(12)

NUSC Technical Report 6639
9 March 1982

On Resonance Extraction and Waveform Fitting for Transient Data; Prony's Method

Albert H. Nuttall
Surface Ship Sonar Department

Naval Underwater Systems Center
Newport, Rhode Island / New London, Connecticut

Preface

This research was conducted under NUSC Project No. A75205, Subproject No. ZR0000101, *Applications of Statistical Communication Theory to Acoustic Signal Processing*"; Principal Investigator, Dr. A. H. Nuttall, Code 3302; Program Manager, CAPT D. F. Parrish, Naval Material Command, MAT 08L.

The Technical Reviewer for this report was Dr. A. H. Quazi, Code 3331.

Reviewed and Approved: 9 March 1982

D. Walters
Head, Surface Ship Sonar Department

The author of this report is located at the New London Laboratory of the Naval Underwater Systems Center, New London, Connecticut 06320.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR 6639	2. GOVT ACCESSION NO.: AD-A112 378
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): ON RESONANCE EXTRACTION AND WAVEFORM FITTING FOR TRANSIENT DATA; PRONY'S METHOD	5. TYPE OF REPORT & PERIOD COVERED:
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Albert H. Nuttall	8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Underwater Systems Center New London Laboratory New London, CT 06320	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: A75205 ZR0000101
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Material Command Code MAT 08L Washington, DC 20362	12. REPORT DATE: 9 March 1982 13. NUMBER OF PAGES: 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION / DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution 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)

Resonance Extraction	Complex Exponentials
Waveform Fitting	Least Squares
Transient Data	Linear Prediction
Prony's Method	Eigenvectors

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

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.

DD FORM 1 JAN 73 1473

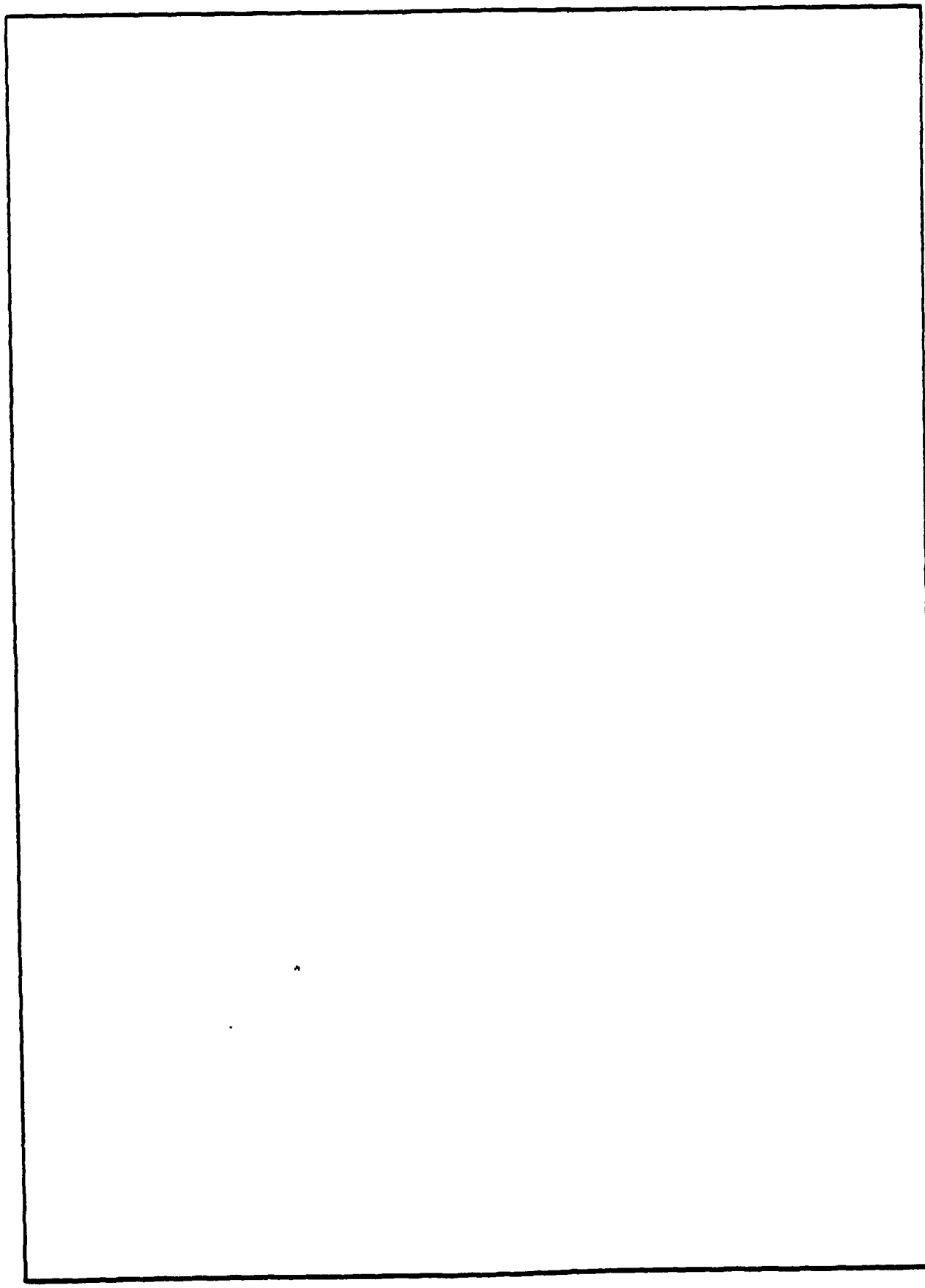
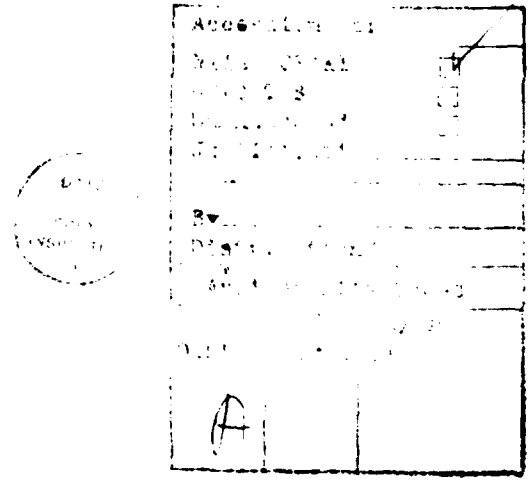


TABLE OF CONTENTS



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_o	smallest eigenvalue of S
e_o	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*

$$g_m = \sum_{k=1}^{n} C_k \exp(a_k m) \equiv \sum_{k=1}^{n} C_k \mu_k^m \quad \text{for} \quad 0 \leq m \leq N-1. \tag{1}$$

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 $\alpha_o \equiv -1$)

$$g_m - \sum_{j=1}^{n} \alpha_j g_{m-j} = - \sum_{j=0}^{n} \alpha_j g_{m-j} = - \sum_{j=0}^{n} \alpha_j \sum_{k=1}^{n} C_k \mu_k^{m-j}$$

$$= \sum_{k=1}^{n} C_k \mu_k^{m-n} \sum_{j=0}^{n} (-\alpha_j \mu_k^{n-j}) = \sum_{k=1}^{n} C_k \mu_k^{m-n} [\mu_k^n - \alpha_1 \mu_k^{n-1} - \ldots - \alpha_{n-1} \mu_k - \alpha_n],$$

(2)

where we substituted (1) and interchanged summations. Now we choose the n linear coefficients $\{\alpha_j\}_1^n$ such that

$$\mu_k^n - \alpha_1 \mu_k^{n-1} - \ldots - \alpha_{n-1} \mu_k - \alpha_n = 0 \quad \text{for} \quad 1 \leq k \leq n. \tag{3}$$

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

$$\alpha_j = (-1)^{j-1} \{\text{sum of all possible products of } j \text{ different } \mu\text{'s}\}$$

$$\text{for } 1 \leq j \leq n; \tag{4a}$$

*This can be generalized to include terms like $C\mu^m + Dm\mu^m$; see appendix A.

that is,

$$\begin{aligned}
\alpha_1 &= \mu_1 + \mu_2 + \ldots + \mu_n \\
\alpha_2 &= -(\mu_1\mu_2 + \mu_1\mu_3 + \ldots + \mu_1\mu_n + \mu_2\mu_3 + \mu_2\mu_4 + \ldots + \mu_{n-1}\mu_n) \\
&\vdots \\
\alpha_n &= (-1)^{n-1}\, \mu_1\mu_2\ldots\mu_n.
\end{aligned} \tag{4b}$$

With this choice of $\{\alpha_j\}_1^n$, (2) and (3) yield

$$g_m - \sum_{j=1}^{n} \alpha_j\, g_{m-j} = 0 \quad \text{for} \quad n \leq m \leq N-1, \tag{5a}$$

or

$$g_m = \sum_{j=1}^{n} \alpha_j\, g_{m-j} \quad \text{for} \quad n \leq m \leq N-1, \tag{5b}$$

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 $n \leq m \leq N-1$. 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

$$\hat{f}_m \equiv \sum_{j=1}^{n} \alpha_j f_{m-j} \quad \text{for} \quad n \leq m \leq N-1, \tag{6}$$

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])*

$$\hat{e}_m \equiv f_m - \hat{f}_m = f_m - \sum_{j=1}^{n} \alpha_j f_{m-j} \quad \text{for} \quad n \leq m \leq N-1. \tag{7}$$

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

$$\frac{\partial \hat{E}}{\partial \alpha_k} = 0 \quad \text{for} \quad 1 \leq k \leq n. \tag{9}$$

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*

$$\max_{n \leq m \leq N-1} \left| f_m - \sum_{j=1}^{n} \alpha_j f_{m-j} \right|. \tag{10}$$

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*

$$\tilde{f}_m \equiv \sum_{k=1}^{n} C_k \mu_k^m \quad \text{for} \quad 0 \leq m \leq N-1. \tag{11}$$

Then we define data error sequence (called the true error in [1])

$$\tilde{e}_m \equiv f_m - \tilde{f}_m = f_m - \sum_{k=1}^{n} C_k \mu_k^m \quad \text{for} \quad 0 \leq m \leq N-1. \tag{12}$$

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

$$\frac{\partial \tilde{E}}{\partial C_j} = 0 \quad \text{for} \quad 1 \leq j \leq n, \tag{14}$$

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

$$\max_{0 \leq m \leq N-1} \left| f_m - \sum_{k=1}^{n} C_k \mu_k^m \right| . \tag{15}$$

*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 $C_1\mu_1^m + C_2\mu_2^m$.

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 sequential 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 nonlinear 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 simultaneous 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

$$\begin{aligned} & g_m - \sum_{j=1}^{n} \alpha_j g_{m-j} - \sum_{j=1}^{p} \beta_j g_{m-n-j} \\ = & - \sum_{j=0}^{n} \alpha_j \left[\sum_{k=1}^{n} C_k \mu_k^{m-j} + \sum_{k=1}^{p} D_k (m-j)\mu_k^{m-j} \right] \qquad (\alpha_o \equiv -1) \\ & - \sum_{j=1}^{p} \beta_j \left[\sum_{k=1}^{n} C_k \mu_k^{m-n-j} + \sum_{k=1}^{p} D_k (m-n-j)\mu_k^{m-n-j} \right] \\ = & - \sum_{k=1}^{n} C_k \mu_k^m \left[\sum_{j=0}^{n} \alpha_j \mu_k^{-j} + \sum_{j=1}^{p} \beta_j \mu_k^{-n-j} \right] \\ & - \sum_{k=1}^{p} D_k \mu_k^m \left[\sum_{j=0}^{n} \alpha_j (m-j)\mu_k^{-j} + \sum_{j=1}^{p} \beta_j (m-n-j)\mu_k^{-n-j} \right]. \end{aligned} \tag{A.2}$$

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 $\{\beta_j\}_1^p$; $\alpha_o = -1$. These equations can be put in the form

$$\alpha_o \mu_k^{n+p} + \alpha_1\mu_k^{n+p-1} +\ldots+ \alpha_n\mu_k^p + \beta_1\mu_k^{p-1} +\ldots+ \beta_p = 0 \quad \text{for} \quad 1 \leq k \leq n, \tag{A.5}$$

$$\alpha_1\mu_k^{n+p-1} +\ldots+ \alpha_n n\mu_k^p + \beta_1(n+1)\mu_k^{p-1} +\ldots+ \beta_p(n+p) = 0 \text{ for } \quad 1 \leq k \leq p. \tag{A.6}$$

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

$$e_m \equiv \sum_{j=0}^{n} \alpha_j \, g_{m-j} \quad \text{for} \quad n \leq m \leq N-1, \tag{B.1}$$

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

$$e_m = \sum_{j=0}^{n} \alpha_j \sum_{k=1}^{n} C_k \, \mu_k^{m-j} = \sum_{k=1}^{n} C_k \, \mu_k^{m} \sum_{j=0}^{n} \alpha_j \, \mu_k^{-j}$$

$$= \sum_{k=1}^{n} C_k \, \mu_k^{m-n} \sum_{j=0}^{n} \alpha_j \, \mu_k^{n-j} \quad \text{for} \quad n \leq m \leq N-1. \tag{B.2}$$

Now let us set

$$\sum_{j=0}^{n} \alpha_j \, \mu_k^{n-j} = \alpha_o \, \mu_k^n + \ldots + \alpha_{n-1} \, \mu_k + \alpha_n = 0 \quad \text{for} \quad 1 \leq k \leq n, \tag{B.3}$$

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

$$g_{m-n} = \alpha_o g_m + \ldots + \alpha_{n-1} \, g_{m-n+1} \quad \text{for} \quad n \leq m \leq N-1. \tag{B.5}$$

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

$$C = \begin{bmatrix} c_0 \\ c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix}, \quad A = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{bmatrix} \tag{B.7}$$

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

$$C^T C = 1 \qquad \text{(or K if desired)}, \tag{B.8}$$

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))

$$d_m \equiv \sum_{j=0}^{n} \alpha_j \, f_{m-j} \quad \text{for} \quad n \leq m \leq N-1, \tag{B.10}$$

where set $\{\alpha_j\}_0^n$ is not yet specified. Define error and data matrices

$$D = \begin{bmatrix} d_n \\ d_{n+1} \\ \cdot \\ \cdot \\ d_{N-1} \end{bmatrix}, \quad F = \begin{bmatrix} f_n & f_{n-1} & \cdots & f_0 \\ f_{n+1} & \cdots & & f_1 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ f_{N-1} & \cdots & & f_{N-1-n} \end{bmatrix} \quad (N-n)\times(n+1). \tag{B.11}$$

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

$$S = F^T F \qquad (n+1)x(n+1) \text{ matrix.} \tag{B.15}$$

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

$$A_o = \lambda S^{-1} C. \tag{B.17}$$

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

$$\Lambda = \begin{bmatrix} \lambda_0 & & & \\ & \lambda_1 & & 0 \\ & & \ddots & \\ & 0 & & \lambda_n \end{bmatrix} , \quad \lambda_0 < \lambda_1 < \ldots < \lambda_n , \tag{B.21}$$

and modal (eigenvector) matrix

$$E = \left[\begin{array}{cccc} | & | & & | \\ e_0 & e_1 & \cdots & e_n \\ \downarrow & \downarrow & & \downarrow \end{array} \right]. \tag{B.22}$$

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_o is the smallest eigenvalue of S; see (B.21). Thus

$$\text{Minimum total quadratic error} = \min_C \{A_0^T S A_0\} = \lambda_0 , \tag{B.31}$$

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

$$A_0 = \frac{\pm S^{-1} e_o}{e_o^T S^{-1} e_o} = \pm \frac{\lambda_o^{-1} e_o}{e_o^T \lambda^{-1}{}_o e_o} = \pm e_0 , \tag{B.32}$$

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.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
ASN (RE&S)	1
OUSDR&E (Research & Advanced Technology)	2
Deputy USDR&E (Res & Adv Tech)	
OASN, Spec Dep for Adv Concept	1
OASN, Dep Assist Secretary (Res & Adv Tech)	1
ONR, ONR-100, -200, -102, -480, -481, -222,	6
CNO, OP-090, -902	2
CNM, MAT-08T, -08T2, -08T24, SP-20, ASW-122	5
DIA, DT-2C	1
NAV SURFACE WEAPONS CENTER, WHITE OAK LAB.	1
DWTNSRDC ANNA	1
DWTNSRDC BETH	1
NRL	1
NRL, USRD	1
NRL, AESD	1
NORDA (Dr. R. Goodman, 110)	2
USOC, Code 241	1
Acoustics Research Branch, Code 240	1
OCEANAV	1
NAVOCEANO, Code 02	1
NAVELECSYSCOM, ELEX 03	1
NAVSEASYSCOM, SEA-003	1
NASC, AIR-610	1
NAVAIRDEVCEN	1
NAVAIRDEVCEN, Code 2052	1
NOSC	1
NOSC, Library, Code 6565	1
NAVWPNSCEN	1
NCSC	1
CIVENGRLAB	1
NAVSURFWPNCEN	1
NUWES, KEYPORT	1
NUWES, San Diego Detachment	1
FLTASWTRACENPAC Tactical Library	1
NAVPGSCOL	1
NAVTRAEQUIPCENT, Technical Library	1
APL/UW, SEATTLE	1
ARL/PENN STATE, STATE COLLEGE	1
CENTER FOR NAVAL ANALYSES (ACQUISITION UNIT)	1
DTIC	12
DARPA	1
NOAA/ERL	1
NATIONAL RESEARCH COUNCIL	1
WEAPON SYSTEM EVALUATION GROUP	1
WOODS HOLE OCEANOGRAPHIC INSTITUTION	1
ENGINEERING SOCIETIES LIB, UNITED ENGRG CTR	1
NATIONAL INSTITUTE OF HEALTH	1
ARL, UNIV OF TEXAS	1
MARINE PHYSICAL LAB, SCRIPPS	1