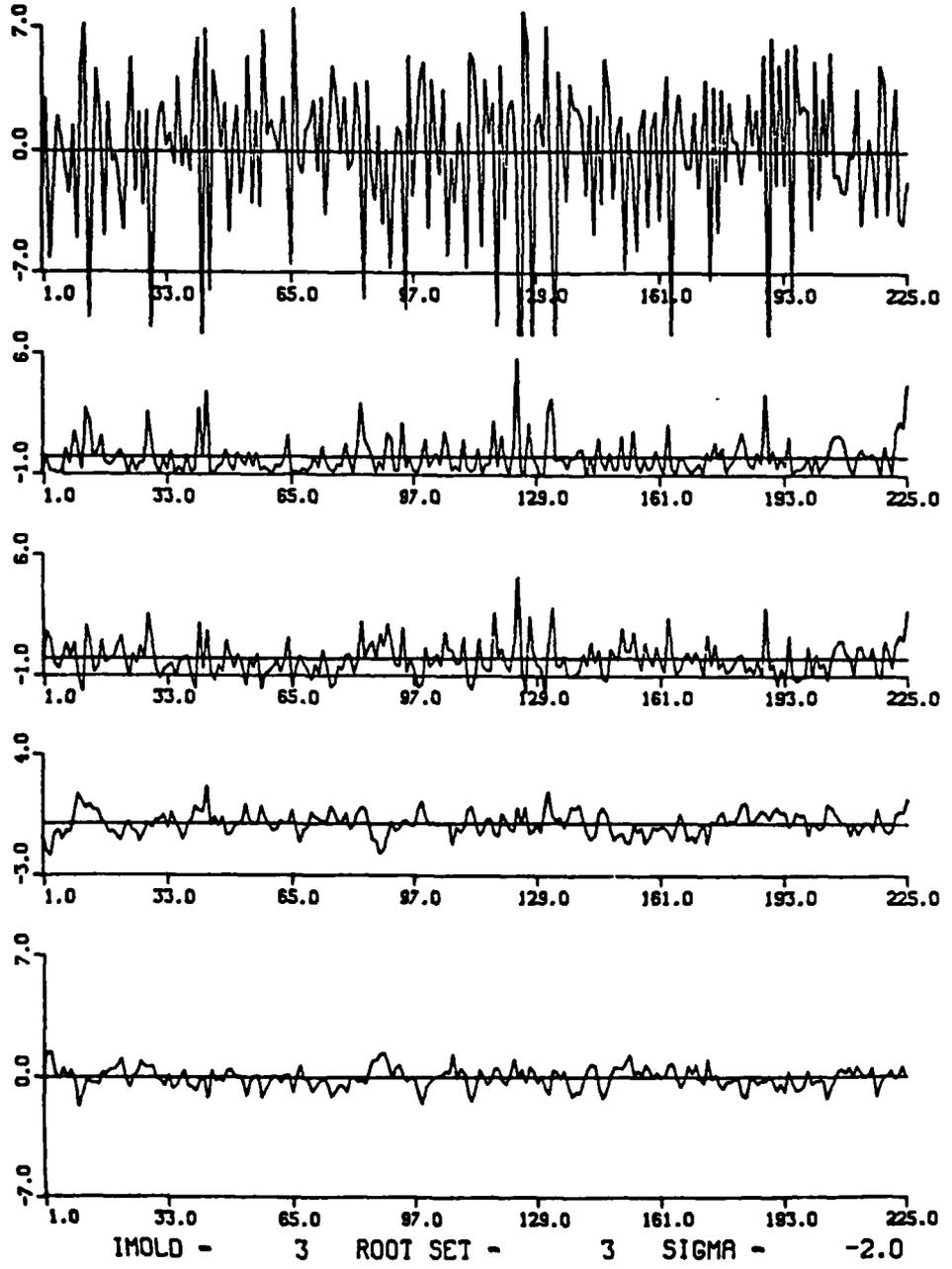
Figure 2.

Y_t. The second line graphs the generating v_t sequence. The
third line gives the result of our naive deconvolution ne-
glecting η. The fourth line gives the error in our decon-
volution. The last line has the result of a minimum phase
deconvolution which is naturally off because one of the roots
is inside the unit disc in the complex plane. Clearly our
naive deconvolution gets worse as σ^2 increases. But even
for $\sigma^2 = 2$, a moderate amount of noise, one can still rec-
ognize certain broad features of the v sequence. If η is
large one won't be able to deconvolve but one can still hope
to estimate a good deal about the α_j's under appropriate con-
ditions.

A simple model. To give some idea of the difficulties that
can arise, let us consider the case in which

(4) $$Y_t = a\, v_t + b\, u_t$$

with the $\{v_t\}$ sequence nonGaussian independent, identically
distributed with

$$E\, v_t \equiv o, \quad E\, v_t^2 \equiv 1, \quad E\, v_t^3 \equiv \gamma \neq 0$$

and $\{u_t\}$ an independent, identically distributed sequence
of $N(0,1)$ random variables. The $\{v_t\}$ and $\{u_t\}$ sequences
are assumed to be independent. It is assumed that $|a|$, $|b|$
and γ are unknown. The counterpart of the problem mentioned
above is that of estimating $|a|$ and $|b|$. $|a|$ and $|b|$
are not identifiable in terms of the problem as specified here
since v_t might be decomposable

(5) $$v_t = v_t^{(1)} + v_t^{(2)}$$

in terms of two nontrivial independent summands, one of which,
say $v_t^{(2)}$, is Gaussian. The problem can be normalized by
insisting that v_t be indecomposable in the sense that a repre-
sentation of the form (5) be impossible. We shall call a
random variable v_t with a nontrivial decomposition of the
form (5), in that a Gaussian summand with positive variance
exists, reducible. This could be expressed in terms of dis-
tribution functions. A distribution will be called reducible
if it has a nontrivial Gaussian component.

PROPOSITION. *A reducible distribution has a maximal decomposition relative to its Gaussian component.*

Let φ be the characteristic function of the reducible distribution. Then there are constants $c_n > 0$ and corresponding characteristic functions ψ_n such that

$$\varphi(t) = \psi_n(t) \exp\left(- c_n \frac{t^2}{2}\right) .$$

Then

$$\mathrm{Re}\ \varphi(t) \leq \mathrm{Re}\ \psi_n(t) \leq 1 .$$

Let the F_n be the distributions corresponding to the characteristic functions ψ_n. We have

$$\int_{|x| \geq 1/u} d\ F_n(x) \leq \frac{7}{u} \int_0^u \{1 - \mathrm{Re}\ \psi_n(v)\} dv$$

$$\leq \frac{7}{u} \int_0^u \{1 - \mathrm{Re}\ \varphi(v)\} dv .$$

We have then uniform bounds on the tails of the distributions F_n. One can therefore choose a subsequence F_{n_k} with $c_{n_k} \uparrow c = \sup_n c$ that converges weakly. Let the limiting distribution be F with corresponding characteristic function ψ. It then follows that

$$(6) \qquad\qquad \varphi(t) = \psi(t) \exp\left(- c \frac{t^2}{2}\right)$$

which corresponds to the unique maximal decomposition.

The problem of estimating $|a|$, $|b|$ now in (4) where v_t is irreducible is one in which $|a|$, $|b|$ are identifiable. However, the problem of deconvolution which is that of estimating v_t is clearly not meaningful.

The representation (6) implies that any distribution with a characteristic function $\varphi(t)$ that doesn't decrease to zero as fast as the Gaussian as $|t| \to \infty$ must be irreducible. This means that all discrete distributions and all gamma distributions must be irreducible.

Let's consider the question of predicting a v_t given Y_t. We shall put this in the form

$$Y = \zeta + \eta$$

where ζ and η are independent with means zero and variances σ_1^2

and σ_2^2 respectively. η is normal and ξ is nonnormal with density $g(\cdot)$. We are typically interested in the case in which σ_2^2 is small compared to σ_1^2. The best linear predictor of ξ in mean square is αY where (assuming σ_1^2 and σ_2^2 known)

$$\alpha = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

and the variance of the prediction error is

$$\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \ .$$

The best predictor of ξ in mean square is $E(\xi|Y)$. Now

$$E(\xi^k|Y) = \frac{\int \xi^k g(\xi) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{1}{2\sigma_2^2}(Y-\xi)^2]d\xi}{\int g(\xi) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{1}{2\sigma_2^2}(Y-\xi)^2]d\xi}$$

$$= \frac{\frac{1}{\sqrt{2\pi}} \int (Y-u\sigma_2)^k g(Y-u\sigma_2) \exp(-\frac{1}{2}u^2)du}{\frac{1}{\sqrt{2\pi}} \int g(Y-u\sigma_2) \exp(-\frac{1}{2}u^2)du} \ .$$

Let

$$m_k(Y) = \frac{\int v^k g(Y-u\sigma_2) \exp[-\frac{1}{2}u^2]du}{\int g(Y-u\sigma_2) \exp[-\frac{1}{2}u^2]du} \ .$$

Then

$$E(\xi^k|Y) = \sum_{j=0}^{k} \binom{k}{j} Y^{k-j}\sigma_2^j (-1)^j m_j(Y) \ .$$

Assuming sufficient smoothness and using a Taylor expansion we have

$$g(Y-u\sigma_2) = g(Y) - u\sigma_2 g'(Y) + \frac{(u\sigma_2)^2}{2!} g''(Y) + \dots$$

and consequently the best predictor is

$$E(\xi|Y) = Y + \sigma_2^2 \frac{g'(Y) + \frac{\sigma_2^2}{2} g^{(3)}(Y) + \dots}{g(Y) + \frac{\sigma_2^2}{2} g''(Y) + \dots}$$

while

$$E(\xi^2|Y) = Y^2 + 2Y \, \sigma_2^2 \; \frac{g'(Y) + \frac{\sigma_2^2}{2} g^{(3)}(Y) + \ldots}{g(Y) + \frac{\sigma_2^2}{2} g''(Y) + \ldots}$$

$$+ \, \sigma_2^2 \; \frac{g(Y) + \frac{3}{2} \sigma_2^2 \, g''(Y) + \ldots}{g(Y) + \frac{\sigma_2^2}{2} g''(Y) + \ldots} \; .$$

We should like to compare,

$$E((\xi-\alpha Y)^2|Y) = E(\xi^2|Y) - 2\alpha Y E(\xi|Y) + \alpha^2 Y^2$$

with

$$E((\xi-E(\xi|Y))^2|Y) = E(\xi^2|Y) - E(\xi|Y)^2 \; .$$

It is clear that

$$E((\xi-E(\xi|Y))^2|Y) \leq E((\xi-\alpha Y)^2|Y)$$

and the difference is given by

$$(\alpha Y-E(\xi|Y))^2 = ((1-\alpha)Y + \sigma_2^2 \, \frac{g'(Y)}{g(Y)} + o(\sigma_2))^2 \; .$$

Such a comparison can give us some idea of the effectiveness of the best predictor versus the linear predictor in the tail region of the g distribution. This tail region may be of greatest interest in certain deconvolution problems (see Wiggins [6]).

Up to this point we have considered prediction in mean square. Suppose we consider instead trying to minimize

$$E|\xi-f(Y)|^4$$

for some appropriate f when $Y = \xi+\eta$ with η $N(o,\sigma_2^2)$ and ξ irreducible nonnormal with mean zero and variance σ_1^2. An analysis can be carried out by considering

$$E(\xi-f(Y))^4 = E(E[(\xi-f(Y))^4|Y]) \; .$$

We can obviously minimize this if we can minimize the conditional fourth moment on the right side for each Y. This suggests that we consider minimizing

$$E(Z-c)^4$$

in c where Z is a random variable with $E\,Z^4 < \infty$ and c is a constant. Let m be the mean of Z with $\alpha = m - c$. Then

$$(7) \qquad E(Z-c)^4 = E(Z-m+\alpha)^4 = \alpha^4 + b\sigma^2\alpha^2 + 4\mu_3\alpha + \mu_4$$

with σ^2 the variance of z and μ_j the jth central moment. Differentiating this with respect to α we get

(8)
$$4\alpha^3 + 12\sigma^2\alpha + 4\mu_3 .$$

If $\mu_3 = 0$ the unique real zero is $\alpha = 0$ and the minimum is attained by $c = m$. If $\mu_3 \neq 0$ we have a unique real zero since the derivative

$$12\alpha^2 + 12\sigma^2$$

of (8) is positive. The zero will be negative if $\mu_3 > 0$ and positive if $\mu_3 < 0$. Set (8) equal to zero and solve for α. By Cardano's formula one obtains

(9)
$$\alpha = 2^{-1/3} [-\mu_3 + \{\mu_3^2 + 4\sigma^6\}^{1/2}]^{1/3}$$
$$- 2^{-1/3} [\mu_3 + \{\mu_3^2 + 4\sigma^6\}^{1/2}]^{1/3} .$$

This means that the function $f(Y)$ minimizing (7) is

$$f(Y) = m(Y) - \alpha(Y)$$

where

$$m(Y) = E(\xi | Y)$$

and $\alpha(Y)$ is given by expression (9) with

$$\mu_3 = E((\xi - E(\xi | Y))^3 | Y)$$

and

$$\sigma^2 = E((\xi - E(\xi | Y))^2 | Y) .$$

Estimation of coefficients. Let $\{Y_t\}$ be a process of type indicated in formula (3), that is, the sum of a linear non-Gaussian process and an independent Gaussian noise process. We shall indicate in a simple way that the coefficients α_j of the linear nonGaussian process can be estimated up to an undetermined multiplier and an undetermined time shift. This is an asymptotic argument.

PROPOSITION. Let $\{Y_t\}$ be a linear nonGaussian process perturbed by independent Gaussian noise. Suppose that $\sum |k||\alpha_k| < \infty$, $\alpha(i^{-i\lambda}) \neq 0$ for all λ and $\gamma_3 \neq 0$. Then the coefficients α_k can be consistently estimated up to an undetermined multiplier c and an unspecified time shift of the index set.

The bispectral density of the process $\{Y_t\}$ is

$$b_3(\lambda_1,\lambda_2) = \frac{\gamma_3}{(2\pi)^2}\,\alpha(e^{-i\lambda_1})\alpha(e^{-i\lambda_2})\alpha(e^{i(\lambda_1+\lambda_2)})\ .$$

As before one can estimate the phase of $\alpha(e^{-i\lambda})$ consistently up to an undetermined additive term $ik\pi$ with k integral. This can be accomplished by using bispectral estimates as in Lii and Rosenblatt [4]. Notice that

$$(10) \qquad b_3(\lambda,0) = \frac{\gamma_3}{(2\pi)^2}\,\alpha(1)\ |\alpha(e^{-i\lambda})|^2\ .$$

Bispectral estimates (using (10)) then allow us to estimate the spectral density of the process $\{X_t\}$ up to an undetermined constant. The proposition follows immediately from these remarks.

Corollary. *Let the assumptions of the proposition be satisfied. Suppose that the Gaussian noise is white and that $|\alpha(e^{-i\lambda})| \not\equiv$ constant for all λ. One can then estimate the multiplier c. There is then only an unspecified time shift of the index set. The variance of r can also be estimated.*

Since the assumptions of the proposition are satisfied the conclusions hold. We have only to estimate the multiplier c. Both the spectral density

$$f_1(\lambda) = \frac{1}{2\pi}\,|\alpha(e^{-i\lambda})|^2 + \frac{\sigma_n^2}{2\pi}$$

of the Y process and

$$b(\lambda,0) = \frac{\gamma_3}{(2\pi)^2}\,\alpha(1)\ |\alpha(e^{-i\lambda})|^2$$

can be estimated consistently. Clearly

$$f_1(\lambda) = \frac{2\pi}{\gamma_3\alpha(1)}\,b(\lambda,0) + \frac{\sigma_n^2}{2\pi}$$

$$= A\,b(\lambda,0) + B\ .$$

The coefficients $A,\ B$ can be consistently estimated given consistent estimates of $f_1(\lambda),\ b(\lambda,0)$. The conclusion then follows.

The argument of a polynomial transfer function. In many cases
it seems reasonable to assume that $\alpha(Z)$ is a polynomial

$$\alpha(Z) = \sum_{j=0}^{p} \alpha_j Z^j .$$

Because the coefficients are assumed real, the roots Z_j are
real or occur in complex conjugate pairs. The argument of
$\alpha(e^{-i\lambda})$ is the sum of the contributions from the factors cor-
responding to the roots. We shall consider the contribution
from a root Z_j with $|Z_j| \neq 1$. Suppose $Z_j = re^{-i\theta}$ with
$0 < r < 1$ and $-\pi < \theta \leq \pi$. Then

$$\log (e^{-i\lambda} - Z_j) = - i\lambda + \log (1 - e^{i\lambda} Z_j)$$

$$= - i\lambda - \sum_{k=1}^{\infty} (e^{i\lambda} Z_j)^k / k$$

and this implies that

(11) $\text{Im} \log (e^{-i\lambda} - Z_j) = - \lambda - \sum_{k=1}^{\infty} \frac{r^k \sin k(\lambda - \theta)}{k}$

$$\text{for } |Z_j| < 1 .$$

Similarly for $Z_j = re^{-i\theta}$ with $r > 1$ one has

$$\log (e^{-i\lambda} - Z_j) = \log (-Z_j) + \log (1 - Z_j^{-1} e^{-i\lambda})$$

$$= i\pi + \log Z_j - \sum_{k=1}^{\infty} (Z_j^{-1} e^{-i\lambda})^k / k$$

and this implies that we have a representation

(12) $\text{Im} \log (e^{-i\lambda} - Z_j) = \pi - \theta - \sum_{k=1}^{\infty} \frac{r^{-k} \sin k(\theta - \lambda)}{k}$

for $|Z_j| > 1$. Let us formally consider the difference between
(11) and (12) for $r = 1$, that is

(13) $\theta - \lambda - \pi + 2 \sum_{k=1}^{\infty} \frac{\sin k(\theta - \lambda)}{k} .$

Notice that the Fourier series of $x - \pi \text{ sgn } x$ is

$$- 2 \sum_{k=1}^{\infty} \frac{2 \sin kx}{k} .$$

This implies that the expression (13) equals zero for $0 < \theta
- \lambda < \pi$ and $- 2\pi$ for $- \pi < \theta - \lambda < 0$. This is consistent
with the indeterminacy of the argument up to a multiple of 2π.
For a real root $|Z_j| < 1$ expression (11) becomes

$$- \lambda \pm \sum_{k=1}^{\infty} \frac{r^k \sin k\lambda}{k}$$

according as to whether $\theta = 0$ or π. In the case of complex conjugate roots of absolute value less than one the sum of the contributions is

$$- 2\lambda - 2 \sum_{k=1}^{\infty} \frac{r^k \sin k\lambda \cos k\theta}{k} .$$

The corresponding remarks for roots greater than one read as follows. If $|z_j| > 1$ is real expression (12) is

$$(1 - \delta_{\theta,\pi})\pi \pm \sum_{k=1}^{\infty} \frac{r^{-k} \sin k\lambda}{k} .$$

with $\pm$ according as to whether $\theta = o$ or π. If there are complex conjugate roots one obtains

$$2\pi + 2 \sum_{k=1}^{\infty} \frac{r^{-k} \sin k\lambda \cos k\theta}{k} .$$

In our case we try to estimate the sum of the contributions to the argument from all zeros and then as a convention readjust the value at zero of the sum so that it is zero there.

This discussion of the relationship between the argument of the transfer function and zeros of $\alpha(Z)$ is natural when $\alpha(Z)$ is a polynomial. A corresponding discussion could be carried out if $\alpha(Z)$ is a rational function in terms of the zeros and poles of $\alpha(Z)$. However, one can easily give examples of analytic functions $\alpha(Z)$ with no zeros such as, for example,

$$\alpha(Z) = \exp \{\sin Z\} .$$

Here, there would be no meaning to such an analysis.

References

1. Beneviste, A., Goursat, M. and Ryget, G. (1980). "Robust identification of a nonminimum phase system," IEEE Trans. on Automatic Control, 25, no. 3, 385-399.

2. Donoho, D. (1981). "On minimum entropy deconvolution," in *Applied Time Series Analysis II* (ed. D. F. Findley) Academic Press, New York, 565-608.

3. Dwyer, R. F. (1982). "Arctic ambient noise statistical measurement results and their implications to sonar performance improvements," NUSC reprint report 6739.

4. Lii, K. S. and Rosenblatt, M. (1982). "Deconvolution and estimation of transfer function phase and coeffients for nonGaussian linear processes," Ann. Statist., 10, 1195-1208.

5. Rosenblatt, M. (1980). "Linear processes and bispectra," J. Appl. Probab. 17, 265-270.

6. Wiggins, R. A. (1978). "Minimum entropy deconvolution," Geoexploration 17.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ---	2. GOVT ACCESSION NO. ADA 164599	3. RECIPIENT'S CATALOG NUMBER ---
4. TITLE (and Subtitle) REMARKS ON NONGAUSSIAN LINEAR PROCESSES WITH ADDITIVE GAUSSIAN NOISE		5. TYPE OF REPORT & PERIOD COVERED Research
		6. PERFORMING ORG. REPORT NUMBER ---
7. AUTHOR(s) K. S. Lii and M. Rosenblatt		8. CONTRACT OR GRANT NUMBER(s) ONR Contract N00014-81-K-0003 and NSF Grant DMS83-12106
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics University of California, San Diego La Jolla, CA 92093		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS --
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE 1984
		13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) ---		15. SECURITY CLASS. (of this report) ---
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

nonGaussian, linear process, additive Gaussian noise.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Techniques have been proposed for the deconvolution of non-Gaussian linear processes. It is shown that if a moderate amount of Gaussian white noise is added, the same techniques are still rather effective in deconvolution of the initial nonGaussian process.

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

END

FILMED

3-86

DTIC