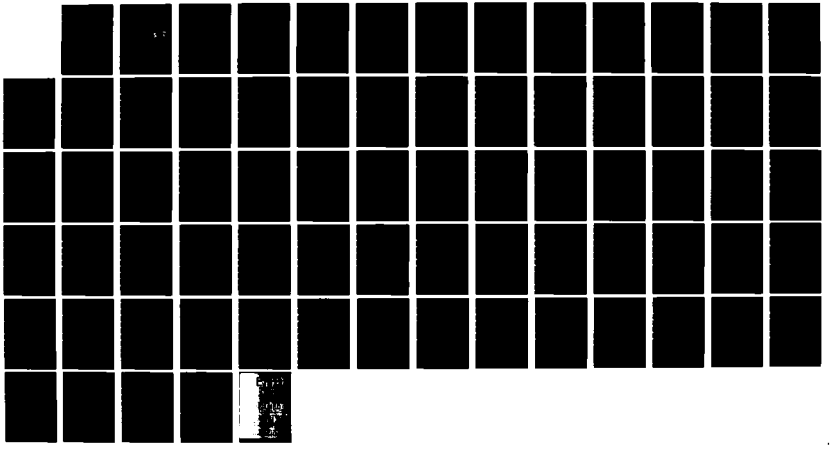
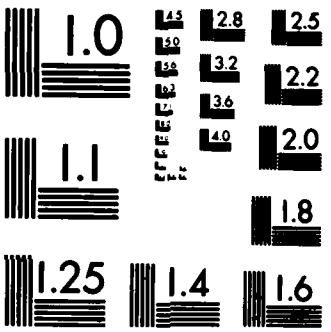
AD-A128 278 OVERSPECIFIED NORMAL EQUATIONS FOR SPECTRAL ESTIMATION 1/1
(U) RESIN RESEARCH LABS INC NEWARK N J D IZRAELEVITZ
JUN 83 TR-498 N00014-81-K-0742

UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

"Overspecified Normal Equations for Spectral Estimation"

David Izraelevitz

Technical Report No. 498

June 1983

M.I.T., R.L.E.

This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 498	2. GOVT ACCESSION NO. A128278	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) "Overspecified Normal Equations for Spectral Estimation"		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David Izraelevitz		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0742
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Laboratory of Electronics Massachusetts Institute of Technology Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-049-506
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22217		12. REPORT DATE June 1983 13. NUMBER OF PAGES 69
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Mathematical and Information Sciences Division, 800 North Quincy Street Arlington, VA 22217		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)

spectral estimation
autoregressive modeling
array processing

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

see other side

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

There is a one-to-one relationship between a set of P normalized positive definite correlation estimates and the P predictor coefficients derived using autoregressive modeling. Several researchers have proposed the use of M > P correlation estimates to provide a better Pth order model. Specifically, the normal equations are augmented to provide M linear equations between the correlation estimates and the predictor coefficients. Since the system of equations is now overspecified, a least squares solution is required.

In this thesis a study is presented of some of the properties of the method of overspecified normal equations as applied to the problem of spectral estimation. The main contribution of this thesis is the derivation of the relationships between the number of correlations used, the model order and the signal to noise ratio of the signal, to the characteristics of the resulting spectral estimate. The characteristics studied are the spectral height, bandwidth and area. The method is shown to be a spectral density estimator like the ME method, where spectral *areas* rather than spectral values should be interpreted as estimates of power.

The relationships derived point to the number of correlations used over the minimum, i.e. model order, as an signal-to-noise enhancer. The resulting spectrum is equivalent to the ME spectrum under higher signal-to-noise conditions. Another result is the requirement of a proportionality constant dependent on the number of correlations and the model order which is necessary for unbiased signal-to-noise measurements. This constant is not required however, for measurements of relative power within the same spectral estimate, as in the power ratio of two sinusoids in noise.

The second part of the thesis presents some empirical studies using computer simulations which verify the theoretical predictions and provide the region of validity of the analysis. Further experiments study the interfering effect of several closely spaced sinusoids. The method of overspecified normal equations is shown to be much more sensitive to this interference than the ME method. Finally, some further empirical studies are made of the resolution capabilities of the method. Using the data derived, an empirical model is derived which seems to agree to some extent with the data.

Overspecified Normal Equations for Spectral Estimation

by

David Izraelevitz

Submitted to the Department of Electrical Engineering and Computer Science,
June, 1983, in partial fulfillment of the requirements for the degree of
Master of Science

ABSTRACT

There is a one-to-one relationship between a set of P normalized positive definite correlation estimates and the P predictor coefficients derived using autoregressive modeling. Several researchers have proposed the use of M > P correlation estimates to provide a better Pth order model. Specifically, the normal equations are augmented to provide M linear equations between the correlation estimates and the predictor coefficients. Since the system of equations is now overspecified, a least squares solution is required.

In this thesis a study is presented of some of the properties of the method of overspecified normal equations as applied to the problem of spectral estimation. The main contribution of this thesis is the derivation of the relationships between the number of correlations used, the model order and the signal to noise ratio of the signal, to the characteristics of the resulting spectral estimate. The characteristics studied are the spectral height, bandwidth and area. The method is shown to be a spectral density estimator like the ME method, where spectral *areas* rather than spectral values should be interpreted as estimates of power.

The relationships derived point to the number of correlations used over the minimum, i.e. model order, as an signal-to-noise enhancer. The resulting spectrum is equivalent to the ME spectrum under higher signal-to-noise conditions. Another result is the requirement of a proportionality constant dependent on the number of correlations and the model order which is necessary for unbiased signal-to-noise measurements. This constant is not required however, for measurements of relative power within the same spectral estimate, as in the power ratio of two sinusoids in noise.

The second part of the thesis presents some empirical studies using computer simulations which verify the theoretical predictions and provide the region of validity of the analysis. Further experiments study the interfering effect of several closely spaced sinusoids. The method of overspecified normal equations is shown to be much more sensitive to this interference than the ME method. Finally, some further empirical studies are made of the resolution capabilities of the method. Using the data derived, an empirical model is derived which seems to agree to some extent with the data.

Thesis Supervisor: Jae S. Lim

Title: Associate Professor of Electrical Engineering

To my parents

Aquí me pongo a cantar,
Al compás de la vigüela,
Que el hombre que lo desvela
Una pena extraordinaria,
Como el ave solitaria,
Con el cantar se consuela.

from "Martin Fierro" by José Hernandez

ACKNOWLEDGEMENTS

There is an important reason why the acknowledgements should be presented at the beginning of the thesis rather than at the end. Displaying them at the beginning should serve as a reminder to the author that the people hereby mentioned not only provided the help needed to complete the thesis, but more importantly motivated and guided the author in the critical period of defining and developing a topic.

My sincere thanks go to Professor Jae Lim for the support, advice and focus required to begin and finish this thesis. His steady guidance was a balancing force to my constant meandering. Professor Bruce Musicus provided my thesis with a firm basis in the literature by pointing out the work of Professor Cadzow. I would also like to thank Farid Dowla for graciously accepting the sometimes painful task of listening to my ideas. Tom Bordley, Webster Dove, and Greg Duckworth displayed on countless occasions the patience necessary to introduce me to our computer facilities. Finally, I'd like to thank my parents, without whose love, understanding and Sunday phone calls, this thesis would not have been completed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Motivation

A very important problem in signal processing is that of information "extrapolation." For example, what may be available is a small set of measurements from which it is desired to make some reasonable guess of what more measurements would have given. By necessity some restrictions must be imposed on the postulated structure of the data in order to restrict its possible behavior. Otherwise, it is not possible to predict future measurements with confidence. The restrictions imposed on the data under study comprise a model. A model can be thought of as a structure for the data with some free parameters which are then tuned to the specific situation. Although the structure may seem too restrictive, it is hoped that by varying the free parameters a reasonable representation can be ascertained.

A model commonly applied to time series data is the autoregressive (AR) model. The basic tenet of this model is that the signal under study is composed of a linear combination of previous signal values plus an independent white noise sequence which allows for some uncertainty in the actual value the signal will take on at any given time. The number and values of the weighing coefficients of previous samples, and the power of the noise sequence, are the means by which we can tune and summarize the signal's structure.

The literature is rich with methods of extracting these model parameters from a data segment, [10]. Most of these can be considered as implementations of the maximum entropy (ME) method introduced by Burg, [1], for the case of one dimensional signals with given contiguous correlations. For this case, the ME method exploits a matrix relationship between the autocorrelation function of an AR sequence and the weighing coefficients of the model. Specifically, there is a one-to-one mapping, up to a scale factor, between the first P correlation coefficients and the P weighing coefficients for a Pth order model. Different methods, such as the autocorrelation and covariance methods, differ on how the autocorrelation sequence is estimated from the data.

There has been much attention given recently to a new method for AR modeling introduced by Cadzow, [2], which can be considered as an extension of the classical AR or ME methods described above. This new method will be referred to in this thesis as the overspecified normal equation (OSNE) method. The method proposed uses many more correlation estimates than the order of the model postulated. Although this method was originally proposed for autoregressive moving average (ARMA) modeling, several authors have recently noted the advantages of the OSNE method for purely autoregressive modeling, [7], [3]. Among these advantages is the method's improved resolution of spectral peaks for the same model order as the ME method, while being less sensitive to correlation estimate inaccuracies than larger order ME model estimates.

The ME and OSNE methods have been applied to the problem of estimating pure tones in noise. The procedure followed is to first extract, using the ME or OSNE methods, an AR model of the signal. From this model the power spectrum, or the Fourier transform of the signal autocorrelation sequence, is estimated. Finally, using this power spectrum estimate, the important characteristics of the sinusoids such as frequency location and power are extracted.

The family of signals of sinusoids in noise is of much importance in the study of monochromatic plane waves being measured by an array of sensors, [5]. These signals also provide a means of characterizing some properties of spectral estimation methods, such as frequency bias and the ability to resolve two sinusoids very close in frequency.

As noted earlier, the link between the data modeling techniques and the estimation of sinusoids in noise is provided by the power spectrum. It has been shown that sinusoids in noise cannot be represented by a finite order AR model, [14]. Therefore, the power spectral estimate that is extracted will never be that of sinusoids in noise as required. In this case, some sort of guidelines must be provided to interpret the power spectrum that is calculated; how to find the sinusoidal power, for example. This problem was studied for the ME case by Lacoss, [8], where he showed that the ME spectrum should be regarded as a spectral density estimator; that is, areas underneath spectral peaks, rather than the peak values themselves are proportional to the power in the spectral region.

In this thesis, the spectral characteristics of the OSNE method are studied. This work provides the necessary link by which power spectral estimates using the OSNE method can be properly interpreted to estimate sinusoid power. In addition, the behavior of the spectral estimate is studied as a function of the parameters of the algorithm and the characteristics of the signal; the order of the model used, the number of correlations used, and the signal-to-noise ratio. Theoretical results are presented for the case of a single complex exponential in noise. These results are then generalized under some conditions to several real sinusoids in noise. The validity of the mathematical approximations and assumptions in the analysis are studied using computer simulations on ideal correlation estimates. Finally, empirical studies are made of the resolution properties of OSNE, that is, under what conditions of signal-to-noise ratio and frequency separation, the OSNE method will be able to detect two distinct sinusoids instead of one.

1.2 Thesis outline

The classic results in spectral analysis using AR modeling are introduced and discussed in chapter 2. The normal equations for the estimation of the model parameters are derived.The well-known results by Lacoss on the shape of the AR spectrum for the estimation of sinusoids in noise are presented. Following a discussion of the paper by Satorius and Ziedler, the Lacoss results are generalized to several sinusoids in noise. Finally, the method of overspecified normal equations due to Cadzow is introduced.

The bulk of the contributions of this thesis are presented in chapter 3. Here the main theoretical results concerning the spectral shape of the OSNE spectrum are derived and discussed.

Chapter 4 contains all the computer simulations which verify the results presented in chapter 3. Furthermore, empirical studies are presented concerning the resolution capabilities of OSNE.

Chapter 5 summarizes the main results of the thesis and provides the conclusions that can be drawn about the behavior of OSNE as a spectral estimator.

CHAPTER 2

BACKGROUND

2.1 Introduction

In this chapter we review some of the classic results in spectral analysis using AR modeling. The presentation begins with a brief derivation of the classical method of extracting model parameters from a data sequence. This method can be considered as an implementation of the maximum entropy (ME) method introduced by Burg [1], for the case of one dimensional signals with given contiguous correlations.

We then restrict our study to the spectral analysis of sinusoids buried in white noise. This will allow for the description of the behavior of different spectral estimation methods when directed toward this widely encountered application of spectral analysis. Here some classic results by Lacoss on the shape of the ME model spectrum for a single complex exponential in noise are discussed. In the case of real data applications, isolated complex exponentials are not found, but rather appear as pairs at symmetric locations about zero frequency. Thus the Lacoss results need to be generalized to several complex exponentials. This is done in a discussion of the results by Satorius and Ziedler,[13], where a peak decoupling property of ME spectra is discussed. Their paper shows that for well separated sinusoids, the ME method estimates each peak separately. This will basically extend the Lacoss results to the case of well separated spectral peaks.

After considering the ME modeling of spectra, the method of overspecified normal equations, first espoused by Cadzow,[2], for ARMA modeling, is introduced. Here a large number of correlation estimates are used to extract the AR model parameters.

2.2 AR modeling and normal equations

Parametric signal modeling, and more specifically autoregressive (AR) modeling of discrete time sequences, has found many important applications in such areas as phased array radar [5], seismic data analysis [12], and speech processing [11]. The basic philosophy of parametric signal modeling is to extract from a finite data segment the underlying characteristics of the signal as expressed by a small set of parameters, and therefore be able to either compress the information presented by this data, or extrapolate the signal, thus "enlarging" the data window.

It is in this last ability that this thesis is basically interested. Since the application at hand is the extraction of power and frequency characteristics of a sinusoid corrupted by noise, enlarging the data window will make the presence of the sinusoid more noticeable, so that we may better estimate its frequency and power.

AR modeling assumes that the present signal sample is well modeled as a time-invariant linear combination of the last P signal samples plus an independent white noise term. This structure may seem restrictive but it has been used successfully in the modeling of speech and estimation of the characteristics of sinusoids buried in noise. The weighing coefficients of previous samples and the noise variance are the characteristic parameters which summarize the signal's structure.

Below is a brief derivation of the relationship between the correlations of an AR sequence and the corresponding model parameters. It is by no means the only derivation available. The reader is referred to the paper by Makhoul [10] for an excellent treatment of the various approaches available to the AR modeling problem. The approach used here is a statistical one, rather than a more data-oriented approach of minimum prediction error for one important reason. It is the intent of the derivation to decouple the relationship between *exact* correlations and model parameters, from the important and difficult problem of properly estimating the correlation sequence from the data. As explained below, different correlation estimates lead to the autocorrelation and covariance methods.

According to the AR model, the signal under study $s[n]$ is postulated to be of the form,

$$s[n] = -\sum_{i=1}^{p} a_i s[n-i] + w[n] \tag{2.2.1}$$

where the a_i are the model parameters, also called predictor coefficients, and $w[n]$ is an additive, zero-mean, white noise sequence.

Let us now multiply both sides of (2.2.1) by $s[n-k]$ where $k>0$, and then take ensemble expectations.

$$E\{s[n]s[n-k]\} = E\{-\sum_{i=1}^{p} a_i s[n-i]s[n-k] + w[n]s[n-k]\} \tag{2.2.2}$$

Using the linearity properties of the expectation operator, and defining the correlations as

$$R(-l,-m) = E\{s[n-l]s[n-m]\} \tag{2.2.3}$$

yields

$$R(0,-k) = -\sum_{i=1}^{p} a_i R(-i,-k) + E\{w[n]s[n-k]\} \tag{2.2.4}$$

We now note that the term $E\{w[n]s[n-k]\}$ should be zero under the model in (2.2.1). The signal sample $s[n-k]$ is a function of previous signal samples, i.e., $s[n-k-1]$, $s[n-k-2]$, etc. and the noise term $w[n-k]$. However, $w[n]$ is added to the sequence *after* $s[n-k]$ occurs since by assumption $k>0$. Therefore, $w[n]$ and $s[n-k]$ are independent. Since $w[n]$ is zero-mean, the expected value of the product $w[n]s[n-k]$ should be zero as well.

Thus we are left with the following exact relationship between the signal correlations $R(i,k)$ and the predictor coefficients a_i, the so called normal equations

$$R(0,-k) = -\sum_{i=1}^{p} a_i R(-i,-k) \qquad k>0 \tag{2.2.5}$$

There are several important comments to be made at this point. The relationship expressed in (2.2.5) is an interesting one, but of little practical value at this stage. What it expresses is not a relationship between the data and the model parameters, but rather between some statistical properties of the data sequence and the model parameters. Thus prior to applying this relationship the statistical properties of $s[n]$ must be estimated. In other words one of the first steps in applying this result will be to decide how $R(i,k)$ should be estimated from the data. Secondly, (2.2.5) is valid for all $k>0$, while there is only a finite number of unknowns. Thus (2.2.5) is actually an infinite number of simultaneous linear equations which need to be solved for the unknown a_i. Finally, there is a striking simi-

larity in structure between (2.2.5) and (2.2.1). By assuming that the signal $s[n]$ is stationary, the relationship between (2.2.1) and (2.2.5) can be strengthened. By stationarity we mean that the statistical characteristics of the signal do not change over time. Since the statistical properties of the signal are time-invariant, any statistical relationship between two signal samples must depend only on the time interval between the two signals. In the case of the correlation between two signal samples, $R(-i,-k)$, must depend only on $(k-i)$ or $R(-i,-k) = R(k-i)$. Under the additional assumption of stationarity then, (2.2.5) becomes

$$R(k) = -\sum_{i=1}^{p} a_i R(k-i) \qquad k>0 \tag{2.2.6}$$

In this case, the correlation sequence obeys the same recursive relationship as the signal sequence except that the noise sequence is missing. Another difference is that the recursive relation in (2.2.6) is valid only for $k>0$.

As noted previously, (2.2.5) and (2.2.6) are actually overdetermined sets of linear equations. Were it not for the assumption that the $R(i,k)$ used are exact, and that $s[n]$ is a Pth order AR process, there would probably be no set of a_i that would satisfy (2.2.5) for all positive k. In the classical or ME approach to the extraction of the predictor coefficients, the above problem is avoided by only using (2.2.5) for k in the range of 1 to P. Thus there are P equations for P unknowns and a solution for the predictor coefficients exists for any well-behaved set of $R(i,k)$.

The ME method can be expressed succinctly in matrix notation. Below, several entities are defined by specifying what each element of the matrix or vector should contain

$$\begin{aligned} [\mathbf{R}]_{ij} &= R(-i,-j) && (P \text{ by } P) \\ [\mathbf{r}]_i &= R(0,-i) && (P \text{ by } 1) \\ [\mathbf{a}]_i &= a_i && (P \text{ by } 1) \end{aligned}$$

Using the above notation, the relationship between exact correlations and predictor coefficients in (2.2.5) can be expressed as

$$\mathbf{Ra} = -\mathbf{r} \tag{2.2.7}$$

Different methods, such as the autocorrelation method and the covariance method, use different estimates of $R(i,k)$ or $R(k-i)$, but exploit the same relationship between the correlations and the

predictor coefficients as expressed in (2.2.7). For example, the autocorrelation method uses

$$R(k) = \sum_{n=0}^{N-1-k} s[n]s[n-k] \tag{2.2.8}$$

as an estimate of the stationary correlation sequence, while the covariance method uses

$$R(-i,-k) = \sum_{n=p}^{N-1} s[n-i]s[n-k] \tag{2.2.9}$$

as an estimate of the correlation values where N is the length of the data sequence.

Once the predictor coefficients have been estimated, the noise variance estimate is found in turn using (2.2.4) but for the case when $k = 0$. In this case the value of $s[n]$ and $w[n]$ are not uncorrelated, but have a covariance equal to the variance of $w[n]$. What results in this case is the following relationship between $R(i,k)$, the a_i set and the noise variance σ_w^2,

$$R(0,0) = -\sum_{i=1}^{p} a_i R(-i,0) + \sigma_w^2 \tag{2.2.10}$$

After all the model estimation calculations have been completed, the power spectral estimate for $s[n]$ can be calculated. If the Fourier transform of the predictor coefficient set is defined by

$$A(\omega) = 1 + \sum_{k=1}^{p} a_k e^{-j\omega k} \tag{2.2.11}$$

then the corresponding power spectrum for $s[n]$ is given by

$$\hat{S}(\omega) = \frac{\sigma_w^2}{|A(\omega)|^2} \tag{2.2.12}$$

The basic idea in using AR modeling for the estimation of sinusoids in noise is that we expect the $\hat{S}(\omega)$ given by (2.2.12) to have a large amount of energy around the frequency of the sinusoid and that the shape of the spectrum around that frequency should somehow be related to the power in that sinusoid.

2.3 Discussion of results by Lacoss

Although the AR modeling procedure is rather simple, requiring basically some evaluations of lagged products and the solution of a set of simultaneous linear equations, what is not obvious is how to interpret the results. As explained in chapter 1, the power spectrum provides the link between data modeling techniques and the estimation of properties of sinusoids in noise. In conventional spectral estimation techniques, such as Welch's method, [15], and in the maximum likelihood method (MLM) introduced by Capon, [4], the value of the spectrum at each frequency is interpreted as an estimate of the power at that frequency.

In a classic paper, Lacoss has shown, [8],that the ME method behaves in a radically different manner than either conventional spectral estimation techniques or MLM. The ME spectral estimate should not be interpreted on a spectral value basis. The value at a spectral peak, Lacoss shows, is proportional to the square of the sinusoid signal power. However, the area underneath the peak, as defined by the product of the spectral peak value and its half-power bandwidth, is proportional to the peak value.

The starting point of his discussion is to assume that we are considering a complex sequence composed of a single complex exponential of radian frequency ω_0 and an additive, zero-mean, white noise sequence, $w[n]$, i.e.

$$s[n] = \sqrt{b}e^{j\omega_0 n} + w[n] \tag{2.3.1}$$

Note that $s[n]$ is also zero-mean as well as stationary. The exact correlation sequence for the above signal is given by

$$R(k) = \delta(k) + be^{j\omega_0 k} \tag{2.3.2}$$

where $\delta(k)$ is the unit sample sequence. If we now insert the above values for $R(k)$ into equations (2.2.7) and (2.2.9), we can solve for the a_i's and ω_p^2. After inserting in (2.2.12), we get the estimated spectrum to be

$$\hat{S}(\omega) = \frac{\frac{1 + b(P+1)}{1 + bP}}{\left|1 - \frac{bP}{1+bP} f_P(\omega_0 - \omega)\right|^2} \tag{2.3.3}$$

where

$$f_P(\lambda) = \frac{1}{P}\sum_{k=1}^{P} e^{-j\lambda k} = e^{j\frac{P+1}{2}\lambda}\frac{\sin P\lambda}{P\sin\lambda} \tag{2.3.4}$$

The notation used above is slightly different from that used by Lacoss in his paper, however it will simplify later comparisons if the results are expressed in this form.

It is readily seen that the value of $\hat{S}(\omega)$ at ω_0 is

$$\hat{S}(\omega_0) = (1 + bP)(1 + b(P + 1)) \approx b^2P(P + 1) \quad \text{for } bP \gg 1 \tag{2.3.5}$$

Thus the value of $\hat{S}(\omega)$ at ω_0 is not proportional to the power in the sinusoid, but actually proportional to b^2. Also, we see that the spectral maximum occurs at $\omega = \omega_0$ so at least for exact correlations, the frequency estimate is unbiased. To further characterize the shape of the spectral peak, Lacoss found an approximation to the bandwidth of the spectral peak using a Taylor series approximation about the peak value. This method yields the following result for the half-power bandwidth of the AR spectrum , for large P,

$$\omega_H \approx \frac{4}{b(P + 1)^2} \tag{2.3.6}$$

Thus the bandwidth of the spectrum decreases as the signal to noise ratio increases, i.e. b, as well as decreasing as the square of the model order. Lacoss proposed the product of $\hat{S}(\omega_0)$ and ω_H as an estimator of the power in the sinusoid. This product we see is basically proportional to b independent of model order.

From (2.3.3) a bias in the noise variance estimate is found. As b goes to infinity, while P stays small, the noise variance estimate does not go to 1, but rather goes to $1 + \frac{1}{P}$. This will introduce a bias in subsequent calculations, especially, as we will find later, this bias increases as the number of sinusoids under study increases. This point was not made in the paper by Lacoss, since a bias in the noise variance estimate will not affect *relative* power measurements, i.e., the ratio in power between two sinusoids.

To avoid this bias, subsequent discussions on the AR model spectrum will concentrate on the "normalized" spectrum $S^N(\omega)$ given by

$$S^N(\omega) = \frac{1}{|A(e^{j\omega})|^2} \tag{2.3.7}$$

In this case, the Lacoss results for the spectral height must be modified slightly. The value of the spectral maximum of $S^N(\omega)$ for a single complex exponential in noise is given by

$$S^N(\omega_0) \approx b^2P^2 \qquad \text{for } bP \gg 1 \tag{2.3.8}$$

2.4 Discussion of spectral peak decoupling property

It is important to keep in mind in interpreting the Lacoss results, that the signal under consideration is a complex exponential. Real sinusoids can be thought of as pairs of complex exponentials and therefore we must be careful to examine where and how the results discussed above apply to real signals.

Although Lacoss studied the interfering effects of multiple complex exponentials on his results, a more elegant method was proposed by Satorius and Ziedler,[13], for studying the characteristics of the ME spectrum for several sinusoids again assuming that exact correlations are available. Using the method of undetermined coefficients, which will be discussed in detail in chapter 3, they showed that $A(\omega)$ in (2.2.12) for a signal consisting of L complex exponentials located at $\omega_1, \ldots, \omega_L$

with power $b_1, \ldots, b_L$ is given by

$$A(\omega) = 1 - \sum_{i=1}^{L} f_p(\omega - \omega_i) d_i \qquad (2.4.1)$$

where $f_p(\,)$ is the same defined in (2.3.4) and the d_i are unknown constants. An important property to take note here is that $f_p(\lambda)$ is a $\frac{\sin x}{x}$ (sinc) type function which goes to zero as λ increases. Thus $A(\omega)$ is a weighed sum of sinc functions centered at the frequency of each sinusoid. The constants d_i provide the weighing of each sinc function. Therefore, for sufficiently well separated ω's, (2.4.1) will look like

$$A(\omega) \approx 1 - f_p(\omega - \omega_i) d_i \qquad (2.4.2)$$

for $\omega \approx \omega_i$ which is independent of both the number of sinusoids L and their frequencies. However, if the value of d_i depended on the characteristics of the other exponentials, the decoupling would not occur. Fortunately, the paper by Satorius and Ziedler shows that for well separated sinusoids, the value of d_i doesn't depend on the location or power of the other sinusoids. In this case, a simple interpretation of the d_i is available. Consider (2.4.2) for the case when $\omega = \omega_i$. Then, $A(\omega)$ will equal $1 - d_i$. Thus, as d_i gets closer to 1, the spectrum will display a peak at ω_i.

It is the variations of $A(\omega)$ which provide the fluctuations and detail as a function of ω for the

AR spectrum. The estimate of the noise variance is only a proportionality constant. Thus for well separated frequencies, equation (2.4.1) extends the Lacoss results to the shape of each individual spectral peak.

As discussed previously, the noise variance estimate is biased for the single complex exponential in noise case. For several sinusoids in noise, the problem is even worse. For well separated sinusoids, the noise variance is given by

$$\sigma_w^2 = 1 + \sum_{i=1}^{L} \frac{b_i}{b_i P + 1} \tag{2.4.3}$$

As P goes to infinity, the value of σ_w^2 goes to the right value 1, but as b_i goes to infinity, σ_w^2 goes to

$$\sigma_w^2 = 1 + \frac{L}{P} \tag{2.4.4}$$

which deviates from 1 more and more as the number of sinusoids being estimated increases. What we can expect from (2.4.4) is that the measure proposed by Lacoss for the signal power will be off by a constant amount as the SNR goes to infinity. For this reason, the estimation of the noise power will not be dealt with further in this thesis, since computer experiments have shown that this bias exists in the OSNE spectrum as well. Without an absolute power measurement, we are forced to restrict our measurement to that of relative power, that is power ratio between two sinusoids in the same spectrum, or the calculation of signal to noise ratio.

2.5 Overspecified normal equations (OSNE)

In this section we will consider a different use of the equations derived in section 2.2 for the purpose of spectral estimation. The key observation to make is that the relationships between the true correlations and the predictor coefficients expressed in (2.2.5) and (2.2.6) hold not only for k between 1 and P, but really hold for all positive k. As described previously, the ME method uses just enough simultaneous linear equations to make the solution unique.

If the model used is the right one, that is if $s[n]$ is a Pth order AR process, and if the correlations are estimated exactly, then the P equations used in the ME method are all that is needed. Subsequent equations, those in (2.2.5) for $k>P$, will not add any more information.

Of course in real applications, neither assumption is satisfied. In fact, even if we are quite confident in using a Pth order AR model for the signal under study, the correlations will never be estimated exactly. Thus, we can be sure that the higher index equations will certainly provide different, although possibly not better, information regarding the predictor coefficients.

The method of overspecified normal equations can therefore be stated as finding a set of predictor coefficients which will minimize in some way the difference between the two sides of (2.2.5) or (2.2.6) for the case of inexact correlation estimates or inaccurate model. If the maximum available index difference in the correlation estimate is M, then the number of available linear equations in (2.2.5) or (2.2.6) is M. To provide some mathematical tractability, a quadratic norm of the difference between the left and right sides of (2.2.5) or (2.2.6) is used as the function to be minimized. Additionally, Cadzow originally proposed the use of a weighing function, arguing that it would be expected that higher correlation estimates would be less accurate than lower correlations, hence implying that the differences in (2.2.5) for equations using higher correlations should be weighed less in the objective function.

We can again express this problem succinctly in matrix notation. The same notation as in (2.2.7) will be used to define R, r, and a. Let the P by M matrix $\mathbf{I}_{MP}$ be defined as the concatenation of a P by P identity matrix and a P by $(M-P)$ zero matrix. That is

$$\mathbf{I}_{MP} = \begin{bmatrix} \mathbf{I}_{PP} \\ \mathbf{0} \end{bmatrix} \qquad (M \text{ by } P) \tag{2.5.1}$$

The basic property of this matrix is that it will add $M-P$ zeroes to a vector of dimension P. Following this notation, $\mathbf{I}_{PP}$ is a P by P identity matrix.

The OSNE problem can be phrased as finding the least squares solution to

$$\mathbf{R}\mathbf{I}_{MP}\mathbf{a} \approx -\mathbf{r} \tag{2.5.2}$$

In this case the dimension of R in (2.5.2) is M by M while in (2.2.7) it was P by P. Similarly, the r vector is now M by 1. The effect of the $\mathbf{I}_{MP}$ matrix can be interpreted as either attaching $M - P$ zeroes to a or "chopping off" the last $M - P$ columns of R. According to the Orthogonality Principle, the least squares solution to (2.5.2) must satisfy

$$\mathbf{I}_{MP}{}^{H}\mathbf{R}\mathbf{R}^{H}\mathbf{I}_{MP}\mathbf{a} = -\mathbf{I}_{MP}^{H}\mathbf{R}^{H}\mathbf{r} \tag{2.5.3}$$

In the above equation the superscript H denotes the Hermitian operation on the corresponding matrix.

The use of a different number of predictor coefficients and equations permits a trade-off between resolution and statistical stability [7]. It has been observed empirically that using M correlations and a Pth order model yields resolution capabilities which lie somewhere between using a Pth order model and an Mth order model by the ME method.

As is usually the case in spectral estimation methods, there is an inherent tradeoff in statistical stability and spectral resolution. That is, in general the higher resolution an algorithm has, the more sensitive it is to the accuracy of the correlation measurements. Thus any variation in the correlation estimates is magnified in the resulting spectrum for higher resolution algorithms. In adherence to this general principle, the OSNE spectrum using a Pth order model and M equations is found to be more statistically stable than the ME spectrum using an Mth order model, but is less statistically stable than the ME spectrum using an Pth order model, in opposite relation to the resolution performance.

The above statements on statistical stability have been purposely vague; their intent is to give a qualitative assessment of the stochastic behavior of the algorithms. Of course if the stochastic behavior of the OSNE method were to be considered in depth, a definition would be needed for the term statistical stability, whether this refers to variability in spectral peak location, or model parameter value, or

model pole locations. Such considerations, although of paramount importance, are outside the scope of this thesis. A partial analysis of the ME method characteristics was conducted by Lang [9] to whose work the reader is referred.

CHAPTER 3

BEHAVIOR OF OSNE AS A SPECTRAL ESTIMATOR

3.1 Introduction

As explained in previous sections, the power spectral estimate provides the link between data modeling techniques and the estimation of sinusoids in noise. The purpose of this thesis is to develop the link connecting the OSNE method and the sinusoid in noise problem in analogy to the work by Lacoss, Satorius and Ziedler for the ME method.

To provide the proper background to interpret the OSNE spectrum the following questions need to be answered,

1. Does OSNE behave like conventional spectral estimates where spectral *values* correspond to spectral power, or does it behave like the ME method where spectral *areas* correspond to power?

2. How do spectral characteristics such as peak value and bandwidth vary as a function of algorithm and signal parameters; that is, the model order, the number of correlations used, and the sinusoid powers?

3. Under what conditions can separate peaks be treated individually; that is, does OSNE satisfy the property discussed by Satorius and Ziedler for the ME method?

In this chapter the theoretical results are developed to help answer these questions.

3.2 Analysis of the OSNE spectral estimate

The problem to be discussed in this section is the description of the OSNE spectrum for the case of several sinusoids in noise. The problem will be considered both for the case of a single complex exponential in noise and the case of several sinusoids. The discussion begins with the consideration of the OSNE spectrum for several sinusoids and the decoupling property of the OSNE spectral analysis is derived. This will motivate the subsequent discussion of the OSNE spectrum of a single complex exponential in noise. In considering the single complex exponential case, a fundamental result will be found linking the characteristics of the ME spectral peak and the OSNE spectral peak.

Consider the case where the exact correlations for $L/2$ sinusoids in noise are available. This is the same case considered in the Satorius and Ziedler paper. The radian frequency of each sinusoid will be denoted by ω_i, and the corresponding amplitude by $2\sqrt{b_i}$. Since scaling both sinusoid powers and noise power by the same amount does not affect the calculation of the a_i set, the noise variance will be assumed to equal 1.

In the above situation, the correlation between a data sample at n and a data sample at $n + k$ is independent of n and is given by

$$R(k) = \delta(k) + \sum_{i=1}^{L/2} 2b_i \cos \omega_i k \tag{3.2.1}$$

Since a cosine function consists of two complex exponentials, (3.2.1) can be rewritten as a sum of weighed complex exponentials. Letting $\omega_{i+\frac{L}{2}} = -\omega_i$, and $b_{i+\frac{L}{2}} = b_i$, we get an equivalent expression for the correlation sequence of $L/2$ sinusoids in noise,

$$R(k) = \delta(k) + \sum_{i=1}^{L} b_i e^{j\omega_i k} \tag{3.2.2}$$

We see in this case that the signal under study is stationary and therefore the formulation in (2.2.6) will be required to find the OSNE AR model.

The next step is to find the matrix R and the vector r as defined in (2.2.7) corresponding to the autocorrelation sequence in (3.2.2). To express the above relationships in matrix notation some auxiliary matrices will have to be defined. Let the M by L matrix E_M be defined as follows,

$$\mathbf{E}_M = \begin{bmatrix} e^{j\omega_1} & e^{j\omega_2} & \cdots & e^{j\omega_L} \\ e^{j2\omega_1} & e^{j2\omega_2} & \cdots & e^{j2\omega_L} \\ \vdots & \vdots & \cdots & \vdots \\ e^{jM\omega_1} & e^{jM\omega_2} & \cdots & e^{jM\omega_L} \end{bmatrix} \tag{3.2.3}$$

We see that $\mathbf{E}_M$ is composed of powers of the complex exponentials corresponding to each frequency ω_i. The L by L diagonal matrix B contains the power in each complex exponential, i.e.,

$$\mathbf{B} = diag(b_1, b_2, \ldots, b_L) \tag{3.2.4}$$

Finally, let the L by 1 vector j consist of all ones. With all the notational machinery defined above, we can now express succinctly the matrix R defined in section 2.2 corresponding to (3.2.2),

$$\mathbf{R} = \mathbf{I}_{MM} + \mathbf{E}_M \mathbf{B} \mathbf{E}^H_M \tag{3.2.5}$$

Similarly, the vector r can be expressed as

$$\mathbf{r} = \mathbf{E}_M \mathbf{B} \mathbf{j} \tag{3.2.6}$$

After inserting (3.2.5) and (3.2.6) into (2.5.2) we can state the OSNE problem for sinusoids in noise given the sinusoid locations and amplitudes,

$$(\mathbf{I}_{MM} + \mathbf{E}_M \mathbf{B} \mathbf{E}^H_M)\mathbf{I}_{MP}\mathbf{a} \approx -\mathbf{E}_M \mathbf{B} \mathbf{j} \tag{3.2.7}$$

The obvious next step would be to solve directly for the a_i set using (2.5.3). However, inverting the R matrix is a difficult task except for the simple case of $L = 1$ where the Woodbury identity could be used. A more insightful approach was introduced by Satorius and Ziedler and later used by Lang [9] which is similar in philosophy to the method of undetermined coefficients used in solving ordinary linear differential equations. In this situation we make the assumption that the solution to (2.5.2) is given by a linear combination of the columns of $\mathbf{E}_P$. Specifically,

$$\mathbf{a} = -\frac{1}{P}\mathbf{E}_P \mathbf{d} \tag{3.2.8}$$

where d is an L by 1 vector and $\mathbf{E}_P$ has the same form as $\mathbf{E}_M$ but truncated at the Pth row. Substituting (3.2.5), (3.2.6) and (3.2.8) into the least squares solution condition expressed in (2.5.3) yields (after some tedious manipulation) an optimality condition for the unknown vector d,

$$\mathbf{E}_P[\frac{1}{P}\mathbf{I}_{LL} + 2\mathbf{B}\mathbf{F}_P + M\mathbf{B}\mathbf{F}_M\mathbf{B}\mathbf{F}_P]\mathbf{d} = \mathbf{E}_P[\mathbf{I}_{LL} + M\mathbf{B}\mathbf{F}_M]\mathbf{B}\mathbf{j} \tag{3.2.9}$$

where the newly introduced L by L matrices $\mathbf{F}_P$ and $\mathbf{F}_M$ are the matrix equivalent of the function $f_P(\,)$

defined in (2.3.4), and are given by

$$\mathbf{F}_Q = \frac{\mathbf{E}_Q{}^H \mathbf{E}_Q}{Q} \tag{3.2.10}$$

Considering each element of $\mathbf{F}_Q$, we find that

$$[\mathbf{F}_Q]_{ij} = f_Q(\omega_i - \omega_j)$$

The above equation can be simplified by noting that this expression is equivalent to requiring that the expression below,

$$[\frac{1}{P}\mathbf{I}_{LL} + 2\mathbf{B}\mathbf{F}_P + M\mathbf{B}\mathbf{F}_M\mathbf{B}\mathbf{F}_P]\mathbf{d} - [\mathbf{I}_{LL} + M\mathbf{B}\mathbf{F}_M]\mathbf{B}\mathbf{j} = \mathbf{n} \tag{3.2.11}$$

be in the null space of $\mathbf{E}_P$, i.e.,

$$\mathbf{E}_P\mathbf{n} = \mathbf{0} \tag{3.2.12}$$

Since $\mathbf{E}_P$ is a Vandermonde matrix, the null space of $\mathbf{E}_P$ consists of the zero vector only as shown in [6]. Therefore, (3.2.9) is simplified (somewhat) to the following relation for the unknown $\mathbf{d}$ vector,

$$[\frac{1}{P}\mathbf{I}_{LL} + 2\mathbf{B}\mathbf{F}_P + M\mathbf{B}\mathbf{F}_M\mathbf{B}\mathbf{F}_P]\mathbf{d} = [\mathbf{I}_{LL} + M\mathbf{B}\mathbf{F}_M]\mathbf{B}\mathbf{j} \tag{3.2.13}$$

It is from studying (3.2.13) that the properties of the OSNE spectrum will be derived.

Recalling that the matrix $\mathbf{B}$ is diagonal and that M and P are scalars, we see that the interaction between the different elements of the vector $\mathbf{d}$ occurs via the two matrices with non-zero elements in the off-diagonals, $\mathbf{F}_P$ and $\mathbf{F}_M$. As *both* M and P are increased, these two matrices approach the identity matrix $\mathbf{I}_{LL}$ (consider the relationship in (3.2.10)). Thus for large enough P, each value of d_i will be calculated from an algebraic relation which will depend only on M, P, and b_i. The location and amplitude of any other sinusoid will not affect the value of this d_i. This is similar to the result for ME discussed in section 2.4. However, this does not mean that the vector $\mathbf{a}$ will in any way reflect this decoupling. Each a_i is a linear combination of *all* the elements of $\mathbf{d}$, as well as depending on the locations of each sinusoid via the matrix $\mathbf{E}_P$. As we will find later, considering the $\mathbf{d}$ vector rather than the set of predictor coefficients will be a more natural choice for examining the properties of the OSNE spectrum.

This decoupling property of the elements of $\mathbf{d}$ leads directly to the decoupling of the spectral peaks. Recall that the predictor Fourier transform is given by

$$A(\omega) = 1 + \sum_{k=1}^{P} e^{-j\omega k} a_k \tag{3.2.14}$$

Substituting the elements of d for the a_k (3.2.14) becomes

$$A(\omega) = 1 - \sum_{i=1}^{L} f_P(\omega - \omega_i) d_i \tag{3.2.15}$$

where the function $f_P(\,)$ is the now familiar friend previously defined. Equation (3.2.15) is the same as (2.4.1) where the decoupling property of the ME spectrum was discussed. Following the same reasoning as proposed in section 2.4, for well separated sinusoids,

$$A(\omega) = 1 - f_P(\omega - \omega_i) d_i \tag{3.2.16}$$

for $\omega \approx \omega_i$. Again, it must be emphasized that increasing the number of correlations used, i.e. M, alone is not enough to decouple the spectral peaks. The value of P must be large enough for two reasons; first, to make sure that $f_P(\omega - \omega_i)$ in (3.2.16) decreases fast enough, and second, to decouple the calculation of each d_i in (3.2.13). Obviously, increasing P will force an increase in M.

Considering (3.2.16) further, we see that for well separated sinusoids, d_i defines the shape of the peak at ω_i. This value of d_i is calculated as if the data consisted of only one complex exponential at ω_i of power b_i. This leads us to study in detail the case of a single complex exponential in noise, since any results which are derived in this case are applicable under some conditions to multiple sinusoids.

In the case of a single complex exponential in noise, L equals 1, the d vector is a scalar d_1, the matrix B is a scalar b_1 and the $\mathbf{F}_P$ and $\mathbf{F}_M$ matrices become the scalar 1. The value of d_1 is now easily found,

$$d_1 = \frac{(1 + Mb_1)b_1}{\frac{1}{P} + 2b_1 + Mb_1^2} = \frac{1 + Mb_1}{\frac{1}{Pb_1} + 2 + Mb_1} \tag{3.2.17}$$

As a check for the above relation, setting $M = P$ yields

$$d_1 = \frac{b_1 P}{1 + b_1 P} \tag{3.2.18}$$

Comparing this result with the ME spectrum found in (2.3.3), we see that they agree.

For large model order P or large signal-to-noise ratio (SNR) b_1, a rather interesting property

appears; the value of d_1 becomes dependent only on the SNR and the number of correlations used. For large model order P or for large signal power b_1, we can make the following approximations,

$$1+Mb_1 \approx Mb_1 \tag{3.2.19}$$
$$\frac{1}{Pb_1} \approx 0$$

Thus

$$d_1 \approx \frac{Mb_1}{1+Mb_1} \qquad \text{for } b_1 \gg 1 \text{ or } P \gg 1 \tag{3.2.20}$$

What (3.2.20) implies is that for large SNR or P, the value of d_1 is the same that would have been calculated if instead of OSNE, the ME method had been used with an Mth order model. The reader is warned not to infer from this that the spectral peak will have the same shape for OSNE and ME. Recalling (3.2.16), the characteristics of $A(\omega)$ near ω_i depend not only on the value of d_1, but also on the behavior of $f_P(\,)$. For an Mth order ME spectrum, the function would be $f_M(\,)$, not $f_P(\,)$. It is an interesting behavior of the OSNE spectrum that the effect of the model order used is reflected in the function $f_P(\,)$, while the effect of the number of correlations used and SNR is reflected in the value of d_1 solely. Also we can deduce some trade-offs that OSNE provides between the number of correlations and the SNR. Suppose that we consider two situations where the model order used is the same; first, the SNR of the signal is b_1 and the number of correlations used M; second, the SNR of the signal is $\frac{b_1}{k}$ while the number of correlations used is Mk. Considering (3.2.20) and (3.2.16), we see that the value of d_1 remains the same, while the function $f_P(\,)$ is unaffected. Therefore, the shape of the spectral peak will be virtually the same both in spectral value and bandwidth. Thus the effect of increasing the number of correlations is to make the *effective* SNR larger.

The above discussion presented an interpretation of what increasing the number of correlations did in terms of artificially boosting the SNR. Another effect which is of interest is to find what trade-offs occur in varying the model order as well as the number of correlations used. The discussion below provides an interesting result which also provides another link between the ME and OSNE spectra.

Consider now the $A(\omega)$ corresponding to the Mth order ME spectrum; that is, if the M correlation values had been used to extract an Mth order model.

In this case, from (2.3.3)

$$A(\omega) = 1 - \frac{b_1 M}{1 + b_1 M} f_M(\omega_0 - \omega) \tag{3.2.21}$$

where $f_M(\,)$ was previously defined as

$$f_M(\lambda) = e^{j\frac{M+1}{2}\lambda} \frac{\sin M\lambda}{M \sin\lambda} \tag{3.2.22}$$

The task at hand is to find a reasonable approximation to $f_M(\,)$ for small λ. For small λ, $\sin\lambda \approx \lambda$, so $f_M(\,)$ could be considered to depend only on the product $M\lambda$ or $(M+1)\lambda$. Although both approximations are valid for small enough λ, the second choice has a wider region of validity as shown in Table 3.2.1. Thus the approximation we will use is that

$$f_M(\lambda) \approx e^{j\frac{M+1}{2}\lambda} \frac{\sin(M+1)\lambda}{(M+1)\lambda} = f[(M+1)\lambda] \qquad (M+1)\lambda \ll 1 \tag{3.2.23}$$

With the approximations and assumptions discussed above in mind, $A(\omega)$ for an Mth order ME spectrum is approximated by,

$$A_{ME}(\omega) \approx 1 - \frac{Mb_1}{1 + Mb_1} f[(M+1)(\omega - \omega_i)] \qquad (M+1)(\omega - \omega_i) \ll 1 \tag{3.2.24}$$

The above approximations are now applied to the OSNE predictor Fourier transform. Recall (3.2.15) for $L=1$, which provides the predictor Fourier transform for the use of OSNE on a single complex exponential in noise,

$$A_{OSNE}(\omega) \approx 1 - \frac{1+Mb_1}{\frac{1}{Pb_1} + 2 + Mb_1} f_P(\omega - \omega_i) \tag{3.2.25}$$

If we now apply the approximation in (3.2.20) for the value of d_1 for large P or large SNR, we find,

$$A_{OSNE}(\omega) \approx 1 - \frac{Mb_1}{1 + Mb_1} f_P(\omega - \omega_i) \qquad \text{for } b_1 \gg 1 \text{ or } P \gg 1 \tag{3.2.26}$$

Using the approximation in (3.2.23) for the behavior of $f_P(\,)$, a final form is found for the OSNE predictor Fourier transform,

$$A_{OSNE}(\omega) \approx 1 - \frac{Mb_1}{1 + Mb_1} f[(P+1)(\omega - \omega_i)] \tag{3.2.27}$$

$$\text{for } b_1 \gg 1 \text{ or } P \gg 1 \text{ and } (P+1)(\omega - \omega_i) \ll 1$$

The above equation (3.2.27) is one of the fundamental results of this thesis. This equation shows

Table 3.2.1				
	real part		imaginary part	
λ	M approx. error	M+1 approx. error	M approx. error	M+1 approx. error
1.000e-02	1.030e-03	-3.640e-04	-9.930e-03	-4.021e-05
2.000e-02	4.077e-03	-1.425e-03	-1.944e-02	-3.186e-04
3.000e-02	9.016e-03	-3.090e-03	-2.814e-02	-1.059e-03
4.000e-02	1.564e-02	-5.213e-03	-3.567e-02	-2.454e-03
5.000e-02	2.369e-02	-7.596e-03	-4.169e-02	-4.657e-03
6.000e-02	3.281e-02	-1.001e-02	-4.596e-02	-7.769e-03
7.000e-02	4.265e-02	-1.220e-02	-4.830e-02	-1.183e-02
8.000e-02	5.279e-02	-1.390e-02	-4.861e-02	-1.681e-02
9.000e-02	6.282e-02	-1.485e-02	-4.685e-02	-2.263e-02

Comparison of M and $M + 1$ approximations to the $f_M(\lambda)$ function in equation (3.2.23).
Error is defined as approximation minus actual value.

clearly how each of the parameters of the problem, i.e., the model order used, the number of correlations, and signal power affect the structure of the resulting power spectral estimate.

Comparing (3.2.27) and (3.2.24) provides a very interesting and useful correspondence between the ME and OSNE spectra near where the peak is located. Note that the only difference between the two equations is that in (3.2.24) the factor $(\omega-\omega_i)$ is multiplied by $(M+1)$, while in (3.2.27), $(\omega-\omega_i)$ is multiplied by $(P+1)$. Thus one is just a frequency stretched version of the other. Specifically,

$$A_{OSNE(P,M)}(\omega_i + \Delta) = A_{ME(M)}\left(\omega_i + (P+1)\Delta/(M+1)\right) \qquad (3.2.28)$$

$$\text{for } (P+1)\Delta \ll 1 \text{ and } b_1 \gg 1 \text{ or } P \gg 1$$

An interpretation of (3.2.28) is that the OSNE spectrum for M correlations and Pth order model is equivalent *around the spectral peak*, to the spectrum resulting from ME analysis using an Mth order model, but with a frequency stretching by a factor of $(M+1)/(P+1)$.

As discussed earlier, the number of correlations used and the signal power affect the spectral shape only as a function of their product. The model order, however, controls the "frequency scaling" of the spectrum. For example, suppose the signal power and number of correlations are kept constant, while the model order used is $2P+1$, then the resulting spectrum will have the same maximum value while the spectral peak will be shrunk in the frequency axis by a factor of two. Therefore, the half-power bandwidth will be halved. The effect of varying the model order for constant M and signal power is shown in figure 3.2.1.

Thus using (3.2.28) and the discussion above, we can generalize the Lacoss results to the case of the OSNE spectrum. Recall however, that unlike Lacoss, we do not consider the noise power estimation, but rather use just $S^N(\omega)$. The spectral height then for the OSNE spectrum using M *correlations* and P *predictor* coefficients is the same as the spectral height for ME using M *predictor* coefficients, or from (2.3.7),

$$S^N_{OSNE}(\omega_0) \approx S^N_{ME}(\omega_0) \approx b^2M^2 \quad b_1 \gg 1 \text{ or } P \gg 1 \qquad (3.2.29)$$

The Lacoss result on the bandwidth of the spectral peak can also be generalized yielding from (2.3.6),

$$\omega_H \approx \frac{4}{b(P+1)(M+1)} \quad b_1 \gg 1 \text{ or } P \gg 1 \qquad (3.2.30)$$

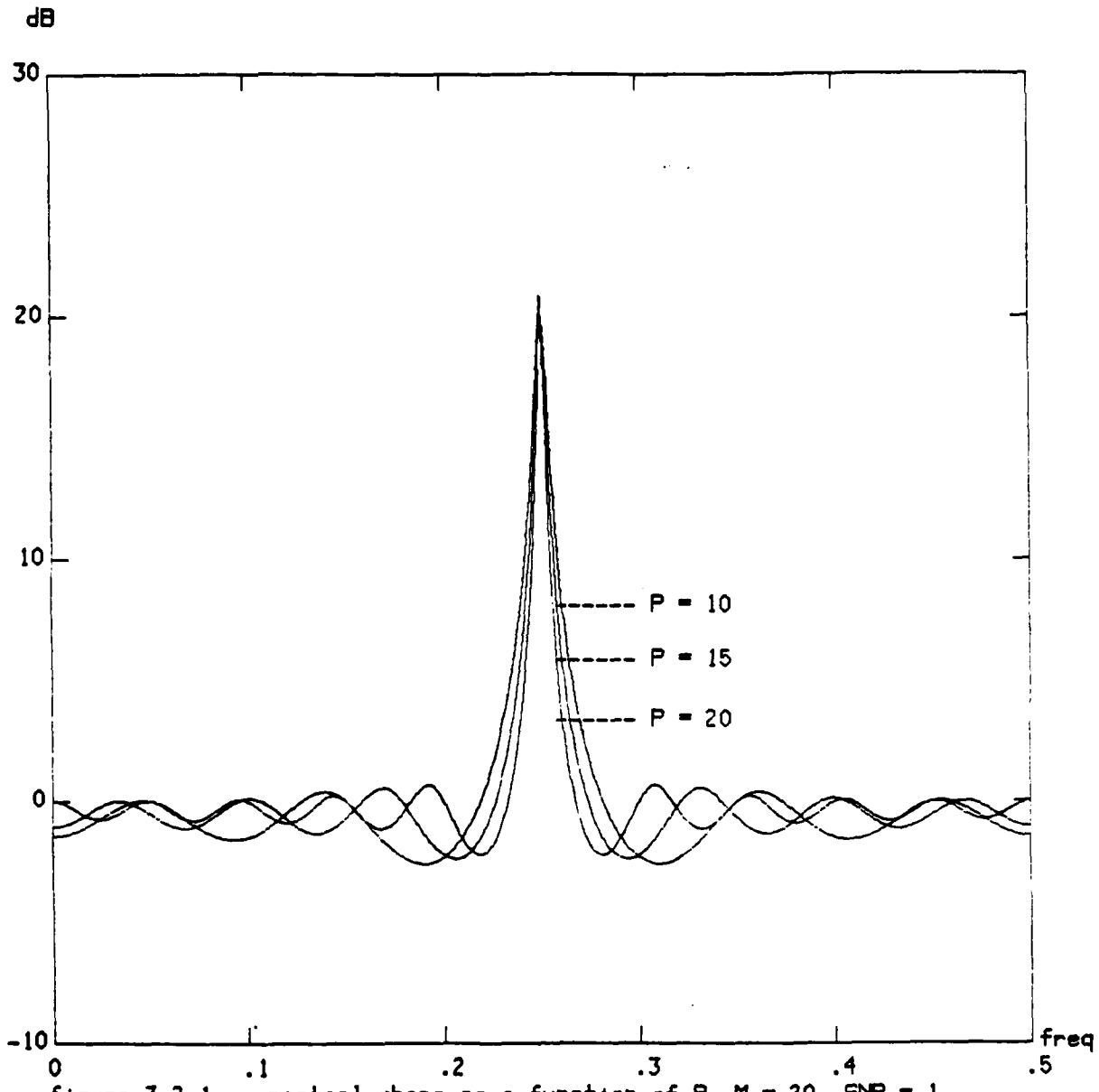


figure 3.2.1 - spectral shape as a function of P. M = 20, SNR = 1

Considering (3.2.29) and (3.2.30), we see that OSNE is a spectral density estimator since the product of spectral height and bandwidth is proportional to signal to noise ratio. However, the spectral area is no longer independent of model order or the number of correlations. There is basically, for large P, a factor of M/P which does not cancel out. Therefore to use OSNE as a power estimator the spectral area must be calculated and a proportionality constant which depends on both the model order and the number of correlations must be introduced. Specifically,

$$S^N(\omega_0) \cdot \omega_B \approx b\frac{4M^2}{(P+1)(M+1)} \tag{3.2.31}$$

This proportionality constant will have no effect if the power ratio between two sinusoids in the same spectrum is required since the constant is independent of the sinusoid characteristics. However, if an SNR value is required, this constant must be accounted for.

One final comment is to note that b is not the power in the sinusoid, but rather the power in the complex exponential. Thus for real sinusoids, (3.2.29), (3.2.30), and (3.2.31) must be altered slightly replacing b by $\frac{\alpha}{2}$ where α is the power in the real sinusoid. The above results are summarized in Table 3.2.2, where the Lacoss results for the spectral bandwidth have been converted to normalized radian frequency.

Table 3.2.2		
	ME	OSNE
Spectral peak of $SN(\omega)$	$\frac{\alpha^2 P^2}{4}$	$\frac{\alpha^2 M^2}{4}$
Spectral bandwidth in radian freq.	$\frac{8}{\alpha(P+1)^2}$	$\frac{8}{\alpha(P+1)(M+1)}$
Spectral area	$\alpha\frac{2P^2}{(P+1)^2}$	$\alpha\frac{2M^2}{(P+1)(M+1)}$

Spectral characteristics as a function of the parameters, P, the model order, M, the number of correlations, and α, the signal-to-noise ratio for a real sinusoid in noise. It is assumed that $\alpha \gg 1$, or $P \gg 1$.

CHAPTER 4

EMPIRICAL STUDIES OF OSNE AS A SPECTRAL ESTIMATOR

4.1 Introduction

The previous chapter provided some theoretical results relating the values of some algorithm parameters and the resulting behavior of the OSNE AR modeling technique when applied to spectral estimation. However, these results only apply under some conditions, namely large model orders and/or large SNR and well separated spectral peaks. The purpose of this chapter is to study what the regions of validity are in the above theoretical results.

The empirical studies will be performed on exact correlations. This requires some explanation. First, for enough data, the correlation estimates do approach the exact correlations, so a study of the behavior of OSNE under exact correlations becomes valid for the large data case. Second, the use of real data introduces some stochastic properties to the problem, i.e., the results will begin to vary statistically. This variation will cloud the underlying relationships among the algorithm parameters. Third, the use of exact correlations has provided some rather elegant relationships among the algorithm parameters, which provide at least a qualitative feel for their interaction. It is important to find where these elegant relationships are valid.

In the first set of experiments, we will consider the region of validity of the results of chapter 3 as summarized in Table 3.2.1. Actual values for peak, bandwidth and spectral area are plotted versus predicted values for a single complex exponential located at $\omega=\frac{\pi}{2}$. However, the choice of the location of the complex exponential is completely arbitrary since there are no other interfering complex exponentials. These experiments study the approximations used in deriving the basic results in (3.2.27) and (3.2.28).

The second set of experiments considers the interaction of sinusoids and therefore provides some feel for the peak coupling that occurs when the peak separation is too small. One experiment studies the interference on the estimation of the characteristics of a sinusoid by the power at negative frequencies. This interference becomes more pronounced as the location of the sinusoid approaches zero

frequency. This will show under what situations the single complex exponential results become valid. A second experiment studies the resolution capabilities of OSNE. For a given P, M and SNR the minimum spectral separation is calculated for which two peaks rather than one appear.

4.2 Verification of theoretical results

The format of the experiments in this section is now briefly described. The correlations used were exact and corresponded to a complex exponential in noise at $\frac{\pi}{2}$. The values of SNR used ranged from .125 (in absolute measurements, not dB) to 16, in multiplicative steps of 2. In the set of experiments labeled "a", the number of correlations M, was kept constant at 40. The model order was then increased from 10 to 40 in additive steps of 10. In the set of experiments labeled "b", the model order P was kept constant at 10. Then the number of correlations used was increased from 10 to 40 in steps of 10. The empirical values are denoted by circles connected by dashed lines, while the theoretical prediction are solid straight lines. Although the specific range of values used is somewhat arbitrary, they do include both a region where the theoretical results are valid and a region where the empirical calculations deviate from the predicted values.

4.2.1 Peak value vs. P, M, and SNR

In this set of plots the peak value for several combinations of M, P, and SNR were calculated. In the first plot, fig. 4.2.1a, the number of correlations was kept constant and the model order increased. The empirical results are compared with (3.2.25) which indicate that there is excellent agreement for SNR larger than 1. In figure 4.2.1b, the model order was kept constant and the number of correlations increased. Again for large SNR, the theoretical predictions and the empirical results agree. In these plots we can see clearly the "SNR increasing" property of increasing M. Note that for constant $M \cdot SNR$, the peak value remains the same. This further verifies the theoretical results.

4.2.2 Bandwidth vs. P, M, and SNR

In this section, the empirical half-power bandwidth is calculated and compared to the theoretical predictions in (3.2.26). The radian frequency bandwidth was multiplied by $1/(2\pi)$ to convert the result to normalized frequency. In figure 4.2.2a we find increasing agreement between the empirical and

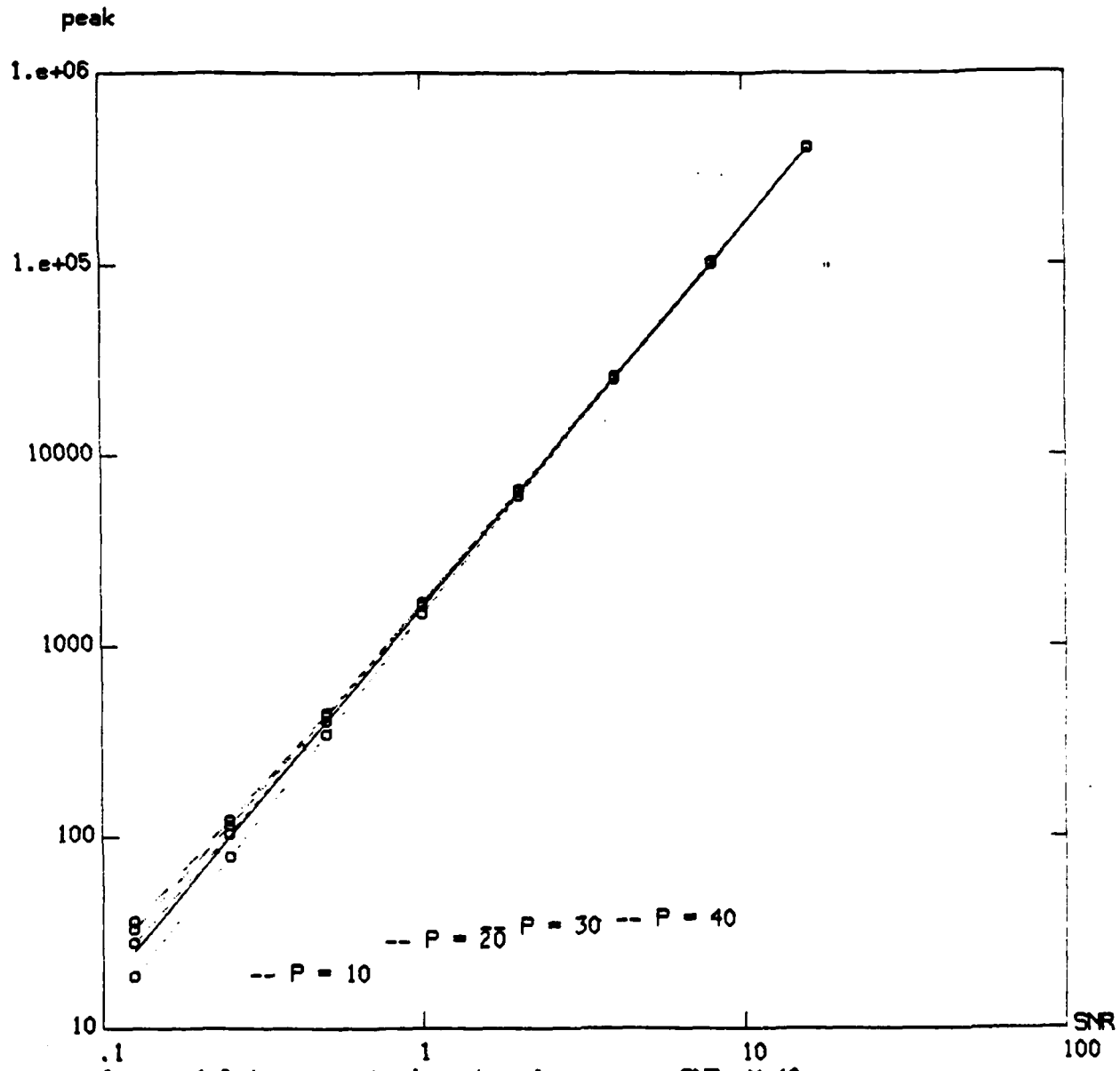


figure 4.2.1a - spectral peak value versus SNR, M=40

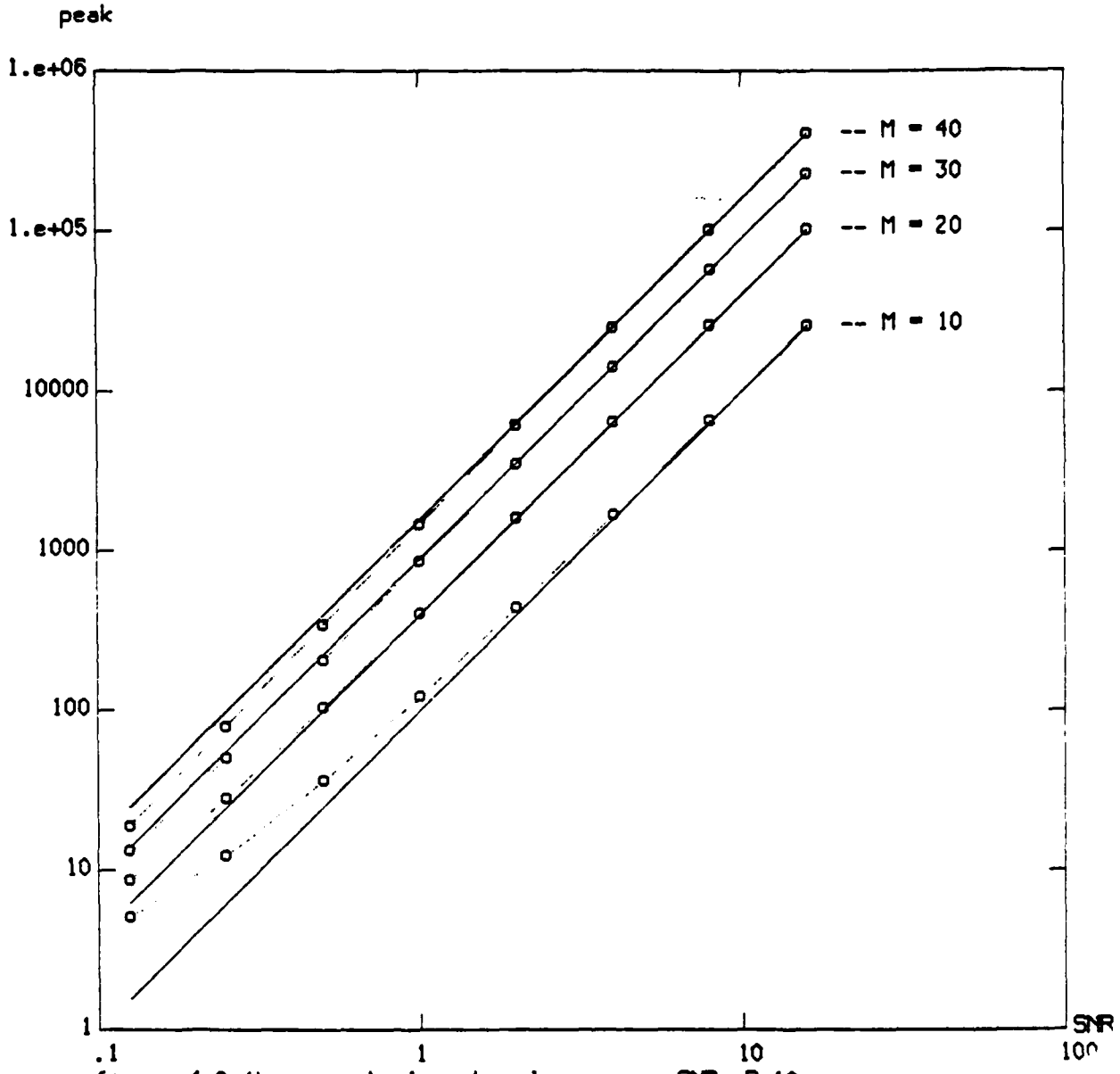


figure 4.2.1b - spectral peak value versus SNR, P=10

theoretical results as the SNR is increased (the miraculous agreement over all SNR for $P=20$ is accidental and due to the change in convexity of the empirical curve.) The second plot, figure 4.2.2b, shows the empirical calculations for the case where the model order is kept at $P=10$. As M increases, the asymptotic behavior of the spectral bandwidth approaches the predicted values. Finally, if we consider constant $M \cdot SNR$ combinations, we find that the bandwidth of the spectral peak remains constant as predicted in section 3.2.

4.2.3 Spectral Area vs. P, M, SNR

The following plots study the use of OSNE as a spectral density estimator. In figures 4.2.3a and 4.2.3b, the spectral area is plotted as a function of SNR, P, and M. Unlike the ME case, the asymptotic behavior of the spectral area depends on the values of M and P even for large P. It is not surprising that good agreement is found between empirical and theoretical results since both were calculated from the results presented in sections 4.2.1 and 4.2.2. A more interesting result is shown in figures 4.2.3c and 4.2.3d. Here each value of the spectral area was multiplied by $(P+1)(M+1)/4M^2$ to get rid of the dependence on the algorithm parameters. We note here that since we are dealing with a single /fIcomplex/fR exponential, the formulas in table 3.2.2 must be modified accordingly to yield the above proportionality constant. We see that for large SNR, multiplying the spectral area by the above constant produces a good estimate of the signal to noise ratio of the sinusoid power to the surrounding noise.

4.2.4 Interference among sinusoids vs. spectral separation

The experiments in previous sections considered the case of a single complex exponential in noise. This section studies the effect the interference of the energy at negative frequencies on the characteristics of the spectral peak of a single sinusoid in noise. A study of this phenomenon also points out the minimum separation required between two sinusoids so that the peak decoupling property becomes valid.

The OSNE method using $P=20$ and $M=40$ was applied to a signal with SNR = 4. From

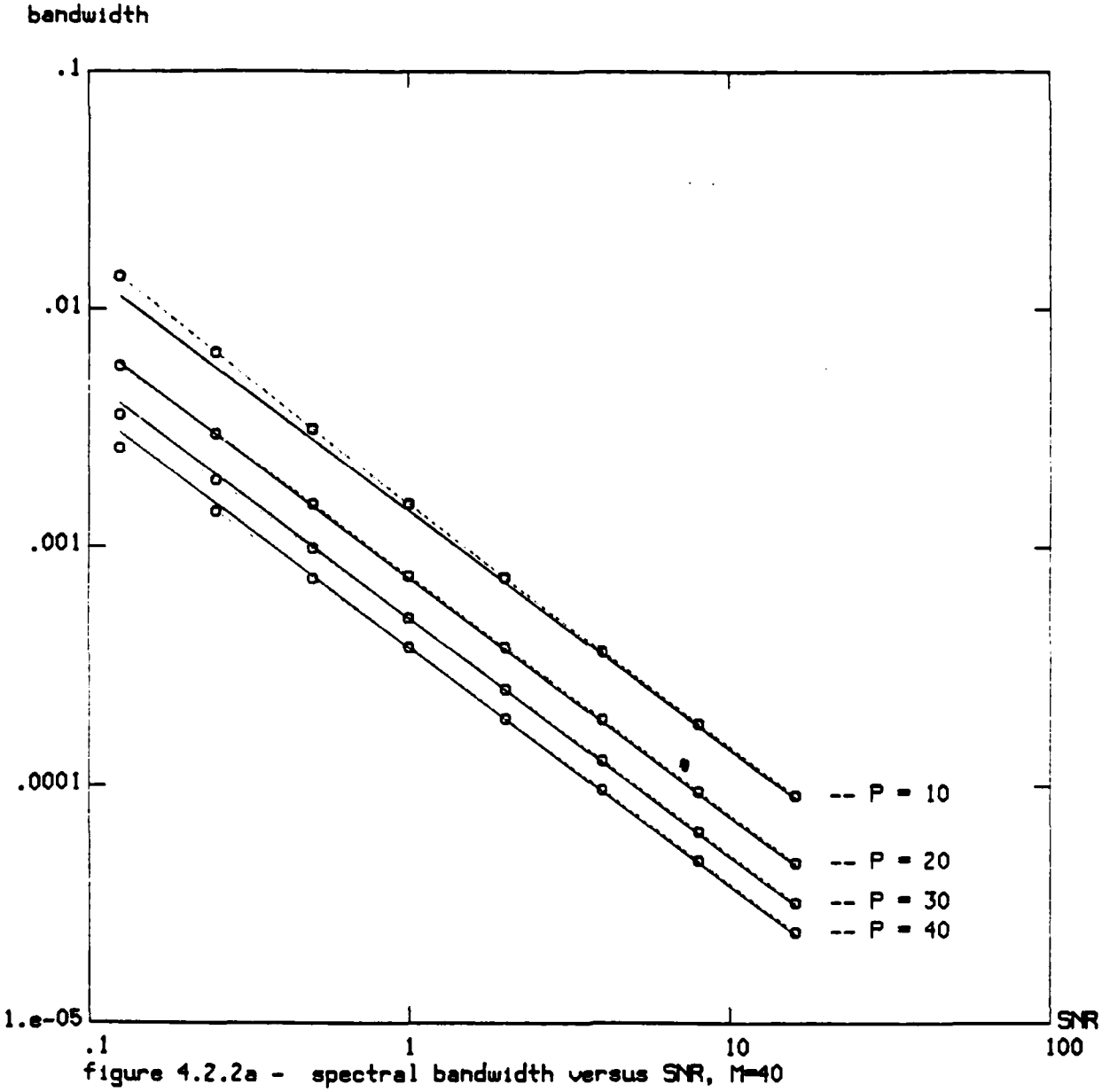


figure 4.2.2a - spectral bandwidth versus SNR, M=40

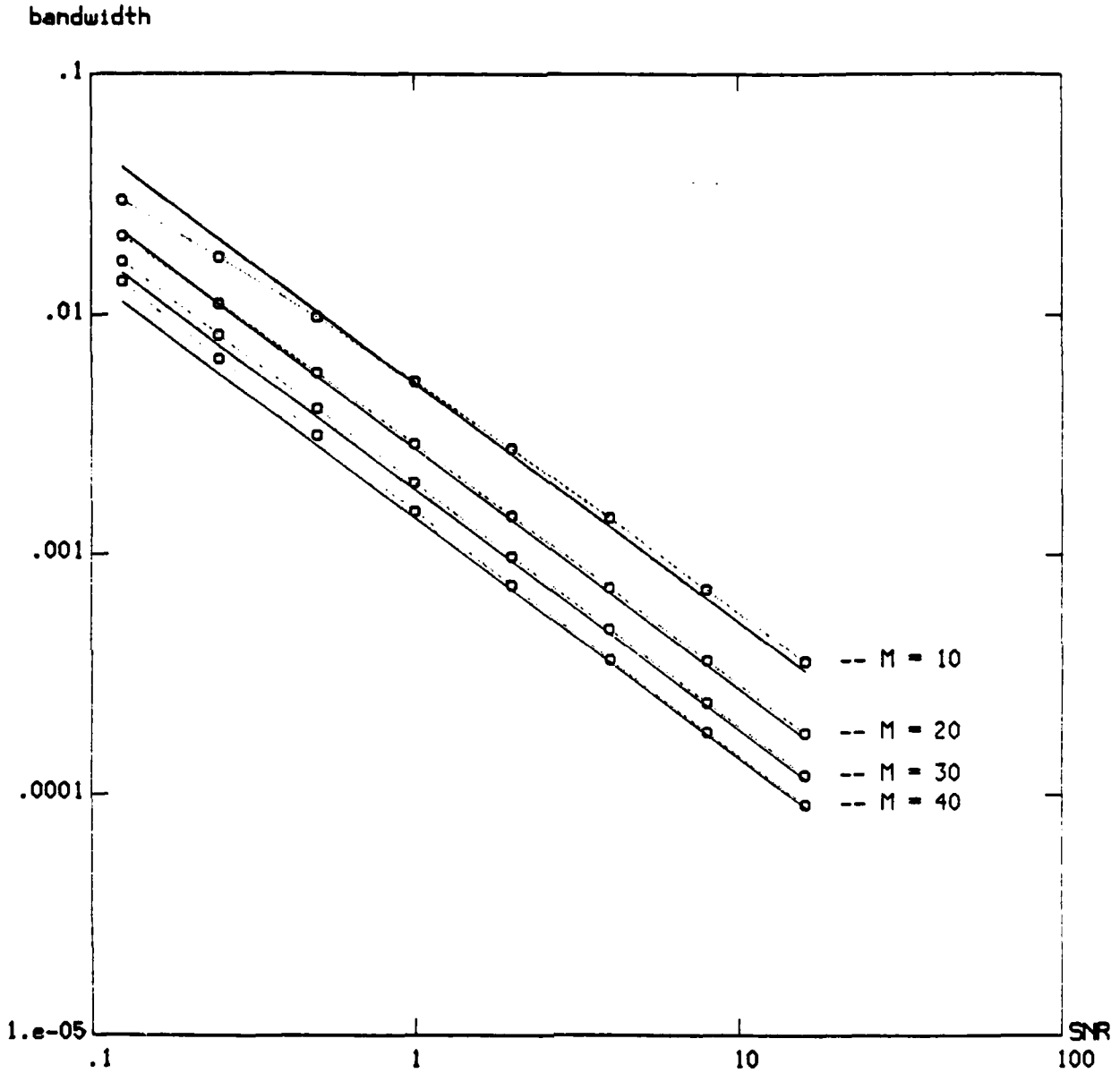


figure 4.2.2b - spectral bandwidth versus SNR, P=10

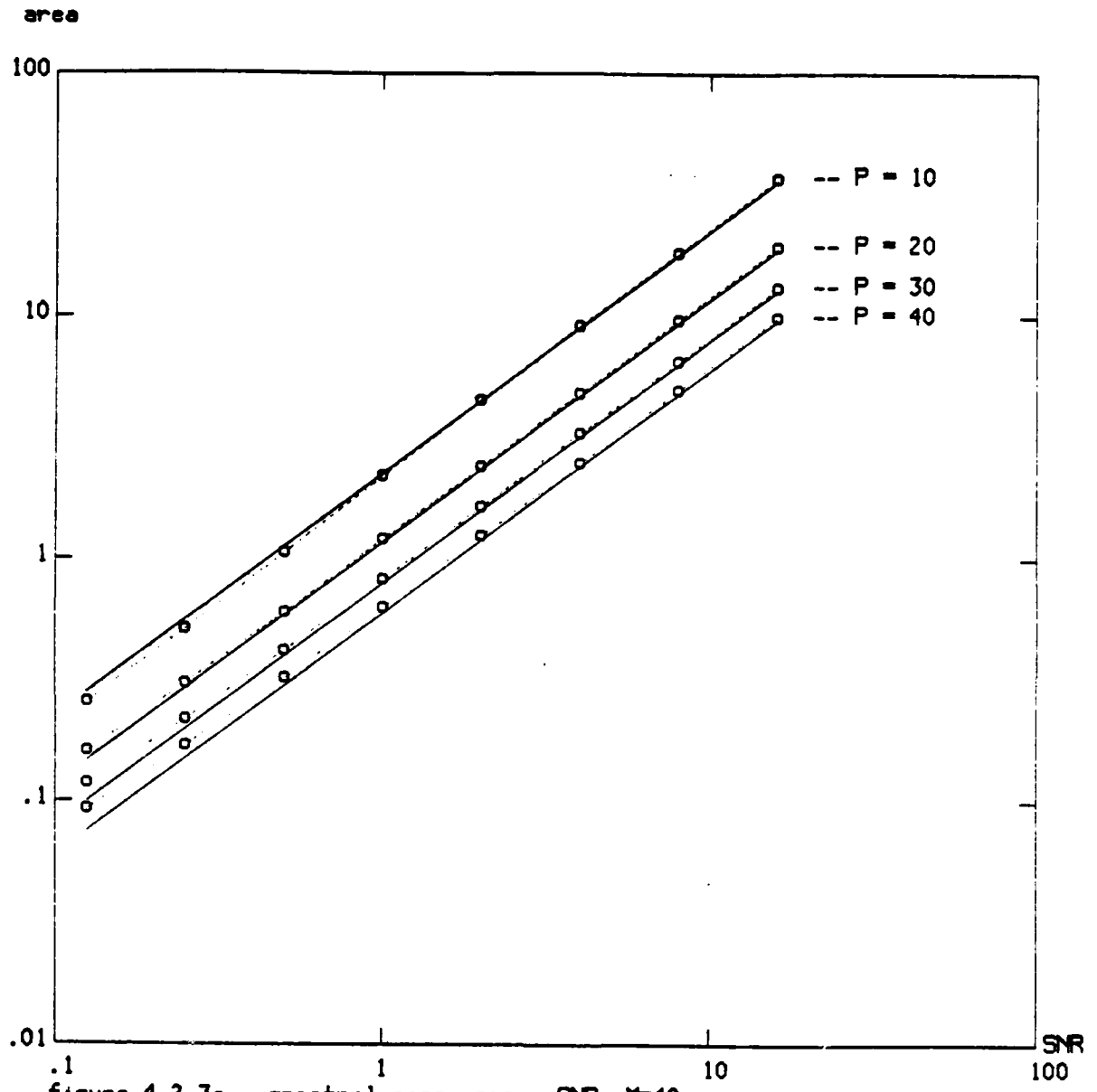


figure 4.2.3a - spectral area versus SNR, M=40

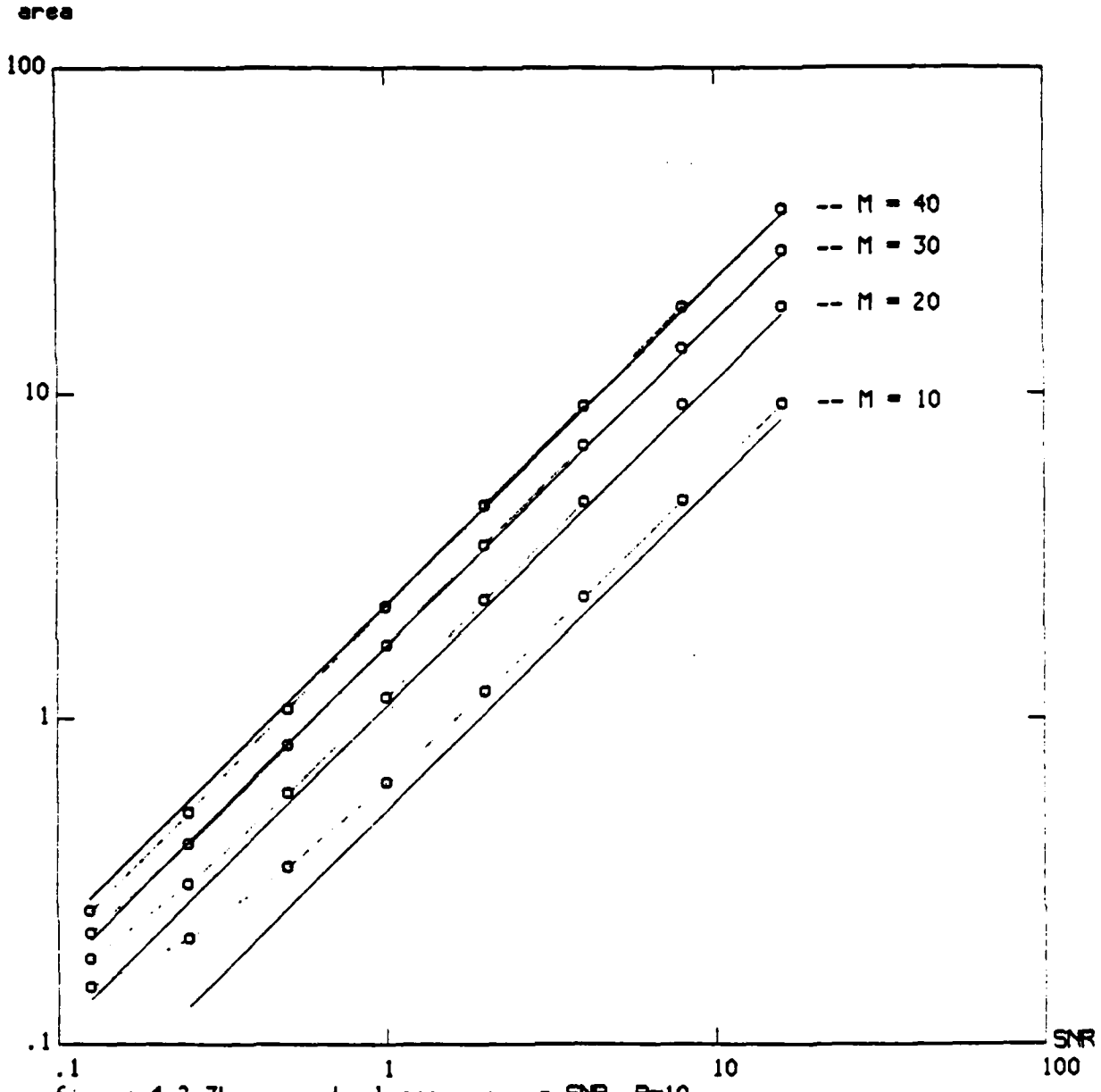


figure 4.2.3b - spectral area versus SNR, P=10

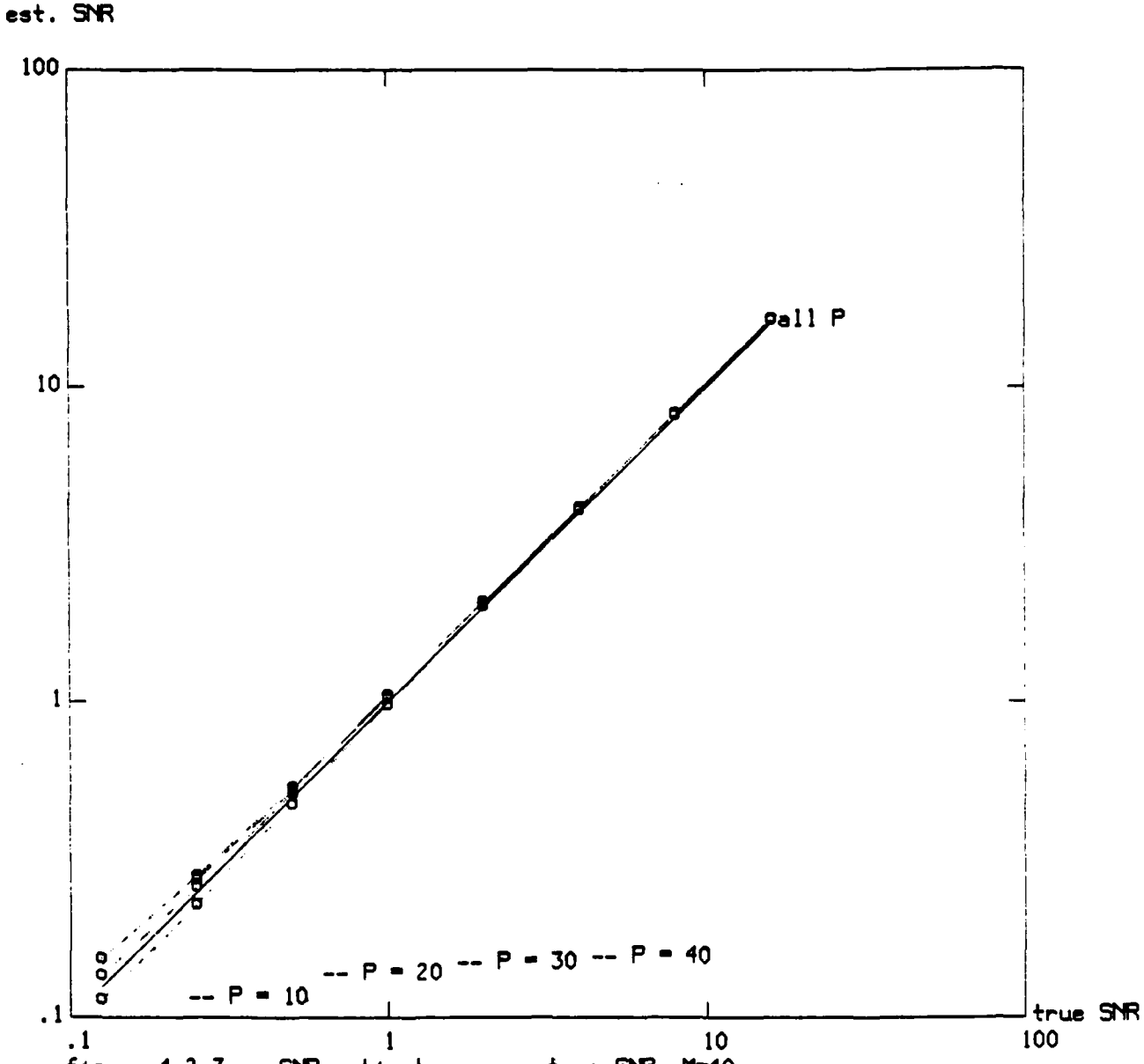


figure 4.2.3c - SNR estimate versus true SNR, M=40

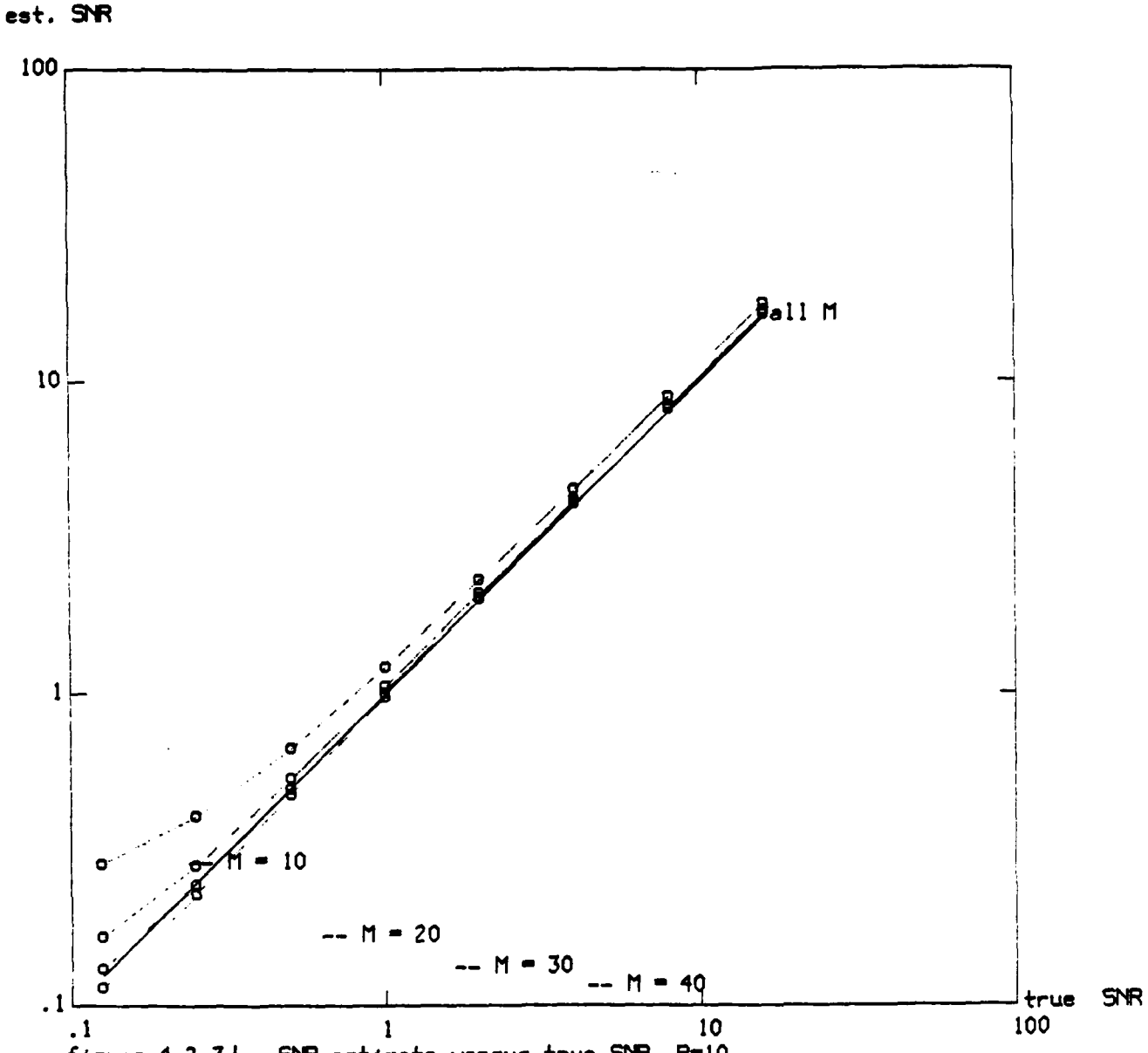


figure 4.2.3d - SNR estimate versus true SNR, P=10

previous experiments, we see that these values fall in the region of validity of the theoretical predictions. The normalized frequency, f, of the sinusoid was changed from .01 to .25 in steps of .01. The values for peak height, bandwidth, and spectral area were converted to a dB scale and normalized to the value $f=.25$. The results of the experiments are shown in figures 4.2.4a, 4.2.4b, and 4.2.4c. From 4.2.4c we see that the spectral area does not vary by more than .5 dB until a value of f of about .05. That means that a minimum separation of .1 is needed for independent estimation of power for the above choice of P, M, and SNR.

For comparison, the same experiment was conducted using $P=20$, $M=20$, i.e., the ME method, figures 4.2.4d, 4.2.4e, 4.2.4f. In this case, although the peak value and bandwidth vary tremendously, the spectral area remains very stable, varying by about .01 dB down to f=.04. We can conclude from this that ME is a more "stable" power estimator; that is, it is much less sensitive to interference from other sinusoids. As a final comparison, the above experiment was repeated for $P=40$, $M=40$, figures 4.2.4g, 4.2.4h, 4.2.4i. Here the spectral area estimate is even more stable, further supporting the assertion that ME is a "stable" power estimator.

4.2.5 Resolution of two sinusoids vs. P, M, SNR

The final experiment performed to study the behavior of OSNE was to gain some idea of the resolution capabilities of the method. It was found that an approximate "rule of thumb" could be extracted which described the relationship between P, M, and SNR and the resolving frequency. The experimental data is plotted in figures 4.2.5a to figure 4.2.5d along with predictions based on this approximate formula. Before discussing these results however, some definition of resolution must be presented. The definition used here is in terms of the convexity of the spectrum midway between the location of two equal amplitude complex exponentials. When the separation between the two exponentials is such that the convexity of the spectrum at the midpoint is about to change sign then the spectrum is about to change from displaying one to two peaks. In this experiment, a real sinusoid near zero was used and the problem of resolution was rephrased as whether the resulting spectrum showed a peak at zero, implying that the spectral peaks at positive and negative frequencies were not

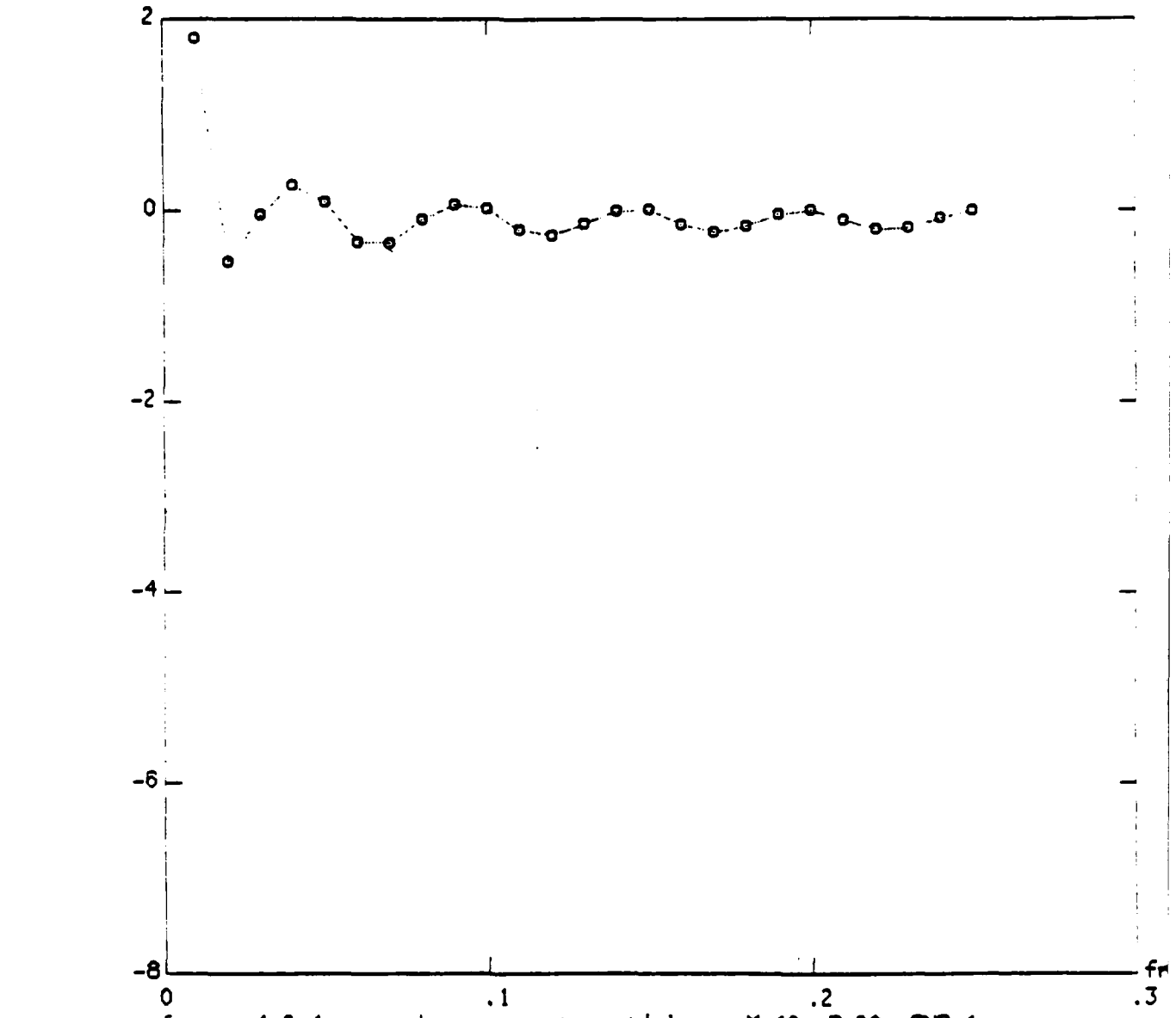


figure 4.2.4a - peak versus sinusoid loc., M=40, P=20, SNR=4.

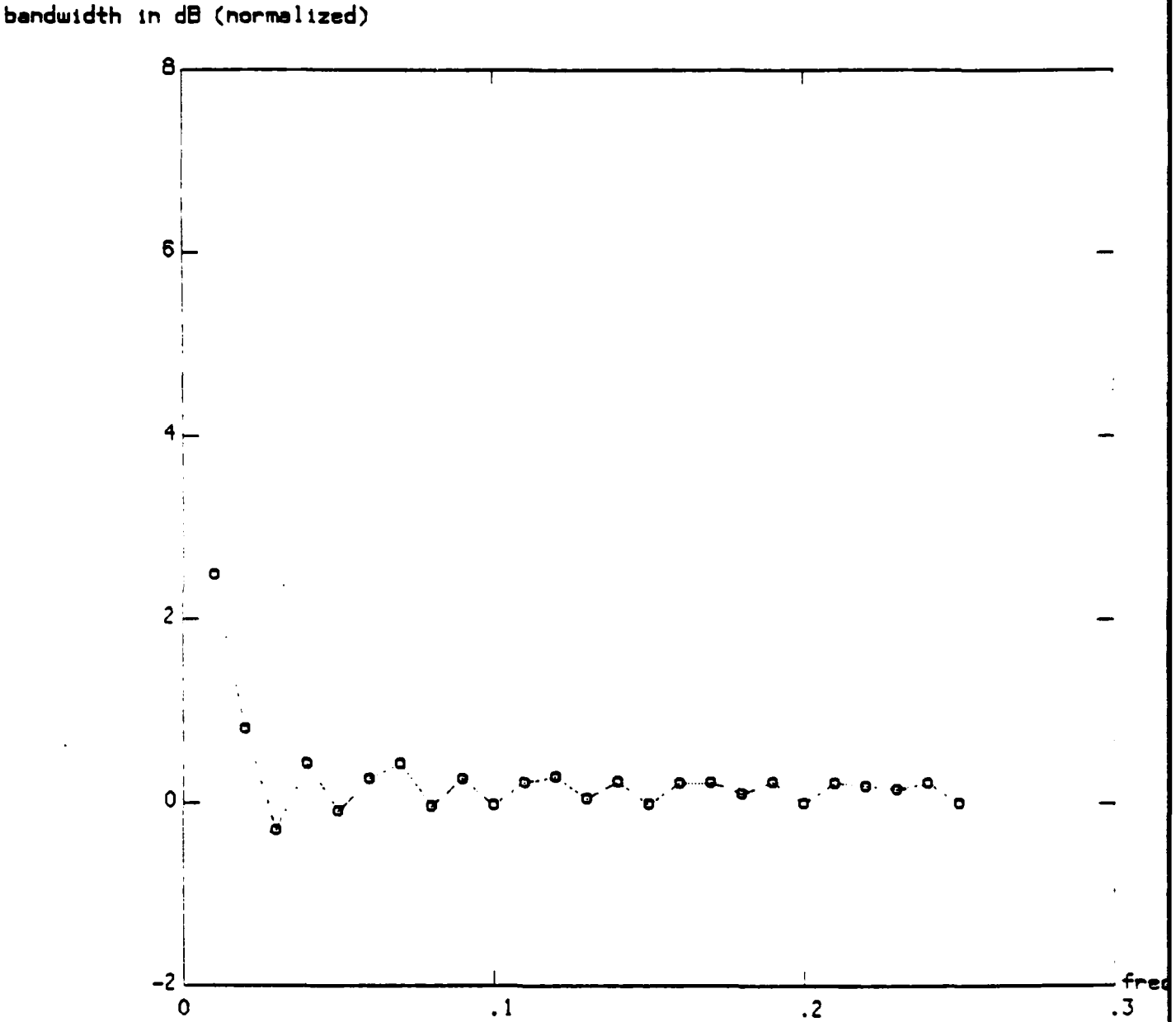


figure 4.2.4b - bandwidth versus sinusoid loc., M=40, P=20, SNR=4.

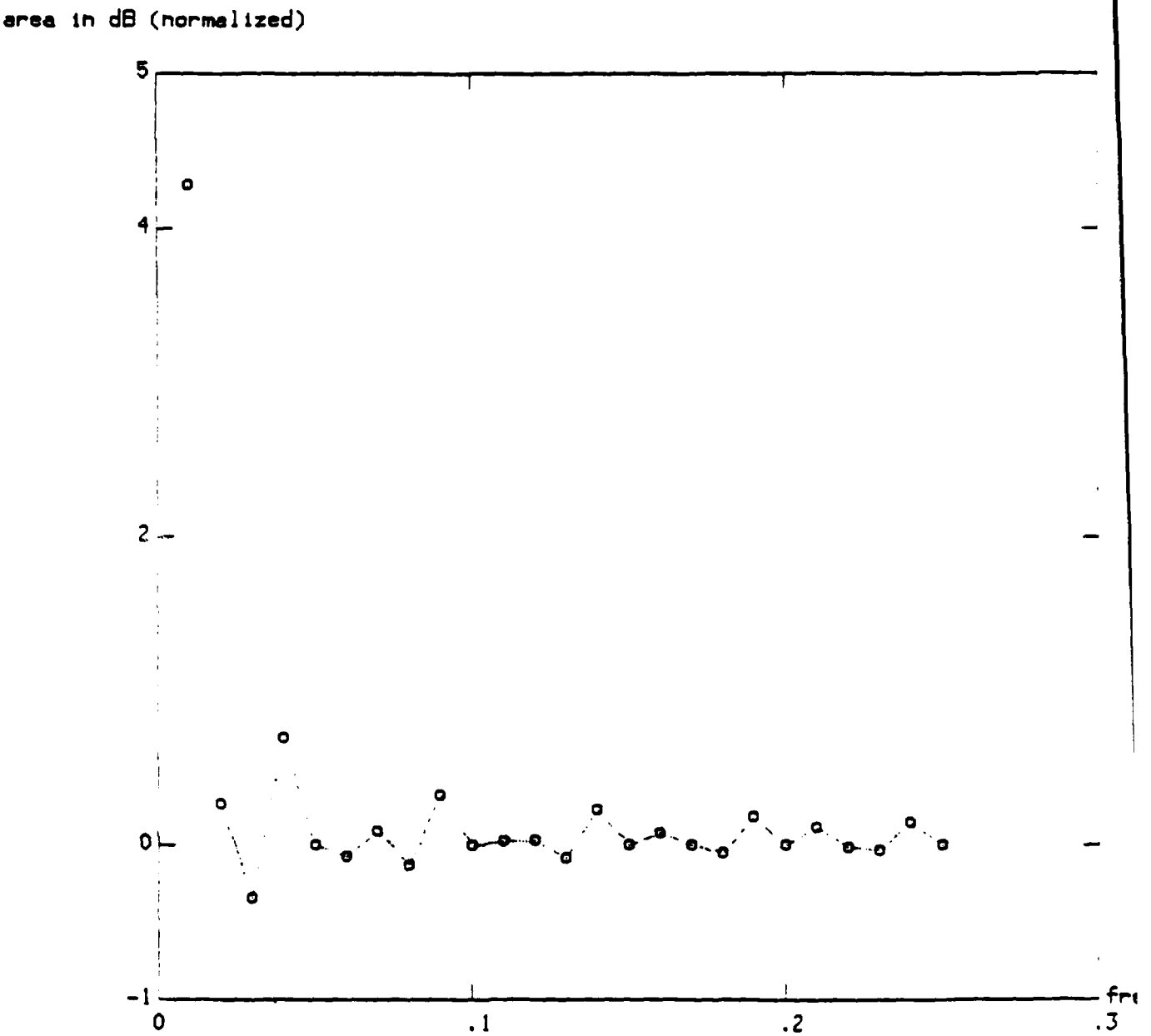


figure 4.2.4c - area versus sinusoid loc., M=40, P=20, SNR=4.

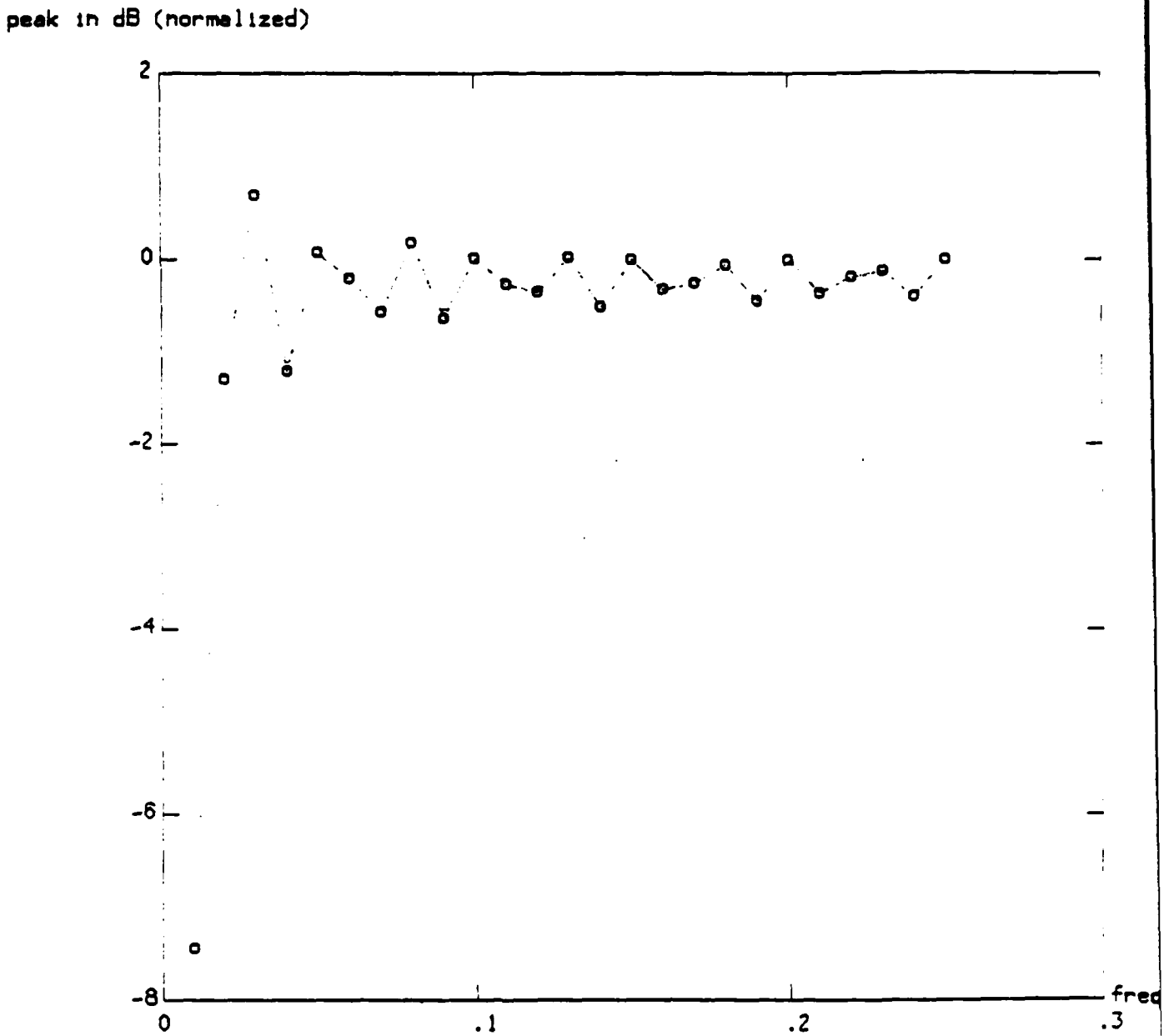


figure 4.2.4d - peak versus sinusoid loc., M=20, P=20, SNR=4

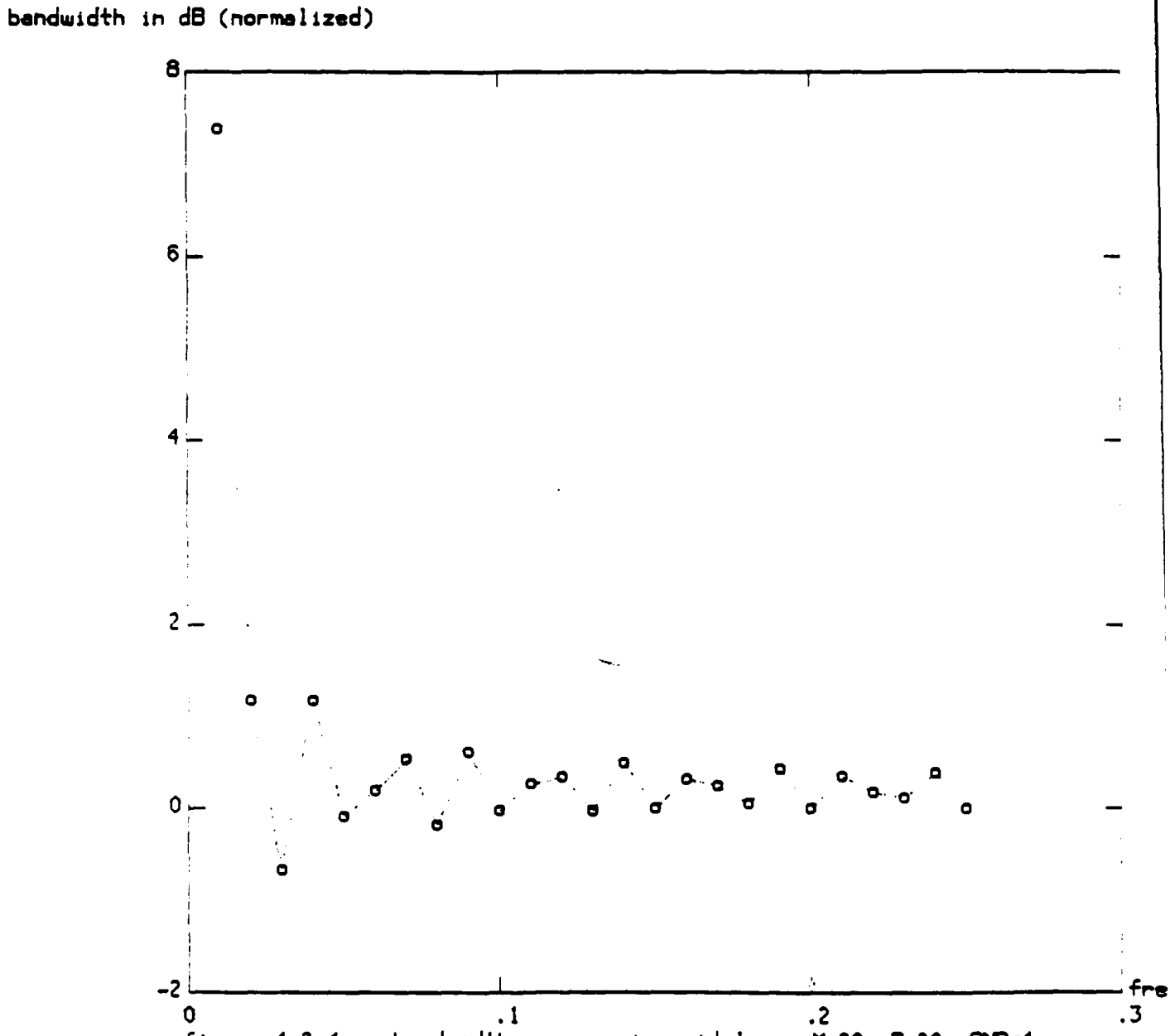


figure 4.2.4e - bandwidth versus sinusoid loc., M=20, P=20, SNR=4

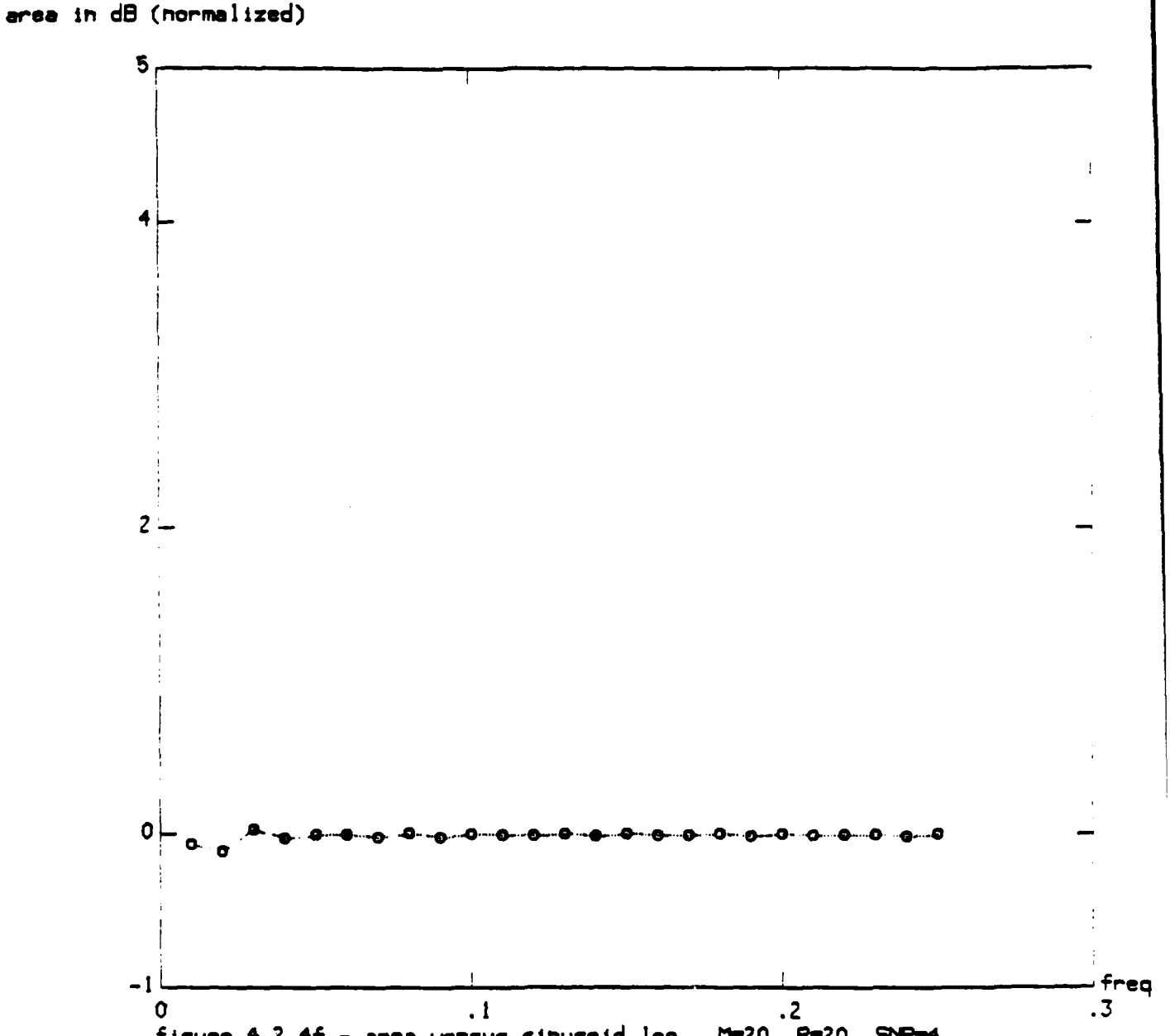


figure 4.2.4f - area versus sinusoid loc., M=20, P=20, SNR=4.

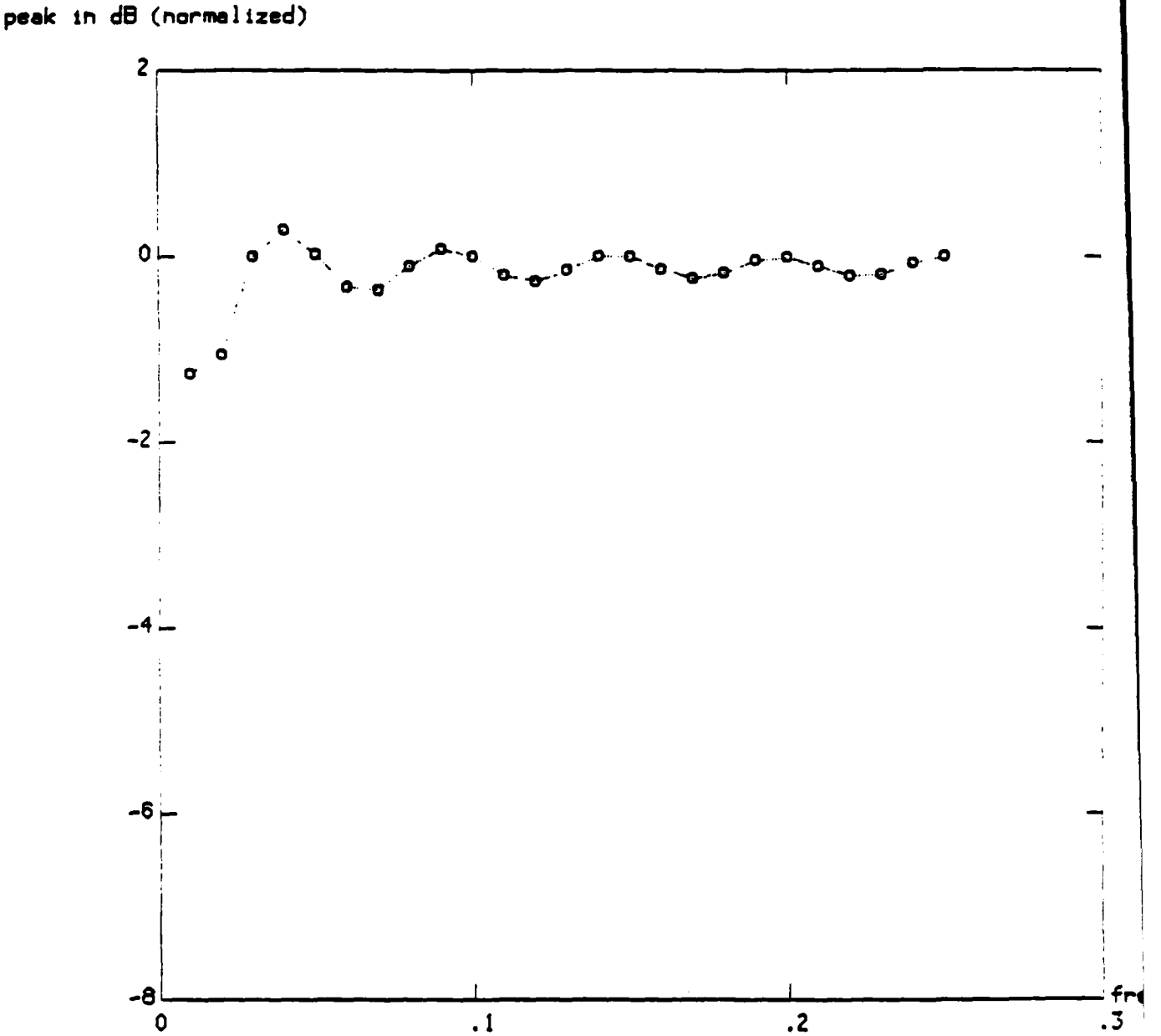


figure 4.2.4g - peak versus sinusoid loc., M=40, P=40, SNR=4.

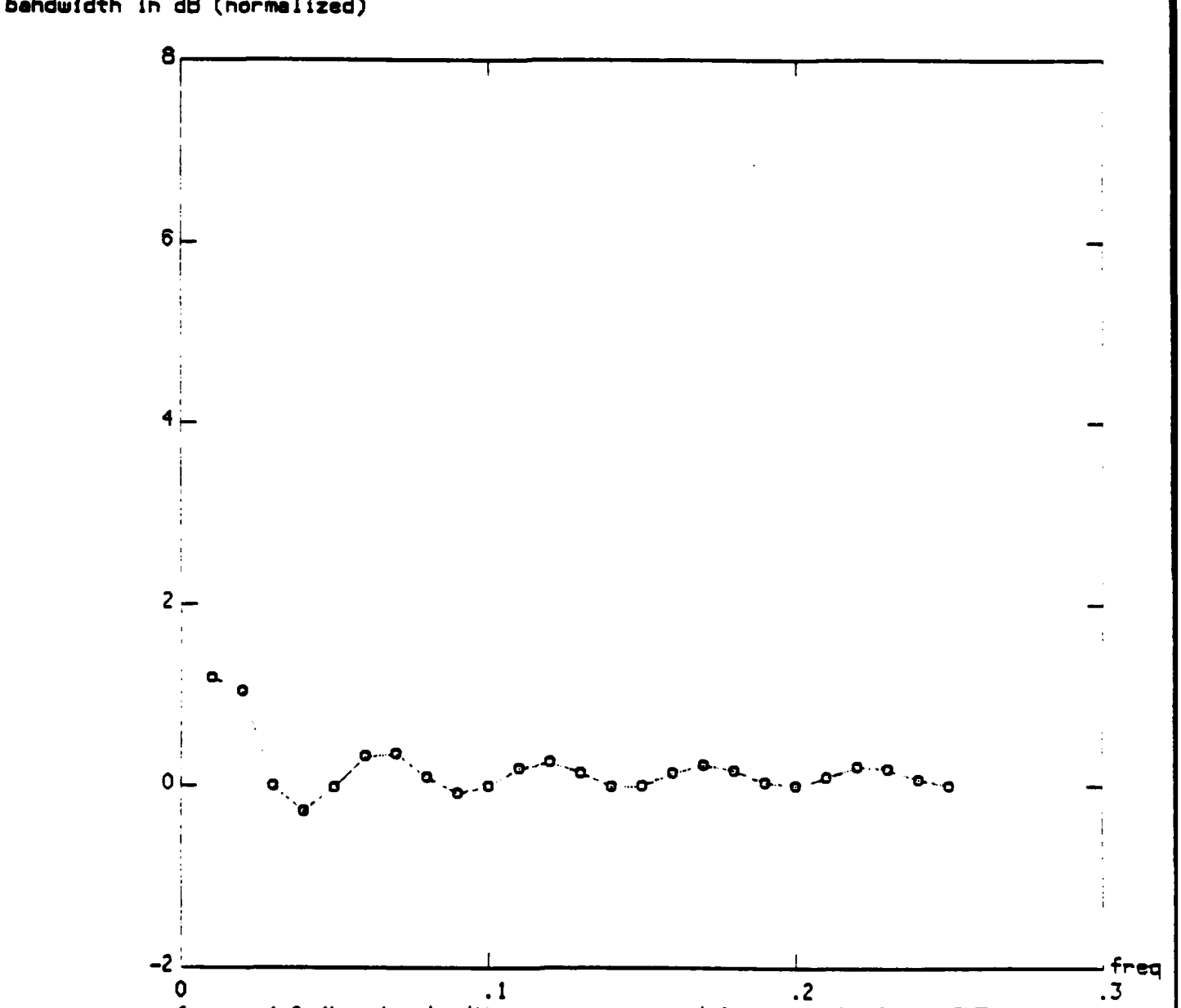


figure 4.2.4h - bandwidth versus sinusoid loc., M=40, P=40, SNR=4.

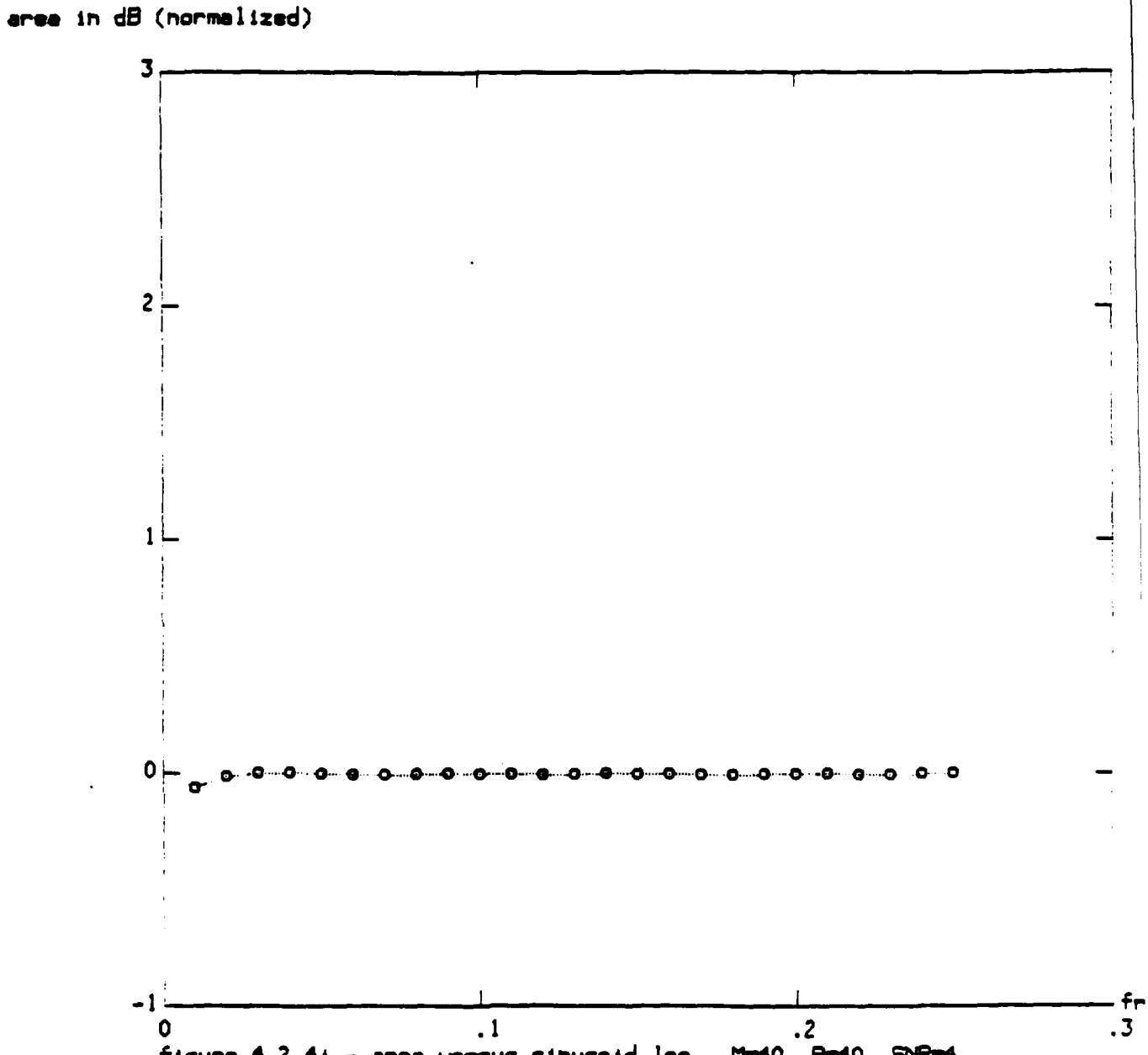


figure 4.2.4i - area versus sinusoid loc., M=40, P=40, SNR=4.

separated, or whether the peak occurred at a positive frequency, so that the two peaks at positive and negative frequencies appeared.

Using the expression below for the predictor Fourier transform,

$$A(\omega) = 1 + \sum_{k=1}^{P} e^{-j\omega k} a_k \tag{4.2.5.1}$$

the second derivative of $|A(\omega)|$ is easily calculated as

$$\left.\frac{d^2|A(\omega)|^2}{d\omega^2}\right|_{\omega=0} = 2[(S_2)^2 - S_1 S_3] \tag{4.2.5.2}$$

where

$$\begin{aligned} S_1 &= 1 + \sum_{k=1}^{P} a_k \\ S_2 &= \sum_{k=1}^{P} k\, a_k \\ S_3 &= \sum_{k=1}^{P} k^2 a_k \end{aligned} \tag{4.2.5.3}$$

It is readily seen that $\frac{d^2|S^N(\omega)|^2}{d\omega^2}$ is monotonic with $-\frac{d^2|A(\omega)|^2}{d\omega^2}$ and is equal to zero at the same values of ω. Thus we can use (4.2.5.2) and (4.2.5.3) to decide whether a sinusoid located at ω_0 will be resolved for a given value of P, M, and SNR. The procedure is to see if (4.2.5.2) is less than or greater than zero. If greater than zero, then the sinusoid is certainly resolved; if less than zero, there is the possibility that the sinusoid is still resolved if ω_0 is much greater than zero. What could occur is that the oscillations of the spectrum about 0 dB as seen in figure 3.2.1 may yield a negative convexity at zero. However, if we are careful not to wander too far from ω_0=0, checking if (4.2.5.2) is less than or greater than zero will provide an accurate test of resolvability.

For combinations of P, M, and SNR, the sinusoid location was varied until the the value of (4.2.5.2) equaled 0. In figure 4.2.5a, M was kept constant at 40, and P and the SNR varied. In figure 4.2.5b, P was now kept constant at 20 and M and the SNR varied. From the linear characteristic of the resolution as a function of SNR we can deduce that the resolving frequency, for a constant P and M varies as a function of some power of the SNR times a constant which depends on P and M only. In figures 4.2.5c and 4.2.5d, the SNR was kept constant at 16. and the resolving frequency plot-

ted as a function of M and P. In all the plots, we see approximate linear relationships (on a log scale) between M, P, SNR, and the resolving frequency as the SNR, P, and M increase. Such a behavior implies that the relation between the resolving frequency and the above parameters is given by,

$$f_{resol} \approx kM^{\alpha_M} P^{\alpha_P} SNR^{\alpha_{SNR}} \qquad (4.2.5.4)$$

From the experimental data we can extract the above parameters (the data points used were the combinations (P, M, SNR): (40, 40, 16), (40, 40, 8), (30, 40, 16), (30, 30, 16))

$$k = .47$$
$$\alpha_P = -.67$$
$$\alpha_M = -.63$$
$$\alpha_{SNR} = -.32$$

Using the above model the predicted results of the experiments are plotted in the solid lines in figures 4.2.5a, 4.2.5b, 4.2.5c, 4.2.5d. The reader should not be surprised at the agreement between the predicted values and the experimental values for the combinations described in the previous paragraph; after all, it was using that data that the model was derived. However, there is some widespread agreement in other regions of the figures. These examples seem to point to the structure in (4.2.5.4) as being of the right form. However, no more decisive comments than that will be made. A theoretical analysis of the resolution properties of OSNE does not seem intractable using the techniques described in this thesis and may provide some verification of the empirical results presented here.

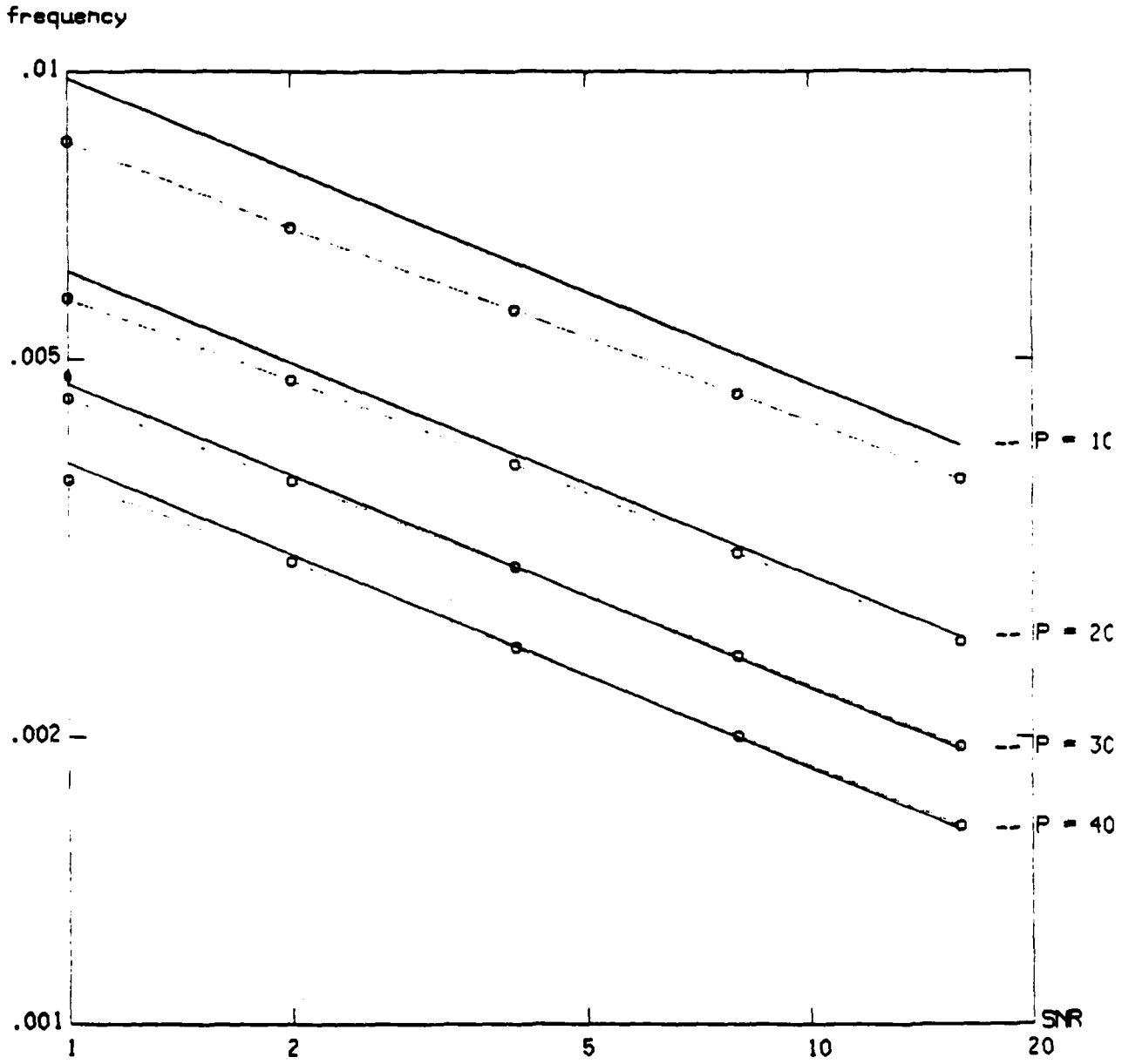


figure 4.2.5a - resolving frequency versus SNR, M=40

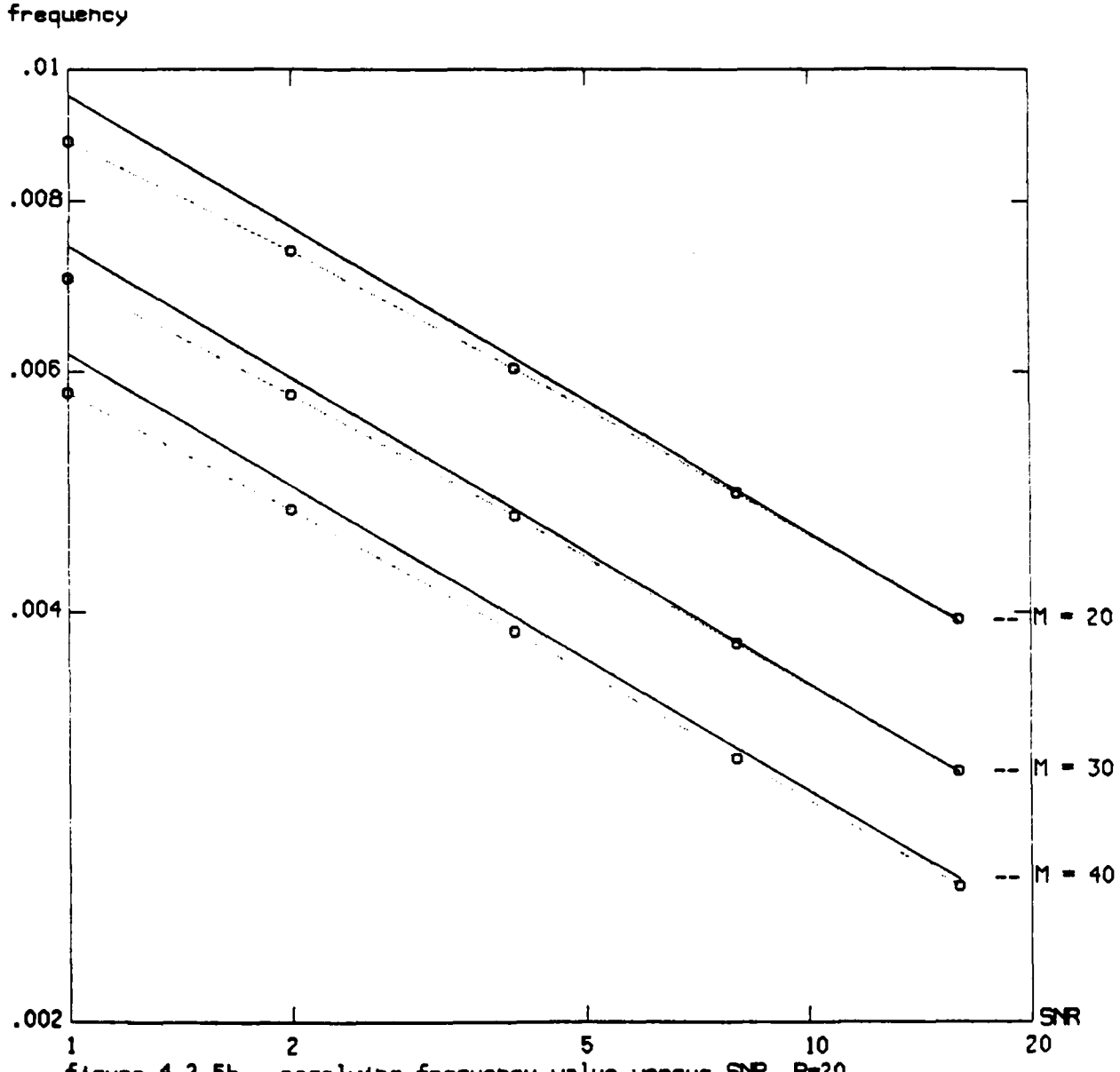


figure 4.2.5b - resolving frequency value versus SNR, P=20

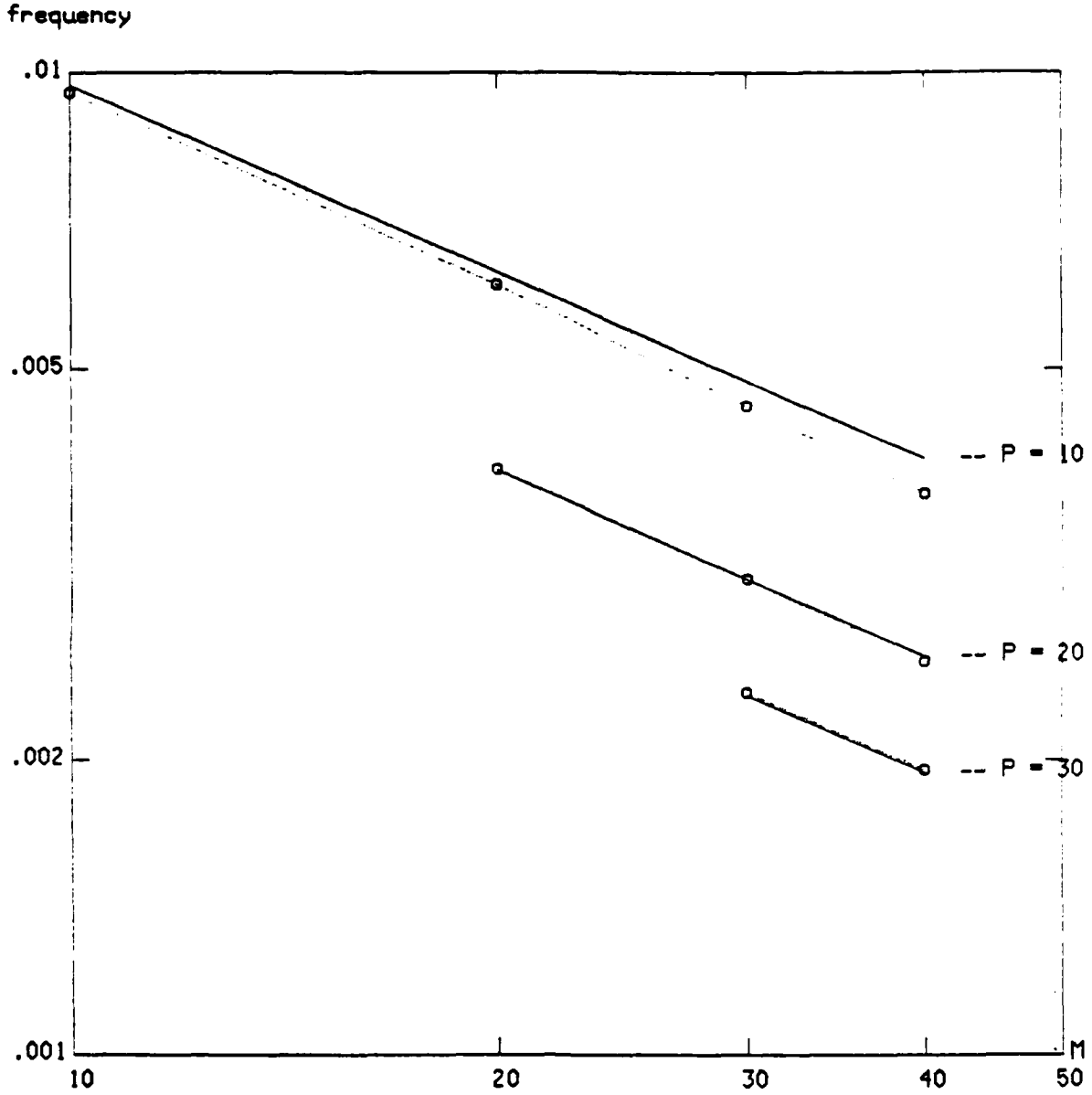


figure 4.2.5c - resolving frequency versus number of correlations, SNR = 16

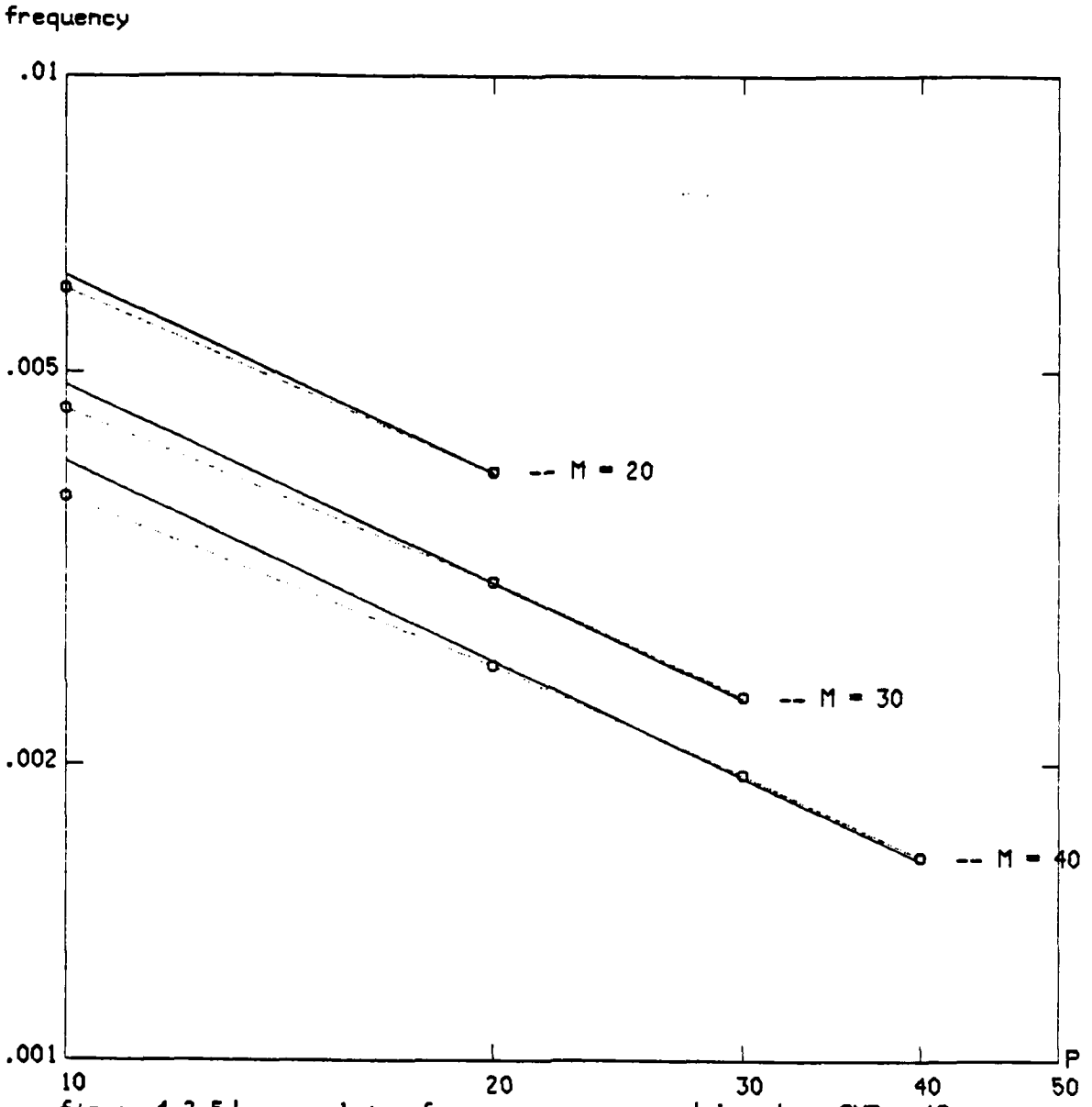


figure 4.2.5d - resolving frequency versus model order, SNR = 16

CHAPTER 5

SUMMARY AND CONCLUSIONS

In this thesis a study is presented of some of the properties of the method of overspecified normal equations as applied to the problem of spectral estimation. The main contribution of this thesis is the derivation of the relationships between the number of correlations used, the model order and the signal to noise ratio of the signal, to the characteristics of the resulting spectral estimate in terms of the spectral height, bandwidth and area. The method is shown to be a spectral density estimator like the ME method, where spectral *areas* rather than spectral values should be interpreted as estimates of power.

The relationships derived point to the number of correlations used over the minimum, i.e. model order, as a signal-to-noise enhancer, where the resulting spectrum is equivalent to the ME spectrum under higher signal-to-noise conditions. Another result is the requirement of a proportionality constant dependent on the number of correlations and the model order which is required for accurate signal-to-noise measurements. This constant is not required however, for measurements of relative power within the same spectral estimate, as in the power ratio of two sinusoids in noise.

The second part of the thesis presents some empirical studies using computer simulations which verify the theoretical predictions and provide the region of validity of the analysis. Further experiments study the interfering effect of several closely spaced sinusoids. The method of overspecified normal equations is shown to be much more sensitive to this interference than the ME method. Finally, some purely empirical studies are made of the resolution capabilities of the method. Using the data derived, an empirical model is derived which seems to agree to some extent with the data.

REFERENCES

[1] J. P. Burg, "Maximum entropy spectral analysis," *Proc. 37th Meeting Society of Exploration Geophysicists* (Oklahoma City, OK), 1967.

[2] J. A. Cadzow, "High Performance Spectral Estimation - A New ARMA Method," *IEEE ASSP-28*, pp. 524 - 529, 1980.

[3] J. A. Cadzow, "Spectral Estimation: An Overdetermined Rational Model Equation Approach," *Proc. IEEE*, vol. 70, pp. 907 - 938, 1982.

[4] J. Capon, "High Resolution Frequency-Wavenumber Spectrum Analysis," *Proc. IEEE*, vol. 57, pp. 1408 - 1418, 1969.

[5] D. E. Dudgeon, "Fundamentals of Digital Array Processing," *Proc. IEEE*, vol. 65, pp. 898 - 904, 1977.

[6] J. N. Franklin, *Matrix Theory*, Englewood Cliffs, NJ: Prentice - Hall, 1968. See pp 12 - 13.

[7] M. Kaveh and S. P. Bruzzone, "A Comparative Overview of ARMA Spectral Estimation," *First ASSP Workshop on Spectral Estimation*, Hamilton, Ontario, Canada, 1981.

[8] R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," *Geophysics*, vol. 36, pp. 661 - 675, 1971.

[9] S. W. Lang, "Performance of Maximum Entropy Spectral Estimators," S.M. thesis, MIT, Cambridge, MA, 1979.

[10] J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, vol. 63, pp. 561 - 580, 1975.

[11] J. D. Markel and A. H. Gray, Jr, *Linear Prediction of Speech*, New York: Springer - Verlag, 1976.

[12] E. A. Robinson, "Spectral Approach to Geophysical Inversion by Lorentz, Fourier, and Radon Transforms," *Proc. IEEE*, vol. 70, pp. 1039 - 1054, 1982.

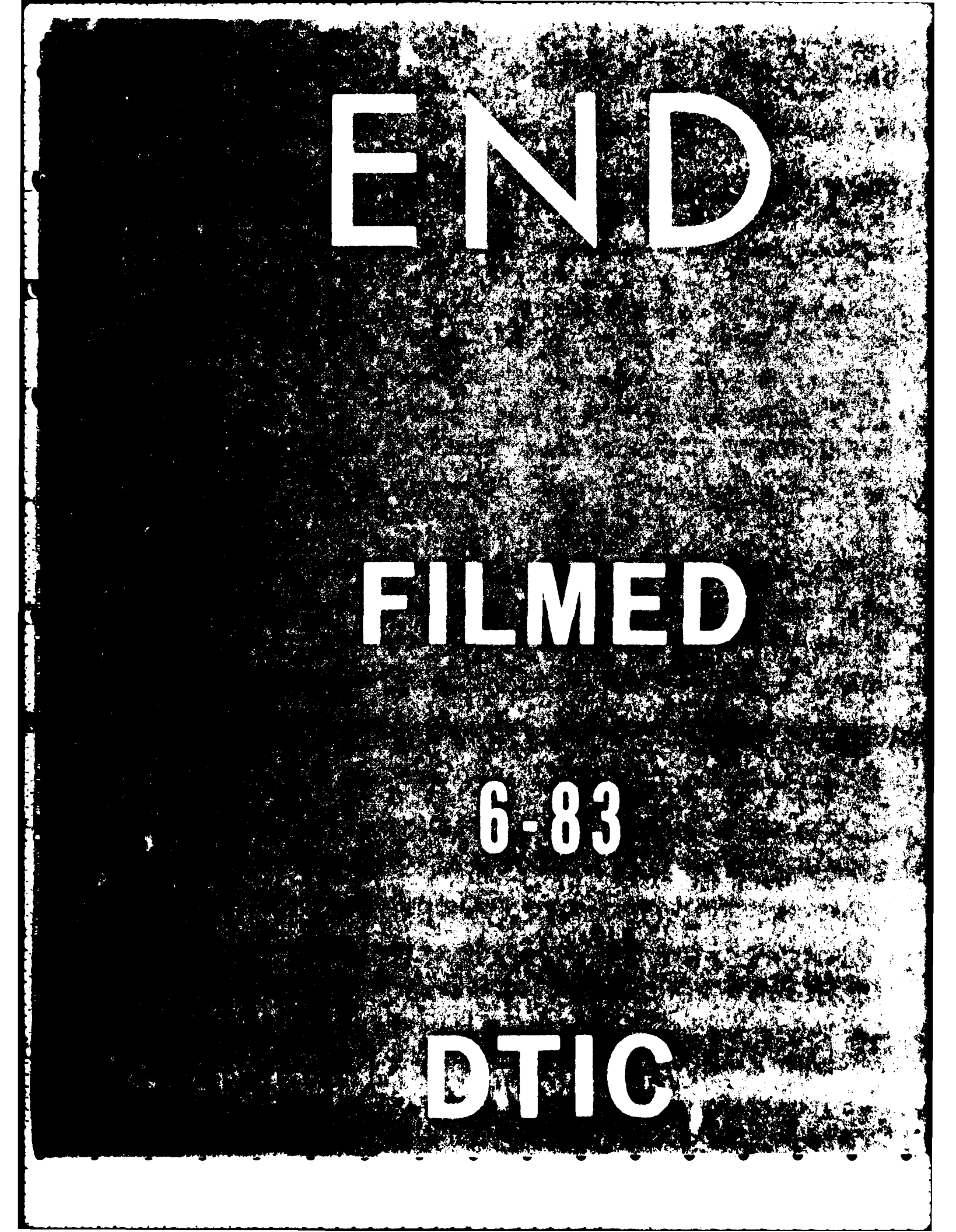
[13] E. H. Satorius and J. R. Zeidler, "Maximum Entropy Spectral Analysis of Multiple Sinusoids in Noise," *Geophysics*, vol. 43, pp. 1111 - 1118, 1978.

[14] T. J. Ulrych and R. W. Clayton, "Time Series Modelling and Maximum Entropy", *Phys. of the Earth and Planet. Interiors*, pp. 188 - 200, 1976.

[15] P. D. Welch, "The Use of the Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time-Averaging over Short Modified Periodograms," *IEEE AU-15*, pp. 70 - 73, 1967.

DISTRIBUTION LIST

	DODAAD Code	
Director Advanced Research Project Agency 1400 Wilson Boulevard Arlington, Virginia 22209 Attn: Program Management	HX1241	(1)
Group Leader Information Sciences Associate Director for Engineering Sciences Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217	N00014	(1)
Administrative Contracting Officer E19-628 Massachusetts Institute of Technology Cambridge, Massachusetts 02139	N66017	(1)
Director Naval Research Laboratory Attn: Code 2627 Washington, D.C. 20375	N00173	(6)
Defense Technical Information Center Bldg. 5, Cameron Station Alexandria, Virginia 22314	S47031	(12)


END
FILMED
6-83
DTIC