Spectral Estimation Based on Singular Value Decomposition Techniques

by

Ismail Mahmoud Al-Farra

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

June, 1990

Order Number 1355716

Spectral estimation based on singular value decomposition techniques

Al-Farra, Ismail Mahmoud, M.S.

King Fahd University of Petroleum and Minerals (Saudi Arabia), 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

SPECTRAL ESTIMATION BASED ON SINGULAR VALUE DECOMPOSITION TECHNIQUES

BY

ISMAIL MAHMOUD AL-FARRA

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

JUNE 1990

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS
DHAHRAN, SAUDI ARABIA

COLLEGE OF GRADUATE STUDIES

This thesis, written by ***ISMAIL MAHMOUD AL-FARRA*** under the direction of his Thesis Advisor and approved by his Thesis Committee, has been presented to and accepted by the Dean of the College of Graduate Studies, in partial fulfillment of the requirements for the degree of ***MASTER OF SCIENCE*** in ***ELECTRICAL ENGINEERING.***

THESIS COMMITEE

DR. M. BETTAYEB (CHAIRMAN)

DR. U. AL-SAGGAF (MEMBER)

DR. M. DAWOUD (MEMBER)

Department Chairman

Dean, College of Graduate Studies

TO MY BELOVED PARENTS, BROTHERS, SISTERS;
THOSE WHO SHARED THEIR CARE AND CONCERN
AND THE MUJAHIDEEN SPECIALLY THE HEROS
OF INTIFADDAH IN PALESTINE

ACKNOWLEDGEMENT

First of all, I thank " *ALLAH* " for all the knowledge and science *HE* has given to us. Also, I thank our Prophet *MOHAMMED* ,Peace Be Upon Him. who encourages us to seek for science.

Acknowledgement is due to **King Fahd University of Petroleum and Minerals** for support of this research.

I would like to express my deep appreciation to my thesis advisor, ***Dr. Maamar Bettayeb*** , for his support and his patient guidance. Also, I would like to express my gratefulness to the other committee members, ***Dr. Ubaid Al-Saggaf and Dr. Mahmoud Dawoud,*** for their valuable suggestions. and their kind cooperation.

Special thanks are to ***Mr. Mohammed Shahin, Mr. Rami Sukaik, Mr. Wael Shehadadh, Mr. Husam Abdawi, Mr. Hadeel Nouman, Mr. Ahmed Abu-Juyab, Abu-Diyab's brothers, and Rayhan's brothers*** for their assistance in producing this work in a nice form, and all my Brothers, Friends and many others for the memorable days we shared together.

خلاصة الرسالة

عنوان البحث : التقدير الطيفي اعتماداً على طرق تحليل القيم المفردة
اسم الطالب : اسماعيل محمود الفرا
التخصص : هندسة كهربائية
تاريخ الدرجة : ذو القعدة ١٤١٠هـ / يونيه ١٩٩٠م

فى كثير من التطبيقات العملية مثل معالجة الكلام ومستقبل الاشارات البعيدة (الرادار) ومستقبل الاشارات الصوتية (السونار) والأجهزة الطبية والأجهزة الجيوفيزيائية ، تكون الأشارات مشوشه ومن المطلوب تطوير نموذج ملائم يمثل هذه الاشارات أو تقدير المحتوى الطيفى لها . فى السابق كان بناء النموذج يقدر على اساس التقدير الخطى بينما يتم تحديد التحليل الطيفى بواسطة تنفيذ حسابات سريعة على الدالة التحويلية ولكن ظهرت فى العقد الأخير العديد من طرق التمثيل وأساليب التقدير الطيفى الحديثة التى حسنت اداء الطرق القديمة ، وقد سجل استعمال هذه الطرق نجاحاً باهراً وخاصة فى الحالات التى تتوفر فيها معلومات قليلة وهى الحالات التى عجزت عن معالجتها الطرق التقليدية .

فى هذا البحث نستعرض طريقة جديدة مثلى فى بناء النماذج وتقدير التحليل الطيفى للأشارات وينتج عن هذه الطريقة دالة قياسية تعتمد على التمازج بين الأقطاب والأصفار فى النموذج وتناسب مجموعة العينات المحدودة التابعة لإشارة مشوشة . وقد تم اقتراح قوة طبيعية للنموذج عن طريق تحليل القيم المفردة لمصفوفة هانكل ومقارنة القيمة النسبية العددية للقيم المفردة المحددة وتعطى هذه الطريقة الترددات المحدده للاشارة تستبعد الترددات المشوشه التى تنتج من استعمال قوة طبيعية غير مجديه للنموذج المشوش .

وقد طبقت هذه الطريقة الجديدة على عدة امثلة عملية تتفاوت فيها عدد العينات ونسب التشوش والتقارب فى ترددات الاشارة وقد قورنت النتائج من هذه الطريقة الجديدة مع نتائج التقليدية مثلا بلاكمان - توكى والحديثة مثل يول - ولكار وغيره ، وقد نوقشت النتائج المميزة لهذه الطريقة الحديثة من ناحية صعوبة الحسبة والمستوى التقريبى للأنجاز تحت ظرف إختلاف النسبة فى التشويش وفى حالة توفر معلومات قليلة عن الاشارة .

درجة الماجستير في العلوم
جامعة الملك فهد للبترول والمعادن
الظهران ، المملكة العربية السعودية
ذو القعدة ١٤١٠هـ - يونيه ١٩٩٠م

ABSRACT

Title : SPECTRAL ESTIMATION BASED ON SINGULAR VALUE DECOMPOSITION TECHNIQUES

By : **Ismail Mahmoud Al-Farra**

Major Field : **Electrical Engineering**

Date : **June 1990**

In many practical situations, such as, in radar, sonar, speech processing, biomedical and geophysics, samples of noisy signals are available and it is required to develop a suitable model or estimate its spectral contents. Traditionally, modelling is done in the context of linear prediction while spectrum estimation is determined via fast calculations on the periodogram. A multitude of modern modelling and spectrum estimation techniques have emerged the last decade to improve the performance of the classical methods. A net success has been registered with these new methods, specially in the often realistic situation of short data length, a case where, generally, classical methods perform poorly.

A new optimal approach developed in the context of model reduction will be applied for the problem of modelling and spectrum estimation of signals. A parametric ARMA model is explicitly derived to fit a finite set of samples from a noisy signal. Singular value decomposition of a Hankel data matrix is performed and a natural order of the model is suggested from the relative magnitude of the determined singular values. Unnecessary high order noise modelling is therefore eliminated and noise frquencies are suppressed to produce the desired deterministic frquencies of the signal.

The new algorithm for optimal signal modelling and spectrum estimation is tested for various practical situations of short data length, different signal to noise ratios and in the presence of close deterministic frequencies in the signal. The performance of the new technique is compared, in all these cases, with the classical Periodogram and Blackman-Tukey methods and most of the popular modern techiques of Yule-Walker, Burg, Cadzow, Beex, Kung, MUSIC, Prony, Pisarenko and Tufts.

The merits of the new technique are shown with respect to computational complexity, relative performance under different noise levels and in the case of short available data length.

MASTER OF SCIENCE DEGREE

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS

DHAHRAN, SAUDI ARABIA

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Spectral analysis plays a major role in modern communication systems . It often forms the basis for distinguishing signals and extracting relavent information in the presence of noise and other interfering signals . In certain applications, measuring the waveform of the incoming signals in time domain does not help in determining the different parameters of the signals . Consequently, energy content or power spectrum in frequency domain of the recevied signals is often measured or estimated .

1.2 HISTORICAL REVIEW

In many practical situations, samples of noisy signals are available and it is required to develop a suitable model or estimate its spectral contents . Consequently, many methods have been proposed and developed achieving the spectrum estimation . Some of these methods are called traditional methods and others are called modern methods .

There are two traditional methods in spectral estimation, both of which are based on Fourier Transform . The first is the power spectral estimate based on the indirect approach via an autocorrelation estimate and was proposed by Blackman and Tukey [27], [29], [41] . The second technique is based on the direct approach via the Fast Fourier Transform operation on the data and is known as periodogram [27], [29], [41], [61] . Schuster was the first to coin the term " periodogram " [29] . He made use of Fourier series fit to the variation in sun - spot numbers in a trial to find " hidden periodicities " in the measured data . Then, Norbert Wiener treated stochastic processes by using a Fourier Transform approach [29] . Later on, Welch [61] suggested a direct computation of a power spectrum estimate using Fast Fourier Transform . For short data records, both traditional techniques perform poorly . Both of which have the problem of windowing, which is assuming the data outside the interval to be zero . Furthermore, they are poor in estimating very close frequencies [29] .

To overcome the aforementioned problems, new techniques were developed based on assuming certain models for various systems . The parameters of these models are estimated from the set of received data . Rational functions can be used as models . These rational functions can be all pole rational functions (Auto-regressive (AR) models) . The AR models were, firstly, used by Yule [62] and Walker [60] to forecast trends in economic time series . Then Burg [10], [11], [12] in 1967 and Parzen [13] in 1968 employed these models in achieving spectral estimation . The Burg's method which is known as Maximum Entropy Method (MEM) led to many concepts that are now standard tools in spectral estimates . Theoretical considerations concerning the development of MEM have been introduced by Barnard (in 1969), Edward and Fitelson (in 1973), and Smylie et al (in 1973) . Van den Bos [59] showed that MEM is equivalent to AR power spectral density(PSD) estimator .

Also, rational functions can be all zero rational functions (Moving - average (MA) models) [13], [29] . Or, in many applications, the models can be a combination of both zeros (notches) and poles (peaks) (Auto - regressive Moving - average (ARMA) models) [13], [29] . The ARMA model is a composite of an AR model and an MA model . Although, the ARMA model might give better resolution and performance in spectral estimation for different applications, many practitioners prefer to use either AR or MA models .

Sinusoids at high signal to noise ratios (SNR) can be modelled by all pole model . While, at low SNR, the AR model of the signal containing both sinusoids and additive noise will give poorer resolution [64] . Pisarenko [45], in 1973, came up with an alternative spectral estimation technique based on a time series model of sinusoids plus additive white noise . It is known as Pisarenko Harmonic Decomposition (PHD) and is based upon a special case ARMA model . MUSIC algorithm, proposed by Schmidt [52] , generalized Pisarenko's method by relaxing the uniform sampling restriction and is also capable of utilizing all available autocorrelation lags in its solution . Recently, a new proposed approach to the signal parameter estimation problem called ESPRIT [68] (estimation of signal parameters via rotational invariance techniques) is used to estimate the parameters of complex sinusoids observed in noise . MUSIC method, Pisarenko's method, and ESPRIT exploit the underlying signal model . ESPRIT has advantages over MUSIC in direction finding applications and in spectral estimation with nonuniform sampling . Also, it and MUSIC have advantages over Pisarenko's method in utilizing all available lags obtained from the data .

Prony, in 1795, developed a technique for modelling data of equally spaced samples by a linear combination of exponentials while studying the effects of alcohol vapor pressures . Later on, Prony's method has been extended to estimate the power spectrum . It is applicable for a process consisting of damped sinusoids in noise and it is known as Prony Energy Spectral Density Estimation [29] . A special variants of Prony's method has been investigated for a process containing undamped sinusoids in noise . This approach is known as Prony Spectral Line Estimation [29] . Beex and Scharf [3] applied the modal decomposition to a covariance sequence data instead of a data as Prony's method . Also, they generalized the Pisarenko approach by fitting a deterministic damped sinusoid to the empirical correlations .

Capon [14] , in 1969, developed a technique for seismic array frequency - wave number analysis . This technique can be applied for estimating the power spectral density (PSD) . It is well known as the Maximum Likelihood Spectral Estimation (MLSE) [14], [29] . Lacoss [40] compared MEM with MLSE and other conventional methods .

A common problem with all the parametric methods is the determination and the selection of the model order . Different criteria have been proposed to determine the appropriate model order [29], [57] . The singular value decomposition (SVD) technique can play an important role in selecting the model order and estimating the power spectrum .

Based on the above, new different techinques in spectral estimation depending on SVD approach have been proposed . The SVD has been used by Tufts and Kumaresan (TK) [34] in estimating the frequencies of damped or undamped sinusoids . Also, it has been used in the Suboptimal Hankel Method

(SHM) for estimating the power spectrum by Kung et al [37] and others . Furthermore, Cadzow [13] has developed an efficient spectral estimation technique based on ARMA modeling and has proposed that SVD can be applied . This SVD based modelling perfoms well for the problem of short data records and results in low variance . It improves the accuracy of estimation and it gives good results . In the context of model order reduction of complex systems, Bettayeb [5], [53] developed a model based on SVD approach known as Optimal Hankel Model .

1.3 SCOPE OF THE THESIS

All the work in this thesis has been carried out on different sets of real data . This data matches practical situations . Simulations were done using Fortran programs .

In this thesis, most of the popular methods of power spectral estimation and frequency estimation will be reviewed and programmed in a single package . Also, an algorithm based on Hankel approach developed in the context of model reduction [5], [53] will be applied to the problem of power spectral estimation. This new optimal Hankel (OHM) algorithm will be evaluated and compared with some of the above methods . These methods will be evaluated on various different examples containing either AR process, MA process, ARMA process, sinusoids (damped or undamped), or a combination of both sinusoids and rational models. The evalution will take place in the presence of noise. The above will be considered for both direct data approach and covariance approximation approach . Many important aspects of the problem will be investigated . These include the effects of data length and the signal to noise ratios (SNR) on the performance of the algorithms . Resolution in terms of clossceness of frequencies in the spectrum will, also, be investigated .

1.4 THESIS ORGANIZATION

In chapter two, the traditional methods of power spectrum estimation are described . Also, some of the aforementioned modern methods will be explained in brief .

The algorithm (OHM) for spectrum estimation will be presented and explained in chapter three . Also, some sensitivity studies and model selection criteria will be presented .

In chapter four, a comprehensive simulation study will be carried on for all algorithms using different examples . Some conclusions and remarks will be drawn .

Some of the frequency estimation techniques will be presented in chapter five . Also, they will be tested for several examples and some conclusions could be drawn .

Finally, conclusions and suggestions for further work are given in chapter six .

CHAPTER 2

POWER SPECTRAL ESTIMATION TECHNIQUES

2.1 TRADITIONAL METHODS

There are two spectral estimation techniques based on Fourier Transformation.

2.1.1 Periodogram Method:

2.1.1.1 Definition of the periodogram:

Estimation of the power spectral density of a discrete sampled deterministic or stochastic data usually depends on algorithms based on the Discrete Fourier Transform (DFT) . When the process x(m) is a wide sense stationary stochastic process, the autocorrelation function

$$R_x(n)=\frac{1}{N}\sum_{m=0}^{N-|n|-1} x(m)x(n+m) \qquad ; \quad n=0,\pm 1,\ldots,\pm(N-1) \tag{2.1}$$

provides the basis for spectrum analysis rather than random process x(m) itself . The Wiener-Kinchine theorem relates the autocorrelation function via Fourier Transform to the power spectral density,

$$I_x(f) = \sum_{n=-(N-1)}^{(N-1)} R_x(n) e^{-j2\pi fn} \tag{2.2}$$

With the assumption that the process is ergodic and knowing the Fourier Transform of the sequence, it is possible to rewrite equation (2.2) in another form . The DFT of the real finite-length sequence x(m), $0 \le m \le N-1$,is:

$$X(e^{j\omega}) = \sum_{m=0}^{N-1} x(m) e^{-j2\pi fm} \tag{2.3}$$

Then, substituting (2.1) into (2.2), we get :

$$I_x(f) = \frac{1}{N} | X(e^{j\omega}) |^2 \tag{2.4}$$

where the discrete $I_x(f)$ has been termed periodogram spectral estimate .

2.1.1.2 Welch method - Averaging over short, modified periodograms:

Welch [61] has suggested a direct computation of a power spectrum estimate using FFT .

Let x(i), i= 0,1, ..., N-1 be a sample from a wide-sense stationary sequence with zero mean . Taking segments, possibly overlapping, of length L with the starting points of these segments D units apart, it will lead to :

$$x_1(i) = x(i) \quad (2.5a)$$

$$x_2(i) = x(i+D) \quad (2.5b)$$

.

.

.

$$x_K(i) = x(i+(k-1)D) \quad (2.5c)$$

where (K-1)D + L = N and i = 0,1, ..., L-1

For each segment of length L, the modified periodogram will be calculated That is, a data window, W(i), i = 0,1, ..., L-1, will be selected, and we form the sequences $x_1(i)W(i),...,x_K(i)W(i)$. Different choices for the data window or weighting function, such as Hanning and Bartlett, have been suggested . The choice of one depends on the type of applications . It is usually chosen so that the resultant spectral window area is unity . Weighting functions have the common feature of multiplying the unmeasured samples by zero . The particular weightings applied to the known samples governs the shape of the corresponding frequency window which when convolved with the true power spectrum gives the estimated spectrum . The Fourier Transform will be computed for the above sequences . That is,

$$J_k(n) = \frac{1}{L}\sum_{i=0}^{L-1} x_k(i)\, W(i)\; e^{-j2\pi n\frac{i}{L}} \quad (2.6)$$

Finally, the K modified periodogram will be obtained .

$$I_k(f_n) = \frac{L}{U}\, |J_k(n)|^2 \qquad ; \; k = 1, 2, \ldots, K \quad (2.7)$$

where

$$f_n = \frac{n}{L} ; \quad n = 0 , 1 , \dots , \frac{L}{2}$$

and

$$U = \frac{1}{L} \sum_{i=0}^{L-1} W^2(i) \tag{2.8}$$

The spectral estimate will be the average of these periodograms.

$$P_x(f) = \frac{1}{K} \sum_{k=1}^{K} I_k(f_n) \tag{2.9}$$

This method will be applied for different examples in the fourth chapter.

2.1.2 Blackman - Tukey Method:

This method will estimate the power spectral density (PSD) via an autocorrelation estimate . With a finite data record, only a finite number of discrete autocorrelation function values can be computed . Blackman and Tukey proposed the spectral estimate [29]:

$$P_{BT}(f) = \Delta t \sum_{n=-N}^{N} R_x(n)\, e^{-j2\pi f n \Delta t} \tag{2.10}$$

where $\frac{-1}{2\Delta t} \leq f \leq \frac{1}{2\Delta t}$ and $R_x(n)$ can be computed from equation (2.1).

The above two methods have advantages and disadvantages as mentioned before . Both of them are computation efficient . The power spectral density estimate is directly propotional to the power for sinusoid processes . On the other hand, the frequency resolution is limited by the available data record duration and the estimated power spectral density will be distorted due to the side-lobe leakage, which is a result of windowing . Also, in the BT approach, some negative PSD values will appear .

2.2 MODERN METHODS

Given any set of data with length N:

$$y_k = s_k + w_k \quad , 1 \le k \le N \tag{2.11}$$

where, y_k is the received signal , s_k is the information signal and w_k is any white random noise . The problem is to estimate the power spectral density (PSD) from the received noisy data . The parametric power spectrum estimation consists of developing a linear model in state space or frequency domain that fits the data in some manner . A rational transfer function model can approximate many deterministic and stochastic discrete - time processes, the output sequence y_k and the input sequence n_k are related by :

$$y_k = \sum_{j=0}^{q} b_j \, n_{k-j} - \sum_{i=1}^{p} a_i \, y_{k-i} \tag{2.12}$$

eqn (2.12) can be rewritten as :

$$\sum_{j=0}^{q} b_j \, n_{k-j} = \sum_{i=0}^{p} a_i \, y_{k-i} \tag{2.13}$$

where $a_0 = 1$

Taking the z-transform for both sides in eqnuation (2.13)

$$N(Z) \sum_{j=0}^{q} b_j z^{-j} = Y(z) \sum_{i=0}^{p} a_i z^{-i} \tag{2.14}$$

Therefore, the general transfer function is :

$$H(z) = \frac{Y(z)}{N(z)} = \frac{B(z)}{A(z)} = \frac{\sum_{j=0}^{q} b_j z^{-j}}{\sum_{i=0}^{p} a_i z^{-i}} \tag{2.15}$$

And, the power spectrum of the output, $P_y(z)$, is related to the input power spectrum, $P_n(z)$, as follows

$$P_y(z) = |H(z)|^2 P_n(z) \tag{2.16}$$

With the assumption that the input is white - noise sequence with zero mean and variance ε^2. Therefore, the input power spectral density function is given by :

$$P_n(z) = \varepsilon^2 \Delta t \tag{2.17}$$

with

$$\frac{-1}{2\Delta t} \leq f \leq \frac{1}{2\Delta t}$$

This will reduce the problem to a matter of estimating the parameters of the filter .

Multiplying both sides of eqn.(2.13) by y_{k-n} and summing over k will give

$$\sum_{j=0}^{q} b_j \sum_{k=0}^{\infty} y_{k-n} n_{k-j} = \sum_{i=0}^{p} a_i \sum_{k=0}^{\infty} y_{k-n} y_{k-i} \tag{2.18a}$$

Taking the expectation, equation (2.18a) will yield :

$$\sum_{j=0}^{q} b_j R_{ny}(n-j) = \sum_{i=0}^{p} a_i R_y(n-i) \tag{2.18b}$$

But $R_{ny}(t) = 0$ for $t > 0$ since a future input to a stable, causal filter does not affect the present output and n_k is white noise process . Therefore, equation (2.18b) will be reduced to :

$$\sum_{i=0}^{p} a_i R_y(n-i) = \begin{cases} \sum_{j=0}^{q} b_j R_{ny}(n-j), & |n| \le q \\ 0 \quad , & otherwise \end{cases} \tag{2.18c}$$

From the derivation of the Yule-Walker equations, it was proven that :

$$R_{ny}(t) = \varepsilon^2 h_{-t}$$

where ε^2 is the variance of the white noise, and therefore,

$$\sum_{i=0}^{p} a_i R_y(n-i) = \begin{cases} \varepsilon^2 \sum_{j=0}^{q} b_j h_{j-n}, & |n| \le q \\ 0 \quad , & otherwise \end{cases} \tag{2.18d}$$

Expression (2.18d) is the general Yule - Walker equations .

2.2.1 The AR PSD Estimation: [13,29]

Given a zero mean signal y_n, we wish to estimate its PSD based on all pole model via Yule-Walker equations . To determine the AR parameters, one need only r equations from (2.18d) for n > 0, solve for { $a_1 . a_2 . ... . a_r$ }, and then find ε^2 from (2.18d) for n = 0 .

Recalling equation (2.18d), when n = 0, it will be reduced to

$$\sum_{i=0}^{r} a_i \; R_y(-i) = \varepsilon^2 \tag{2.19}$$

The iterative Levinson-Durbin algorithm in solving equation (2.19) is given in steps as follows :

1. For k = 0, $\varepsilon_0^2 = R_y(0)$
2. For step k with $a_0(k) = 1$.

   a. $K_k = \frac{-1}{\varepsilon_{k-1}^2} \sum_{i=0}^{k-1} a_i(k-1) \; R(k-i)$

   b. $a_k(k) = K_k$

   c. For i = 1 to k-1

   $$a_i(k) = a_i(k-1) + K_k \, a_{k-i}(k-i)$$

   d. $\varepsilon_k^2 = \varepsilon_{k-1}^2 \, (1 - K_k^2)$

Where K_k are known as reflection coefficients. Then, the estimated PSD for model order r will be :

$$P_y(f) = \frac{\epsilon_r^2 \Delta t}{\left|\sum_{i=0}^{r} a_i(r)\, z^{-i}\right|^2} \tag{2.20a}$$

where

$$z = e^{j2\pi f i \Delta t} \tag{2.20b}$$

The problem of AR parameter identification is related to the theory of linear prediction. Given an AR process, y_n, we wish to estimate $\hat{y}_k$ as a linear combination of the past values and the model is all-pole predictor . The linear predictor [29] is :

$$\hat{y}_k = -\sum_{i=1}^{r} s_i\, y_{k-i} \tag{2.21a}$$

Then $\{ s_1 , s_2 , \ldots , s_r \}$ can be chosen to give the minimum prediction error power .

The forward error is :

$$y_k - \hat{y}_k = f_k = \sum_{i=0}^{r} a_i\, y_{k-i} \tag{2.21b}$$

and, the backward error is :

$$e_k = \sum_{i=1}^{r+1} b_i\, y_{k-i} \tag{2.21c}$$

By using the orthogonality principle,

$$E\{f_k\, y_l\} = 0 \ , \text{ for } l = k\text{-}1 , k\text{-}2 , \ldots , k\text{-}r \tag{2.21d}$$

the minimum prediction error power can be found . Also, it was found that the best linear predictor is just when $s_i = a_{ri}$ for $i = 1 , \ldots , r$.

2.2.2 Maximum Entropy Method:

Burg [10] has proposed an identical method for AR PSD method . Both of them are identical for Gaussian random process and known autocorrelation sequence of uniform spacing . It can be shown that the entropy rate for a Gaussian random process with zero mean and bandwidth $B = \frac{1}{2T}$ is propotional to

$$\int_{-B}^{B} \ln P_{MEM}(f)\ df \qquad (2.22)$$

where, $P_{MEM}(f)$ is the PSD of the recevied data .

$P_{MEM}(f)$ is found by maximizing equation (2.22) subject to the constraint that this function will be consistent with the given set of (r + 1) lags through the Wiener-Kinchine relationship . The solution is found by using the Lagrange multiplier technique and it is all-pole model . Then, the problem is to find the prediction coefficients a_i's of this model . So, we will be finding the value of K_r which minimizes the sum of the squares of both forward and backward errors . This is

$$E = \frac{1}{2(N-r)} \sum_{i=r}^{N-1} [\ f_i^2(r) + e_{i+1}^2(r)\ ] \qquad (2.23)$$

where the forward and backward error are related by :

$$f_k(r) = f_k(r\text{-}1) + K_r e_k(r\text{-}1) \tag{2.24a}$$

$$e_k(r) = e_{k\text{-}1}(r\text{-}1) + K_r f_{k\text{-}1}(r\text{-}1) \tag{2.24b}$$

Therefore equation (2.23) will become :

$$E = \frac{1}{2(N\text{-}r)} \sum_{i=r}^{N\text{-}1} [\, |\, f_i(r\text{-}1) + K_r e_i(r\text{-}1)\, |^2 + |\, e_i(r\text{-}1) + K_r f_i(r\text{-}1)\, |^2\,] \tag{2.25}$$

Differentiating equation (2.25) with respect to the reflection coefficient K_r:

$$\frac{\partial E}{\partial K_r} = \frac{1}{(N\text{-}r)} \sum_{i=r}^{N\text{-}1} [\, K_r\, [f_i^2(r\text{-}1) + e_i^2(r\text{-}1)\,] + 2e_i(r\text{-}1)\, f_i(r\text{-}1)\,] \tag{2.26}$$

Equating equation (2.26) to zero gives the following formula for K_r :

$$K_r = -2\frac{\sum_{i=r}^{N\text{-}1} e_i(r\text{-}1)\, f_i(r\text{-}1)}{\sum_{i=r}^{N\text{-}1} [f_i^2(r\text{-}1) + e_i^2(r\text{-}1)\,]} = -2\frac{S}{D} \tag{2.27}$$

The Burg's algorithm can be summarized as follows :

1. For k=0

   a. $R_y(0) = \frac{1}{N}\sum_{i=1}^{N} y_i^2$

   b. $f_i(0) = y_i$ and $e_i(0) = y_{i\text{-}1}$

2. For step k

a. $K_k = -2\frac{S}{D}$

b. $a_k(k) = K_k$

c. For j=1 to k-1

$$a_j(k) = a_j(k-1) + K_k\, a_{k-j}(k-1)$$

d. $R_y(k) = -\sum_{i=1}^{k} a_i(k) R_y(p-i)$

e. For i=k to N-1

$$f_i(k) = f_i(k-1) + K_k e_i(k-1)$$

$$e_i(k) = e_{i-1}(k-1) + K_k f_{i-1}(k-1)$$

Then the estimated PSD via MEM for model order r will be :

$$P_y(f) = \frac{\varepsilon_r^2 \Delta t}{|\sum_{i=0}^{r} a_i(i)\; z^{-i}|^2} \tag{2.28a}$$

where

$$z = e^{j2\pi f \Delta t} \tag{2.28b}$$

2.2.3 Cadzow's Method:

Cadzow [13] has proposed that the singular value decomposition (SVD) technique can be applied to ARMA modeling . Recalling equation (2.18d), we find that :

$$R_y(n) = \begin{cases} \varepsilon^2 \sum_{i=0}^{q_e} b_i h_{i-n} - \sum_{i=1}^{p_e} a_i R_y(n-i) \ , & |n| \le q_e \\ - \sum_{i=1}^{p_e} a_i R_y(n-i) \ , & otherwise \end{cases} \tag{2.29}$$

Since, this model is rational with $p_e > q_e$, the extended Yule-Walker equations [13] can be found for $n \ge q_e + 1$. It follows :

$$\sum_{i=0}^{p_e} a_i R_y(n-i) = 0 \qquad , \text{ for } n \ge q_e + 1 \tag{2.30}$$

The extended Yule-Walker equations will be evaluated for l distinct values of n which are satisfying $l \ge p_e + 1$ and $n \ge q_e + 1$. Equation (2.30) can be written in matrix form as follows :

$$\begin{bmatrix} R_y(q_e+1) & R_y(q_e) \ldots & R_y(q_e - p_e + 1) \\ R_y(q_e+1) & & \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ R_y(q_e + l) & \cdot \quad \ldots & R_y(q_e - p_e + l) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_{p_e} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{2.31a}$$

or,

$$R_e A = 0 \tag{2.31b}$$

The Cadzow's algorithm for power spectral density estimation is given in steps as follows :

1. Perform the SVD to the autocorrelation matrix R_e . R_e is an $l \times p_e + 1$ matrix .

$$R_e = U \Sigma V^T \tag{2.32}$$

where, U and V are $l \times l$ and $(p_e + 1) \times (p_e + 1)$ unitary matrices respectively, and Σ is a $l \times (p_e + 1)$ matrix . The required AR order r is obtained by examining the computed singular values . Therefore, the net result of this step will be a rank R optimum approximation of the $l \times (p_e + 1)$ extended order autocorrelation matrix, that is,

$$R_e^r = U \Sigma_r V^T \tag{2.33}$$

A simple matrix manipulation reveals that this rank r approximation may be equivalently represented as:

$$R_e^r = \sum_{i=1}^{r} \sigma_{ii} u_i v_i^T \tag{2.34}$$

where u_i and v_i are i^{th} column vectors of U and V matrices .

2. Estimate the autoregressive parameters .

   a. For ARMA(p_e , q_e) model .

      The matrix R_e^r can be rewritten as follows :

$$R_e^r = [R_1^r \quad R_a^r] \tag{2.35}$$

where, R_1^r is a $l \times 1$ column vector of R_e^r and R_a^r is the $l \times p_e$ matrix of R_e^r . The AR parameter vector A can be computed, with first component equals to one, as follows [13] :

$$\begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{r_e} \end{bmatrix} = -[\,[R_a^r]^T R_a^r]^{-1} [R_a^r]^T R_1^r \tag{2.36}$$

Or, the minimum norm solution can be simplified to [13] :

$$A = \frac{\sum_{i=r+1}^{p_e} \bar{v}_i(0) v_i}{\sum_{i=r+1}^{p_e} |\bar{v}_i(0)|^2} \tag{2.37}$$

where, v_i is the i^{th} column vector of the matrix V and $\bar{v}_i(0)$ is the first component of the vector v_i .

b. For ARMA(r,q) model

Here, we will seek for an AR model with order r and $r < p_e$. The matrix S^r will be formed as follows :

$$S^r = \sum_{i=1}^{p_e - r + 1} [R_i^r]^T R_i^r \tag{2.38}$$

where R_i^r is a $l \times (r+1)$ submatrix of R_e^r and it is composed of the columns (i) through (i + r) of R_e^r . The AR parameter vector A can be computed by solving the following equations :

$$S^r A = \alpha e_1 \tag{2.39}$$

where the constant α can be found by letting the first component of A equals to one and e_1 is the first column vector of the identity matrix .

3. Estimate the moving average parameters .

These parameters can be obtained directly by passing y_k through A(z) whose coefficients correspond to AR parameters obtained above . The filtering yields the so-called residual time series . This filtering causes the residual time series to be a moving average process of order q . The autocorrelation lags of this residual time series can be computed as follows :

$$R_{res}(n) = \begin{cases} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j R_y(n+j-i) , & |n| \leq q \\ 0 , & otherwise \end{cases} \qquad (2.40a)$$

4. Compute the power spectral density

$$P_y(f) = \frac{\sum_{n=-q}^{q} R_{res}(n) e^{-j2\pi fn}}{|\sum_{i=0}^{r} a_i(r) z^{-i}|^2} \qquad (2.40b)$$

2.2.4 Suboptimal Hankel Method:

Given any set of data with length N,

$$y_k = s_k + w_k \quad 1 \leq k \leq N \qquad (2.41a)$$

where, y_k is the received signal, s_k is the information signal, and w_k is any white random noise . A state space representation is given as :

$$x_{k+1} = A x_k + b u_k \tag{2.41b}$$

$$y_k = c x_k \tag{2.41c}$$

where, x_k is the state vector and it is a $r \times 1$ vector process . A, b, and c are constant matrices of sizes $r \times r$, $r \times 1$, and $1 \times r$ respectively . Also, y_k and u_k are the output and the input vectors . Finally, r is the model order . The transfer function related to the above state space representation as :

$$\frac{Y(z)}{U(z)} = F(z) = c(zI-A)^{-1}b = \sum_{i=0}^{\infty} h_i z^{-i} \tag{2.42}$$

where, F(z) is the frequency domain transfer function . Therefore, the relationship between the impulse response of the model and state space parameters is given by :

$$h_k = cA^{k-1}b \tag{2.43}$$

The infinite Hankel matrix is defined as:

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ h_3 & h_4 & h_5 & \cdots \\ \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \\ \cdot & \cdot & \cdot & \cdots \end{bmatrix} = \begin{bmatrix} c \\ cA \\ cA^2 \\ \vdots \end{bmatrix} [\, b \; Ab \; A^2b \ldots]$$

$$= OC \tag{2.44}$$

where O is the infinite observability matrix and C is the infinite controllability matrix . Because our data is limited, then, the infinite Hankel matrix can not be formed . So, we have to pend with different variants to justify the reasonable approach .

After the Hankel matrix has been formed, two steps should be carried in order to determine the state space parameters . This approach is due to Kung *et al* [37] . The algorithm can be summarized as follows :

1. Perform SVD of the $N \times N$ Hankel matrix H and order the singular values of H in descending manner :

$$H = U\Sigma V^T = [\,U_1 \quad U_2\,] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{2.45}$$

where Σ_1 is the size $r \times r$ matrix and it contains all the dominant singular values . While, Σ_2 contains the smaller singular values . As a result of the above, H can be approximated as :

$$\overline{H} = U_1 \Sigma_1 V_1^T = O_1 C_1 \tag{2.46a}$$

where

$$O_1 = U_1 \Sigma_1^{1/2} \quad , \text{ and } \quad C_1 = \Sigma_1^{1/2} V_1^T \tag{2.46b}$$

Also, a balanced realization can be obtained by choosing:

$$O = U \Sigma^{1/2} \quad , \text{ and } \quad C = \Sigma^{1/2} V^T \tag{2.46c}$$

If we define

$$H^{-} = O_1 A C_1 \tag{2.47a}$$

then,

$$A = O_1^{+} H^{\leftarrow} C_1^{+} = O_1^{+} O_1^{\uparrow} = C_1^{\leftarrow} C_1^{+} \qquad (2.47b)$$

where

$$O_1^{+} = (O_1^{T} O_1)^{-1} O_1^{T} \qquad (2.48a)$$

$$C_1^{+} = C_1^{T} (C_1 C_1^{T})^{-1} \qquad (2.48b)$$

where, $O_1^{\uparrow}$ is the submatrix of O_1 shifted up by one row and $C_1^{\leftarrow}$ is the submatrix of C_1 shifted left by one column .

2. The state space parameters can be formed as :

   a. b is the first column of C_1

   b. c is the first row of O_1

   c. A is given in (2.47b)

3. Obtain $F(z) = c(zI-A)^{-1}b$

4. Find the power spectral density (PSD) :

$$P_y(f) = F(z)F^{*}(z)$$

where, $F^{*}(z)$ is the complex conjugate of $F(z)$.

2.2.5 Covariance Approximation Method:

The parametric covariance sequence approximation for the purposes of model identification and spectrum estimation was proposed by Beex and Scharf

[3] . The order - r covariance sequence approximant can be written as :

$$R_k(r) = \sum_{i=1}^{r} A_i z_i^k \tag{2.49}$$

$$R_k(r) = R_{-k}(r) \tag{2.50a}$$

$$z_i = \rho_i e^{j2\pi f_i} \tag{2.50b}$$

where, z_i is the i^{th} complex parameter for frequency f_i, with radius ρ_i, and A_i is the corresponding mode weight . The problem is to estimate a finite covariance string ($R_y(0)$, $R_y(1)$, ... , $R_y(N\text{-}1)$) from a noisy data record and identifying the parameters (A_i , z_i) to fit $R_k(r)$ to the string . The covariance sequence can be estimated as follows :

$$R_k = \frac{1}{N-|k|} \sum_{i=1}^{N-|k|} y_i y_{i+k} \quad , k = 0, \pm 1, \ldots, \pm (N\text{-}1) \tag{2.51}$$

The problem arises in fitting R_k with $R_k(r)$.

1. Least squares :

The squared error between $R_k(r)$ and R_k is defined as follows :

$$E = \sum_{k=-(N-1)}^{(N-1)} | R_k - R_k(r) |^2 \tag{2.52}$$

Taking the derivatives of E with respect to (A_i , z_i), i = 1,2,...,r, and setting them to zero will yield a discrete form of the Aigrain - Williams equation .

$$\frac{\partial E}{\partial A_i} = \sum_{k=-(N-1)}^{(N-1)} [\, R_k - R_k(r) \,] \, z_i^{|k|} = 0 \qquad (2.53a)$$

$$\frac{\partial E}{\partial z_i} = \sum_{k=-(N-1)}^{(N-1)} [\, R_k - R_k(r) \,] \, z_i^{|k-1|} A_i \, |k| = 0 \qquad (2.53b)$$

for $i = 1 , 2 , \ldots , r$

By using the symmetry of R_k and $R_k(r)$ sequences, the above system of equations can be written as follows :

$$P^T Q R_y = P^T Q P A \qquad (2.54a)$$

$$\overline{P}^T Q R_y = \overline{P}^T Q P A \qquad (2.54b)$$

$$P = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & & z_r \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ z_1^{r-1} & \cdot & \cdots & z_r^{r-1} \end{bmatrix}$$

$$Q = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 2 & 0 & & & & 0 \\ 0 & 0 & 2 & & & & 0 \\ \cdot & \cdot & & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot & & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 2 \end{bmatrix}$$

where,

$$R_y^T = (R_0 , R_1 , \ldots , R_{r-1})$$

$$A = (A_1 , A_2 , \ldots , A_r)$$

$$\bar{P} = \sum_{i=1}^{r} A_i \frac{\partial P}{\partial z_i}$$

2. The modified squared error :

The modified squared error can be written as follows :

$$E_m = \sum_{k=r}^{(N-1)} \left| \sum_{i=o}^{r} a_i \left[R_{k-i} - R_{k-i}(r) \right] \right|^2 , a_0 = 1 \tag{2.55}$$

If we select the filters weights to give :

$$\sum_{i=0}^{r} a_i z^{-i} = \prod_{i=1}^{r} \left[1 - z_i z^{-1} \right] \tag{2.56}$$

The approximating sequence $R_k(r)$ will satisfy the following homogenous equation on its tail :

$$\sum_{i=1}^{r} a_i R_{k-i} = 0 , \quad \text{for} \quad k \geq r \tag{2.57}$$

This will reduce equation (2.55) to be :

$$E_m = \sum_{k=r}^{(N-1)} \left| \sum_{i=o}^{r} a_i \; R_{k-i} \right|^2 \qquad , a_0 = 1 \tag{2.58}$$

Minimization with respect to a_i will lead to a covariance method of linear prediction on the tail of the covariance sequence [3] :

$$R^T R a = - R^T r_r \tag{2.59}$$

where,

$$R = \begin{bmatrix} R_{r-1} & \cdots & R_0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ R_{N-2} & \cdots & R_{N-r-1} \end{bmatrix}$$

$$r^T_r = (R_r , R_{r+1} , \ldots , R_{N-1})$$

$$a^T = (a_1 , a_2 , \ldots , a_r)$$

The algorithm will start in part(2) as follows:

1. Find the covariance lags from equation (2.51)
2. Solve for a_i in equation (2.59)
3. Find the coressponding modes z_i from equation (2.56)
4. Solve equation (2.54) to get the mode weights A_i
5. Perform the DFT for $R_k(r)$ to get power spectral density

2.2.6 MUSIC Method:

The multiple signal classification (MUSIC) algorithm is basically developed by Schmidt [52] . It estimates the frequencies of the sinusoids as the peaks of the " spectral estimator " [52] . This method is applicable for the data consisting of r complex (real) sinusoids in complex (real) white Gaussian noise [27]

$$y_n = \sum_{i=1}^{r} A_i e^{[j\omega_i n + \theta_i]} + w_n \qquad (2.60)$$

For w_n is complex (real) white Gaussian noise with zero mean and variance ε_n^2, n = 1,2, ..., N.

The autocorrelation function for the random process given in (2.60) is given as follows :

$$r_y(n) = \sum_{i=1}^{r} A_i^2 e^{[j\omega_i n]} + \varepsilon_n^2 \delta(n) \qquad (2.61)$$

With the assumption that the phase is independent random variable and uniformly distributed over the interval [0 , 2 π) . So, the general $N \times N$ auto-correlation matrix R_y can be found as:

$$R_y = \begin{bmatrix} r_y(0) & \cdots & r_y(-N) \\ r_y(1) & & \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ r_y(N) & \cdots & r_y(0) \end{bmatrix} \qquad (2.62)$$

where the matrix R_y is Toeplitz .

The matrix R_y can be written in terms of the autocorrelation matrices of the information signals and the noise .

$$R_y = R_s + \varepsilon_n^2 I \qquad (2.63a)$$

$$R_s = \begin{bmatrix} r_s(0) & \dots & r_s(-N) \\ r_s(1) & & \\ \cdot & & \cdot \\ \cdot & & \\ \cdot & & \cdot \\ r_s(N) & \dots & r_s(0) \end{bmatrix} \quad (2.63b)$$

Where the rank of the matrix R_s will be r while the rank of the matrix R_y is N, which is due to the additive noise . The matrix R_s can be rewritten as follows :

$$R_s = \sum_{i=1}^{r} A_i^2 e_i e_i^* \quad (2.64a)$$

where

$$e_i = [\, 1 \; e^{j\omega_i} \dots e^{j\omega_i(N-1)} \,] \quad (2.64b)$$

and, (*) denotes the complex cojugate transpose .

The identity matrix can be written in terms of orthonormal set of vectors as follows :

$$I = \sum_{i=1}^{N} V_i V_i^* \quad (2.65)$$

Where, V_i is a $N \times 1$ eigenvector corresponding to λ_i eigenvalue of the matrix R_y . The nonprincipal eigenvectors are chosen to be (N - r) eigenvectors of R_y that have smallest eigenvalues [27] . Based on the orthogonality between those vectors V_i's, the power spectral density can be estimated via MUSIC method as follows :

$$P(f) = \frac{1}{e^* \left[\sum_{i=r+1}^{N} V_i V_i^* \right] e}$$

$$= \frac{1}{\sum_{i=r+1}^{N} |e^* V_i|^2} \qquad (2.66)$$

Equivalently, Kay and Demeure [26] showed that P(f) can be computed this way:

$$P'(f) = \frac{1}{1-\hat{P}(f)} \qquad (2.67a)$$

where

$$\hat{P}(f) = \sum_{i=1}^{r} |(e')^* V_i|^2 \qquad (2.67b)$$

and,

$$e' = \frac{1}{\sqrt{N}} e \qquad (2.67c)$$

Theoretically, the estimated power spectrum should go to infinity as f approaches any sinusoidal frequency .

2.2.7 Maximum Likelihood Method:

Capon [14], in 1969, developed the maximum likelihood method (MLM) for seismic array frequency - wave number analysis . This technique can be applied for estimating the power spectral density (PSD) .

Suppose that a discrete time series is to be filtered, the sampled output is obtained as

$$y_k = \sum_{i=1}^{N} a_i \, x_{k+1-i} \tag{2.68}$$

where x is the input and the a_i's are the filter coefficients .

This method is applicable for the data consisting of complex (real) sinusoids in complex (real) Gaussian noise .

$$x_k = A\, e^{j\omega k} + n_k \tag{2.69}$$

The filter weights are chosen to pass $A\, e^{j\omega k}$ and reject n_k . If y_k is to be an unbaised estimate of $A\, e^{j\omega k}$,

$$A\, e^{j\omega k} = \sum_{i=1}^{N} a_i \, A\, e^{j\omega(k+1-i)} \tag{2.70a}$$

or,

$$\sum_{i=1}^{N} a_i \, A\, e^{j\omega(1-i)} = 1 \tag{2.70b}$$

To obtain the coefficients of the filter, one minimizes the variance of y_k, given by

$$\varepsilon^2 = a^* R\, a \tag{2.71}$$

where R is the $N \times N$ covariance matrix of the noise process . Then, the estimated power spectrum is given by

$$P(f) = \frac{1}{e^{*} R^{-1} e} \tag{2.72}$$

where,

$$e = [\, 1 \;\; e^{j\omega} \;\; \dots \;\; e^{j\omega(N-1)} \,] \tag{2.73}$$

and * denotes complex conjugate transpose .

2.2.8 Prony's Method:

Prony, in 1795, developed a technique for modelling data of equally spaced samples by a linear combination of exponentials . Prony's method has been extended to estimate the power spectrum of a process consisting of either damped or undamped sinusoids . Originally, the exact procedure fitted an exponential curve having r exponential terms to 2r measured data . Here, an approximate fit with r exponentials to a data sequence of length N is desired, such that N>2r, a least squares estimation procedure is encountered . The model assumed in the extended Prony method is a set of r exponentials of arbitrary amplitude, frequency, damping factor, and phase . The discrete-time function

$$\hat{y}_k = \sum_{i=1}^{r} b_i z_i^k \qquad ; k = 0 , 1 , \dots , N\text{-}1 \tag{2.74}$$

is the proposed model to approximate the N data measurements. In general , b_i and z_i are assumed to be complex and

$$b_i = A_i e^{j\theta_i} \tag{2.75a}$$

$$z_i = e^{(\alpha_i + j2\pi f_i)} \tag{2.75b}$$

Finding the parameters and r that minimize the squared error

$$E = \sum_{k=0}^{N-1} [y_k - \hat{y}_k]^2 \tag{2.76}$$

is a difficult nonlinear least squares problem . The solution comes by realizing that equation (2.74) is the homogeneous solution to a constant coefficient linear difference equation . Define the polynomial G(z) as

$$G(z) = \prod_{k=1}^{r} (z - z_k) = \sum_{i=0}^{r} a_i z^{r-i} \quad , a_0 = 1. \tag{2.77}$$

Equation (2.74) can be rewritten as follows :

$$\hat{y}_{k-n} = \sum_{i=1}^{r} b_i z_i^{k-n} \qquad , 0 \leq k-n \leq N-1 \tag{2.78}$$

Multiplying equation (2.78) by a_n and summing over the past r+1 products results in :

$$\sum_{n=0}^{r} a_n \hat{y}_{k-n} = \sum_{i=1}^{r} b_i \sum_{n=0}^{r} a_n z_i^{k-n} \tag{2.79}$$

If in equation (2.79) the substitution, $z_i^{k-n} = z_i^{k-r} z_i^{r-n}$ is performed, then

$$\sum_{n=0}^{r} a_n \hat{y}_{k-n} = \sum_{i=1}^{r} b_i z_i^{k-r} \sum_{n=0}^{r} a_n z_i^{r-n} = 0 \tag{2.80}$$

Therefore, equation (2.80) yields the recursive difference equation

$$\hat{y}_n = -\sum_{i=1}^{r} a_i \hat{y}_{n-i} \tag{2.81}$$

defined for $r \leq n \leq N-1$

Therefore, the coefficients a_i's can be determined by any AR parameter estimation algorithm mentioned before . Then, the roots of equation (2.77) can be determined . After that, expression (2.74) reduces to a set of linear equations in the unknown b_i parameters, expressible in matrix form :

$$P B = \hat{Y} \tag{2.82}$$

where,

$$P = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & & z_r \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ z_1^{N-1} & \cdot & \cdots & z_r^{N-1} \end{bmatrix}$$

$$B = [\, b_1 \, b_2 \ldots b_r \,]^T$$

$$\hat{Y} = [\, \hat{y}_0 \, \hat{y}_1 \ldots \hat{y}_{N-1} \,]^T$$

Thus, the amplitudes and phases can be determined by :

$$B = [P^* P]^{-1} P^* \hat{Y} \tag{2.83}$$

Then, compute

$$A_i = | b_i | \tag{2.84a}$$

$$\theta_i = \tan^{-1} [\frac{Im(b_i)}{Re(b_i)}] \tag{2.84b}$$

$$\alpha_i = \ln | z_i | \tag{2.84c}$$

$$f_i = \frac{\tan^{-1} \left[\frac{Im(z_i)}{Re(z_i)} \right]}{2\pi} \tag{2.84d}$$

Finally, the spectrum can be obtained by computing the transform of exponential model and taking the modulus .

$$P_{Prony}(f) = | \hat{Y}(f) |^2 \tag{2.85a}$$

where,

$$\hat{Y}(f) = \sum_{i=1}^{r} A_i e^{j\theta_i} \frac{2\alpha_i}{[\alpha_i^2 + (2\pi[f - f_i])^2]} \tag{2.85b}$$

This method will not be simulated and tested because it is more or less a special case from Tufts-Kumaresan method and related to Covariance Approximation method .

CHAPTER 3

OPTIMAL HANKEL METHOD

3.1 OPTIMAL HANKEL ALGORITHM

The optimal Hankel approximation for a given N samples of noisy impulse response was developed by Bettayeb [5,53] in the context of model reduction and identification of linear dynamical systems .

Let

$$Y(z)=\sum_{i=1}^{N} y_i z^{-i} \tag{3.1}$$

where $y_i's$ are the samples, and let

$$\|f(z)\|_\infty = \max_{0\leq\omega\leq 2\pi} |f(e^{j\omega})|$$

be the infinite norm. Then, the following result is given in [5,53].

Theorem :

The unique approximation $g^r(z)$ which minimizes $\| Y(z)-g(z)\|_\infty$ over all rational functions with r-stable poles is explicitly given by:

$$g^r(z)=\frac{\sum_{J=1}^{N-1} z^{N-J-1}\sum_{i=1}^{J} y_{J-i+1}m_{N-i+1}}{\sum_{i=1}^{N} m_i z^{i-1}} \tag{3.2}$$

where $M = [m_1\ m_2 \ldots m_N]^T$ is the eigenvector corresponding to the $(r+1)^{st}$ eigenvalue of H_N, and

$$H_N = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \\ y_2 & y_3 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & y_N & & \cdot \\ y_N & 0 & \cdots & 0 \end{bmatrix} \qquad (3.3)$$

It is noted [5,53] that the optimal $g^r(z)$ is not necessarily stable. The stable part is obtained for power spectrum estimation by partial fraction expansion. Of course, the new truncated $g_T^r(z)$ is not optimal in the sense that the infinite norm $\| Y(z) - g_T^r \|_\infty$ is not minimized to be σ_{r+1}. But, the corresponding Hankel matrix in the time domain is optimal because it minimizes the spectral norm $\| H - \Lambda^r \|_s$, ($\| X \|_s$ = max *eigenvalue of* $X^T X$), to be σ_{r+1} as any noncausal unstable part of $g^r(z)$ has zero contribution to the elements of the Hankel matrix [5,53]. However, good suboptimality is obtained. A constant term can be added to control the error $\| Y(z) - g_T^{r(r)} \|_\infty$ [64].

The algrithm can summarized as follows :

1. Form the Hankel matrix H_N given in (3.3)
2. Find the eigenvalues of H_N by $|H_N - \lambda I| = 0$
3. Determine the model order (r)
4. Find the eigenvector M associated with the eigenvalue λ_{r+1}, where $|\lambda_{r+1}| = \sigma_{r+1}$ and the singular values of H_N are ordered in the descending manner.
5. Form the transfer function $g^r(z)$given in (3.2)

6. Perform the partial fraction expansion on $g^r(z)$ to to get $g^r_7(z)$

7. Determine the power spectrum $P_y(f) = |g^r_7(z)|^2$

3.2 SENSITIVITY STUDIES

The optimal Hankel Transfer function can be written in terms of state-space parameters as follows [5]:

$$g^r(z) = Y(Z) - \sigma_{r+1} \frac{c(zI-A)^{-1}e_{r+1}}{e^T_{r+1}(I-zA)^{-1}} b \tag{3.4}$$

where,

$$Y(Z) = c(zI-A)^{-1}b \tag{3.5}$$

σ_{r+1} is the $(r+1)^{th}$ singular value, e_{r+1} is the $(r+1)^{th}$ column of the identity matrix. A,b, and c are balanced realizations and can be otained by starting with equation (2.46c) and following the same procedure explained in section(2.2.4).

Derivatives with respect to parameters:

i) with respect to b_k's where $b = [\, b_1 \; b_2 \dots b_r]^T$

It can be shown that :

$$\frac{\partial g^r(z)}{\partial b_k} = Z^T O e_k + \frac{1}{\sigma_{r+1}} \frac{e^T_{r+1}(zI-A)^{-1}e_k}{c(I-zA)^{-1}e_{r+1}} [g^r(z) - Y(z)]^2 \tag{3.6}$$

where,

$$Z = [z^{-1} \; z^{-2} \dots]^T$$

$$\text{and } O = \begin{bmatrix} c \\ cA \\ cA^2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

O is known in the litreture as the observability matrix.

ii) with respect to c_i's where $c = [\, c_1 \; c_2 \dots c_r]$

It can be shown that :

$$\frac{\partial g^r(z)}{\partial c_i} = e_i^T CZ - \sigma_{r+1} \frac{e_i^T (zI-A)^{-1} c_{r+1}}{e_{r+1}^T (I-zA)^{-1} b} \tag{3.7}$$

where C is the controlability matrix.

$$C = [\, b \; bA \; bA^2 \dots]$$

iii) with respect to a_{kl} where

$$A = \begin{bmatrix} a_{11} a_{12} & a_{1r} \\ a_{21} a_{22} & a_{2r} \\ \cdot \;\; \cdot & \dots \\ \cdot \;\; \cdot & \dots \\ \cdot \;\; \cdot & \\ a_{r1} a_{r2} & a_{rr} \end{bmatrix}$$

It can be shown that :

$$\frac{\partial g^r(z)}{\partial c_l} = Z^T O_k C_l Z - \sigma_{r+1} \frac{e^T{}_l(zI-A)^{-1} e_{r+1}}{e^T_{r+1}(I-zA)^{-1} b} Z^T O_k -$$

$$\frac{1}{\sigma_{r+1}} \frac{e^T{}_{r+1}(zI-A)^{-1} e_k}{c(I-zA)^{-1} e_{r+1}} [g^r(z) - Y(z)]^2 zC_l Z \qquad (3.8)$$

where

$$O_k = Oe_k \text{ and}$$

$$C_l = e^T{}_l C$$

The observability Grammian is defined as :

$$W^o = O^T O = \sum_{i=0}^{\infty} (A^T)^i c^T c A^i \qquad (3.9)$$

Similarly, the controllability Grammian is defined as :

$$W^c = CC^T = \sum_{i=0}^{\infty} A^i b b^T (A^T)^i \qquad (3.10)$$

The overall sensitivity is found by combining the sensitivities of the parameters. The parameter sensitivity based on p-norm is defined as [7,9]:

$$||H(z)||_p^p = \frac{1}{2\pi} \int_0^{2\pi} |H(e^{j\omega})|^p d\omega \qquad (3.11)$$

and the overall sensitivity is defined as :

$$S_p = \sum_{k=1}^{r} ||\frac{\partial g^r(z)}{\partial b_k}||_p^2 + \sum_{l=1}^{r} ||\frac{\partial g^r(z)}{\partial c_l}||_p^2 + \sum_{k=1}^{r}\sum_{l=1}^{r} ||\frac{\partial g^r(z)}{\partial a_{kl}}||_p^2 \qquad (3.12)$$

Here, we will study the sensitivity based on l_1-norm and l_2-norm which have been extensively used by Rao.

Using the l_1-norm criterion, it can be shown by using Holder's inequality and Cauchy-Schwartz inequality that the individual sensitivities are bounded . Therefore, the l_1- norm of the deviations with respect to b_k is bounded as follows:

$$\|\frac{\partial g^r(z)}{\partial b_k}\|_1^2 \le W_{kk}^o + \sigma_{r+1}^2 \tag{3.13}$$

where

$$W_{kk}^o = O_k^T O_k = \|O_k\|_2^2$$

Also, the l_1- norm of the deviations with respect to c_l is bounded as follows:

$$\|\frac{\partial g^r(z)}{\partial c_l}\|_1^2 \le W_{ll}^c + \sigma_{r+1}^2 \tag{3.14}$$

where

$$W_{ll}^c = C_l C_l^T = \|C_l\|_2^2$$

Also, the l_1- norm of the deviations with respect to a_{kl} is bounded as follows:

$$\|\frac{\partial g^r(z)}{\partial a_{kl}}\|_1^2 \le W_{ll}^c W_{ll}^o + \sigma_{r+1}^2 (W_{kk}^o + W_{ll}^c) \tag{3.15}$$

Substituting equations (3.13),(3.14), and (3.15) in (3.12), the overall sensitivity will be bounded by :

$$S_1 \le (1 + r\sigma_{r+1}^2)[\sum_{k=1}^{r} W_{kk}^o + \sum_{l=1}^{r} W_{ll}^c] + 2r\sigma_{r+1}^2 + \sum_{k=1}^{r} W_{kk}^o \sum_{l=1}^{r} W_{ll}^c = \Omega \qquad (3.16)$$

where Ω can be minimized for the balanced systems of coordinates by using the following results [9]:

$$\sum_{k=1}^{r} W_{kk}^o \sum_{l=1}^{r} W_{ll}^c \ge (r\Lambda)^2 \qquad (3.17a)$$

$$\sum_{k=1}^{r} W_{kk}^o + \sum_{l=1}^{r} W_{ll}^c \ge 2r\Lambda \qquad (3.17b)$$

As a result, it can be shown that Ω has a lower bound as follows:

$$\Omega \ge (1 + r\sigma_{r+1}^2)(2r\Lambda) + 2r\sigma_{r+1}^2 + (r\Lambda)^2 \qquad (3.18)$$

where

$$\Lambda = \frac{1}{r}\sum_{i=1}^{r} \sigma_i$$

This will lead to bound the overall sensitivity by :

$$S_{Hankel} \le (1 + r\sigma_{r+1}^2)(2r\Lambda) + 2r\sigma_{r+1}^2 + (r\Lambda)^2 \qquad (3.19)$$

From [9] the balanced sensitivity is found to be bounded as follows:

$$S_{balanced} \le 2r\Lambda + (r\Lambda)^2 \qquad (3.20)$$

If σ_{r+1} is small enough the S_{Hankel} will be reduced to the $S_{balanced}$.

Using the l_2-norm, it can be shown that [7]:

$$S_{Hankel} \le (1 + r\sigma_{r+1}^2)(2r\Lambda) + 2r\sigma_{r+1}^2 + (r\Lambda)^2 + 2(r\Lambda)^2 \sum_{i=1}^{\infty} \|A^i\|^2 \qquad (3.21)$$

where the balanced sensitivity is given as [9]:

$$S_{balanced} \le 2r\Lambda + (r\Lambda)^2 + 2(r\Lambda)^2 \sum_{i=1}^{\infty} ||A^i||^2 \qquad (3.22)$$

If σ_{r+1} is small enough the S_{Hankel} will be reduced to the $S_{balanced}$.

3.3 MODEL ORDER SELCTION

Since the best choice of the model order (r) is not generally known a priori, it is usually necessary to assume several model orders . Too high a guess for model order adds spurious detail into the spectrum. Too low an order results in a highly smoothed spectrum [29] .

Different criteria have been developed to achieve the selection of the AR model order . Akaike has introduced two criteria . The first one is the final prediction error (FPE) . This criterion chooses the order so that the average error for a one step prediction is minimized. It is defined as the mean square prediction error . Or, alternatively , if for a given AR process we define $\hat{y}_k$ as the estimated prediction of y_k, then

$$FPE = E\{ (y_k - \hat{y}_k)^2 \} \qquad (3.23)$$

which reduces to :

$$FPE_r = \varepsilon_r^2 \left(\frac{N + r + 1}{N - r - 1} \right) \qquad (3.24)$$

Details for the derivation of equation (3.24) have been presented by Akaike and Ulrych and Bishop [57] . Where N is the number of samples, r is the model order, and ε_r^2 is the prediction error power . Notice that from equation (3.24), ε_r^2, generally, decreases as the assumed model order (r) is increased, whereas the term in parentheses increases . Therefore, the desired r will give the minimum value of FPE . The FPE has been extensively studied for several applications [29 , 57] . FPE works well for AR process . However, it tends to give too low orders when actual geophysical data is processed .

The second criterion proposed by Akaike is known Akaike information criterion (AIC) . It is based on the minimization of the log likelihood of the prediction-error variance as a function of the filter order r . Akaike used maximum likelihood approach to derive the AIC . The AIC estimates the model order by minimizing an information theoretic function . The AIC is.

$$AIC_r = \ln(\varepsilon_r^2) + \frac{2}{N}(r+1) \tag{3.25}$$

The derivation of equation (3.25) based on the assumption that the process has Gaussian statistics . The term (r + 1) can be replaced in some cases by (r), since $\frac{2}{N}$ is only an additive constant . The term ln (ε_r^2) indicates the penalty for using extra AR coefficients that do not result in a substantial reduction in the prediction error power [29] . Again, the order r selected is the one that the AIC is minimized .

Parzen proposed another criterion in selecting the model order . It is known as the criterion autoregressive transfer (CAT) function . The order r is chosen to be the one that minimizes the differance in the mean - square errors between the true prediction error filter and the estimated filter . The CAT is,

$$CAT_r = \frac{1}{N}\sum_{i=1}^{r}\frac{1}{\hat{\varepsilon}_i^2} - \frac{1}{\hat{\varepsilon}_r^2} \tag{3.26a}$$

where,

$$\hat{\varepsilon}_i^2 = \left(\frac{N}{N-i}\right)\varepsilon_i^2 \tag{3.26b}$$

Again, the order r is chosen to give the minimum CAT_r. For short data records, none of the aforementioned criteria work well in estimating the appropriate model order [29]. Also, for harmonic processes in noise, the FPE and the AIC underestimate the model order if the signal to noise ratio is high [65] . As a solution in selecting the order for the case of short data recodrs, Ulyrch and Ooe suggest that an order selection between $\frac{N}{3}$ and $\frac{N}{2}$ often give reasonable results.

Another effective criterion in model order selection, especially when few samples are available, is based on singular value decomposition (SVD) technique . It is effective in the presence of roundoff errors of noisy data . It has been applied to spectral estimation because it provides an efficient method in the dtermination of the effective rank (model order) . It operates on the formed data matrix . Theoretically, the rank of the matrix is the total number of nonzero SV's . Practically, the observed data matrix consists of the data matrix perturbed by noise .

The determination of the effective rank ($r \leq N$) of the observed data matrix by using its SV's has been taken into account previously based on different criteria [31] .

1. Criterion 1 : $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \beta_1 \geq \sigma_{r+1} \geq \ldots \geq \sigma_N$

2. Criterion 2 : $\frac{\sigma_r}{\sigma_1} > \beta_2 > \frac{\sigma_{r+1}}{\sigma_1}$

3. Criterion 3 : $\sigma_r >> \sigma_{r+1}$

4. Criterion 4 : $\sigma_{r+1}^2 + \sigma_{r+2}^2 + \ldots + \sigma_N^2 < \beta_4$

5. Criterion 5 : $d_5(r) = \left[\frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2}{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_N^2} \right]^{½} > \beta_5$

The threshold values of β_2 , β_4 , and β_5 are not based on any explicit analytical expressions . For a specific finite precision roundoff model, an analytical formula for β_1 was proposed by Golub and Van Loan [66] . Konstantinides and Yao [31] derived confidence regions for the perturbed singular values of matrices with noisy observation data . The analysis was based on the theories of perturbations of SV's and statistical significance test . In random model, the threshold bounds rely on the dimension of the data matrix, the noisy variance, and a predefined level of significance [31] . They derived analytically the upper and lower bounds on β_1 .

Clearly, the ratio $d_5(r)$ approaches its maximum value of one as r approaches N . Therefore, for matrices of low effective rank, the ratio $d_5(r)$ is near to one for values of r significantly smaller than N . Whereas, the matrices of high effective rank, the ratio $d_5(r)$ is close to one for higher values of r (i.e. $r \approx N$).

CHAPTER 4

SIMULATION OF THE SPECTRUM ESTIMATION METHODS

The algorithm proposed in chapter three will be applied for different examples . The comparison is done against most of the methods described before . In order to make a complete study , three different aspects have been taken into account . These are :

1. The signal to noise ratio (SNR)
2. The model order (r)
3. Number of data samples (N)

In testing all methods , thirty statistically independent realizations each of length N were generated . These thirty realiztions are used to compare modeling effectiveness of optimal Hankel algorithm (OHM) with the other methods .

4.1 TWO AR MODELS IN WHITE GAUSSIAN NOISE

In this example , we will examine the time series as characterized by [13] :

$$y_n = x_n + z_n + \varepsilon_e e_n$$

which is a composition of two AR(2) time series generated according to :

$$x_n = 0.4\, x_{n-1} - 0.93\, x_{n-2} + \varepsilon x_n$$

$$z_n = -0.5\, z_{n-1} - 0.93\, x_{n-2} + ez_n$$

where e_n , ex_n , and ez_n are mutually uncorrelated white Gaussian processes with zero mean and variance one . And , ε_e is the scaling constant for varying the signal to noise ratios (SNR) . The exact power spectral density is shown in Fig # 1 .

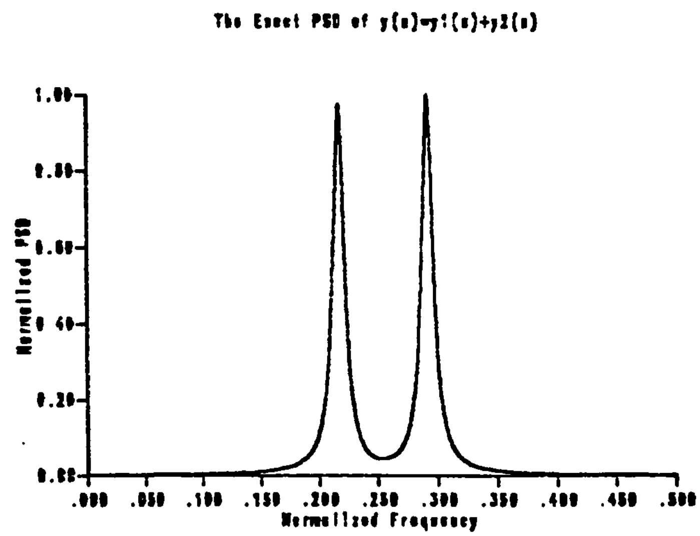

Figure 1: The Exact PSD of the Two AR Models

For short data record , the traditional methods fail to give good approximation . The periodograms with different window lengths , different number of samples , and different FFT lengths are shown in Fig # 2 . Among all trials in varying the aforementioned variants , the relatively best estimation is given with 64 window length and 128 samples . Fig # 3 illustrates the Blackman - Tukey spectrum . When the number of autocorrelation lags is small the spectrum will not show the exact peaks . Increasing the number of both samples and lags will improve the results .

The AR PSD's based on Burg's algorithm are shown in Fig # 4 . For higher model orders , and large number of samples , these estimated power spectral densities give good results , even for low signal to noise ratios . As . the number of samples decreases the performance of the method decreases , even for higher model orders and high SNR's .

The ARMA modeling via Cadzow's algorithm is one of the used modern techniques . The estimated PSD's are shown in Fig # 5 . Cadzow's method fails to give a good estimate , when the number of samples equals 32 . As the number of samples increases , the estimates are improved depending on different ARMA model orders . However , this is a big disadvantage , becuase chere is no solid critrion to judge the appropriate orders .

The suboptimal Hankel method (SHM) behaves in a manner that is slightly better than the the previous methods . As the model order increases with respect to a fixed SNR , the power spectral densities are estimated more accurately . Grnerally , increasing the number of data samples improves the results . As seen from Fig # 6 , when eight samples are used , SHM fails completely to estimate the power spectrum .

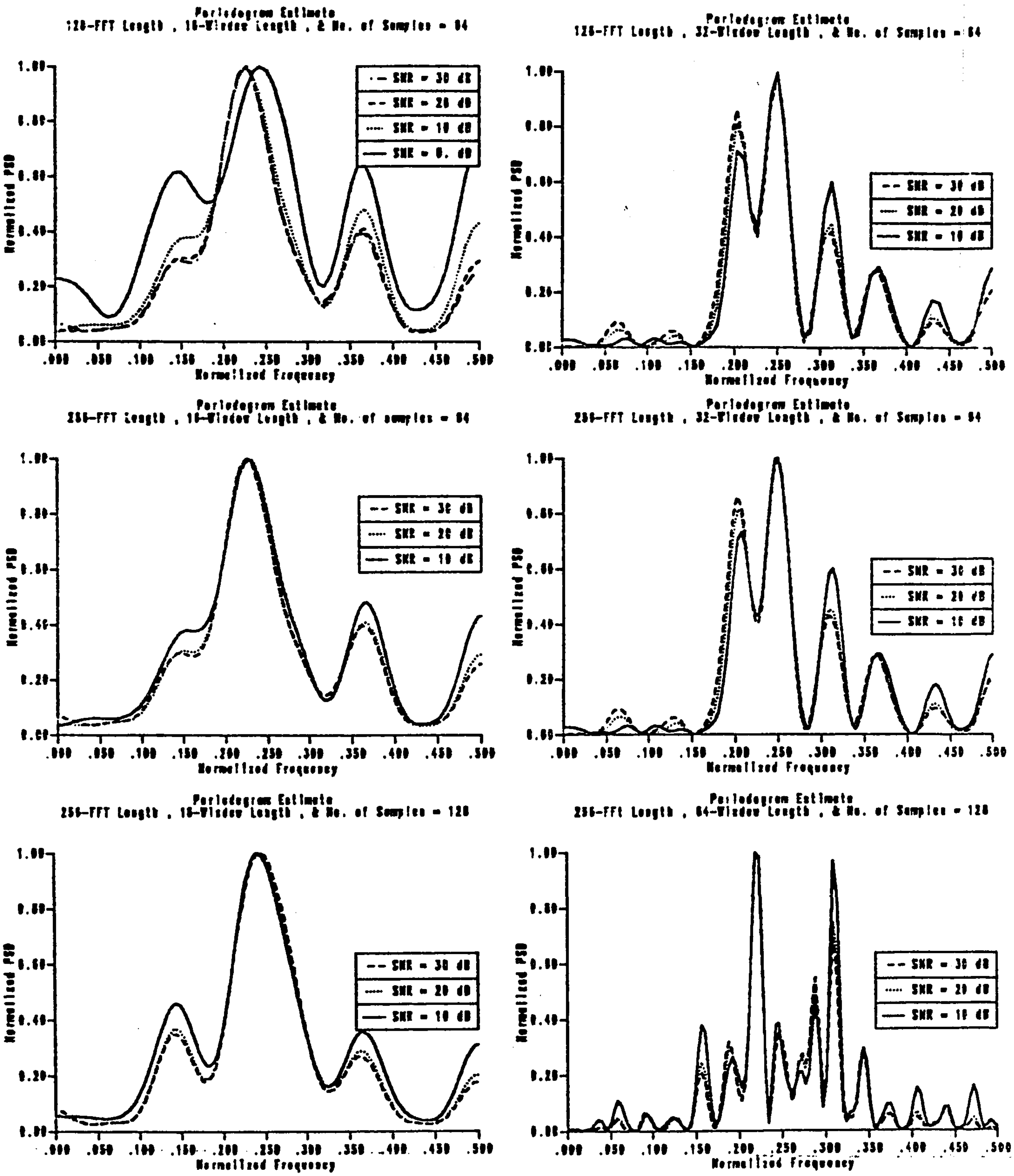

Figure 2: PSD's of two AR models based on Periodogram estimate

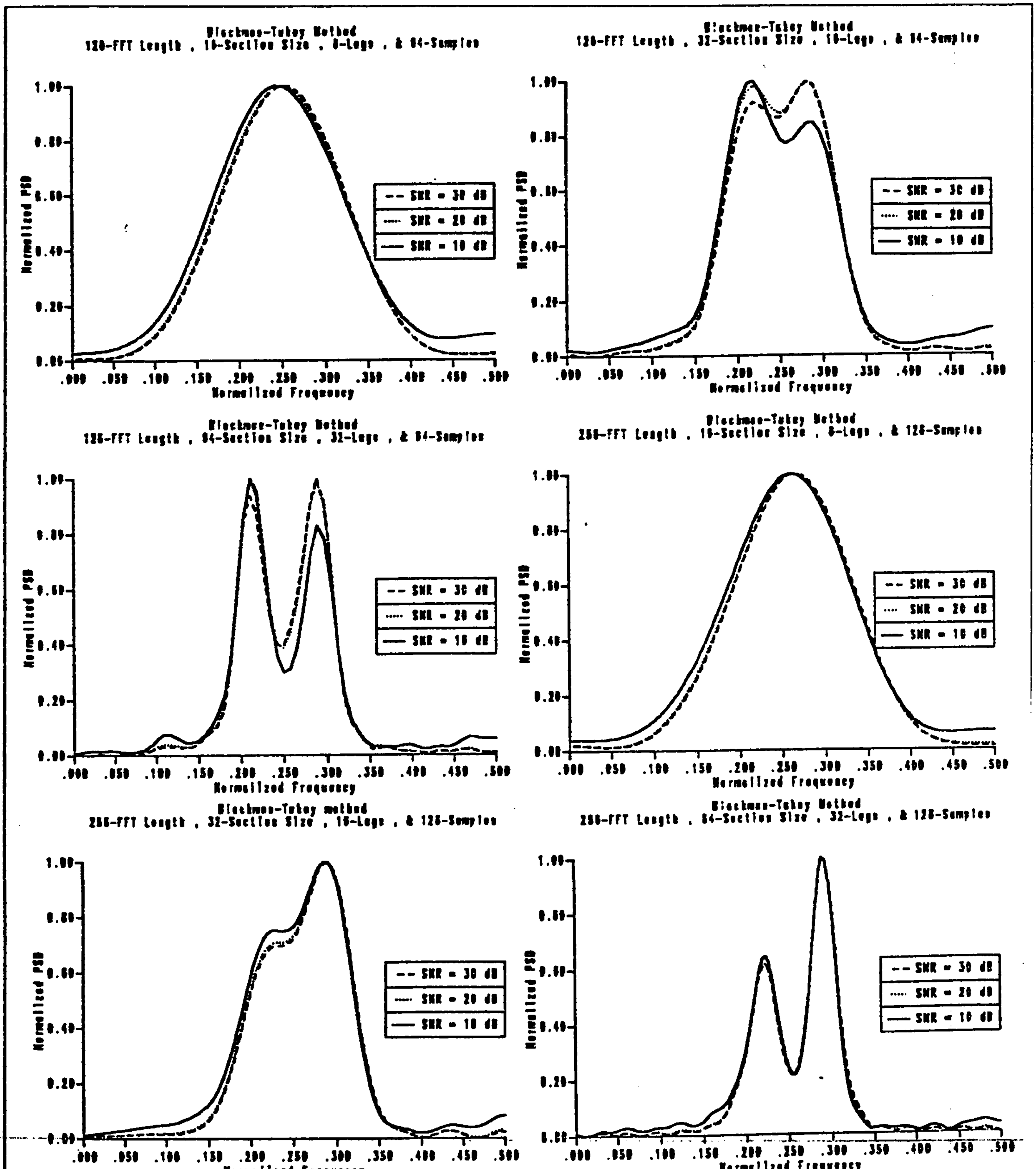

Figure 3: BT PSD's of two AR models

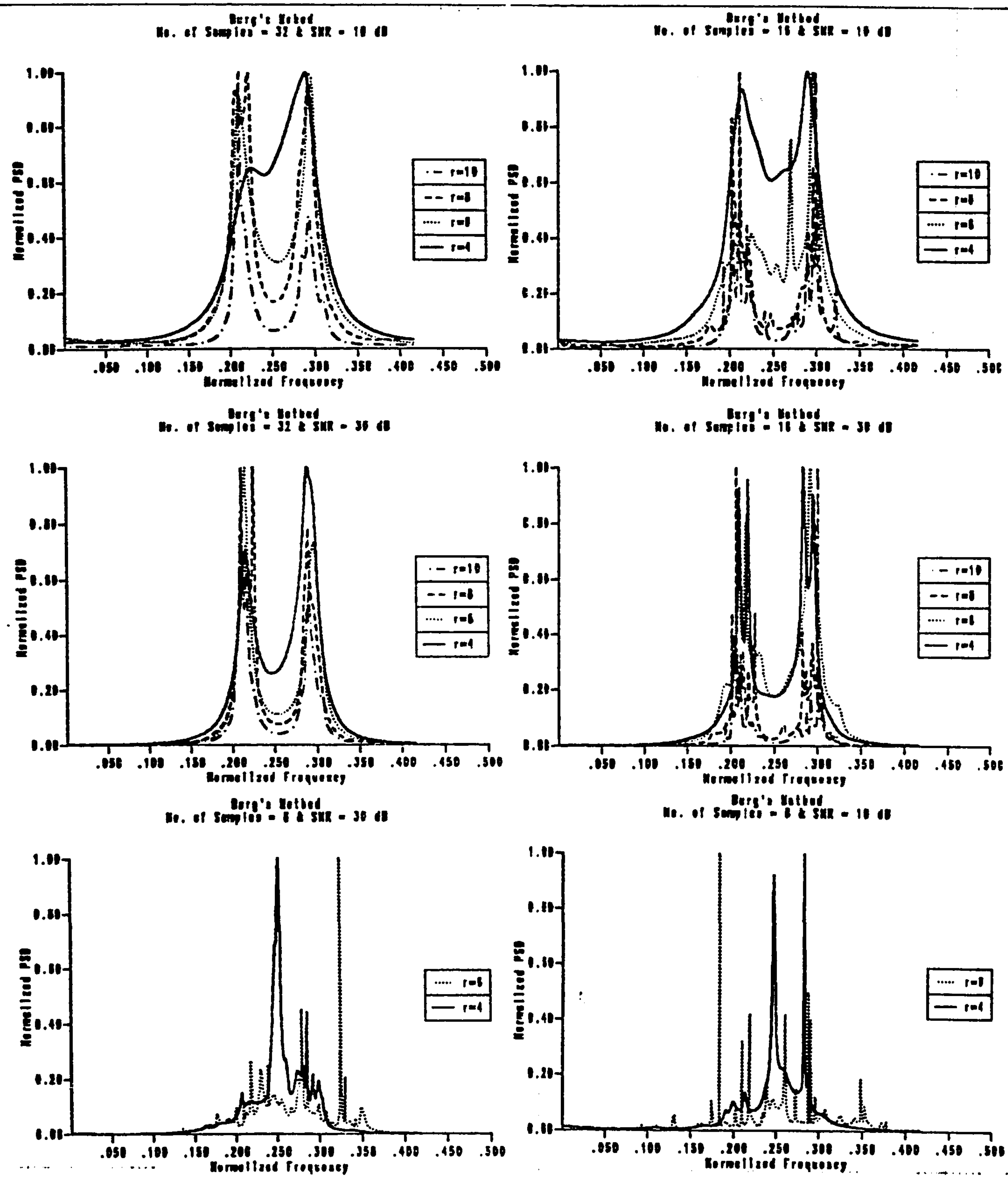

Figure 4: Autoregressive PSD's of two AR models via Burg algorithm

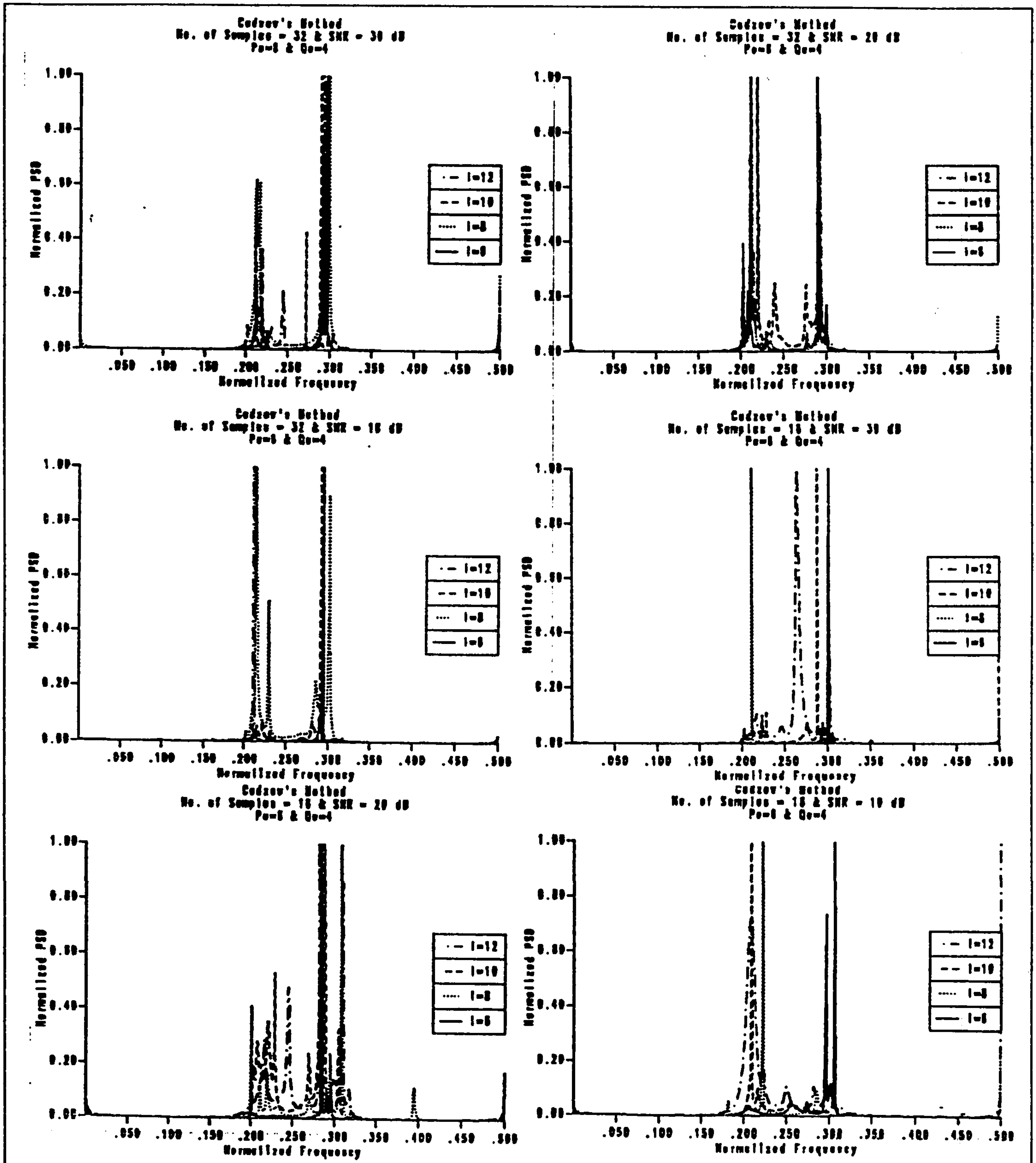

Figure 5: ARMA PSD's of two AR models via Cadzow algorithm

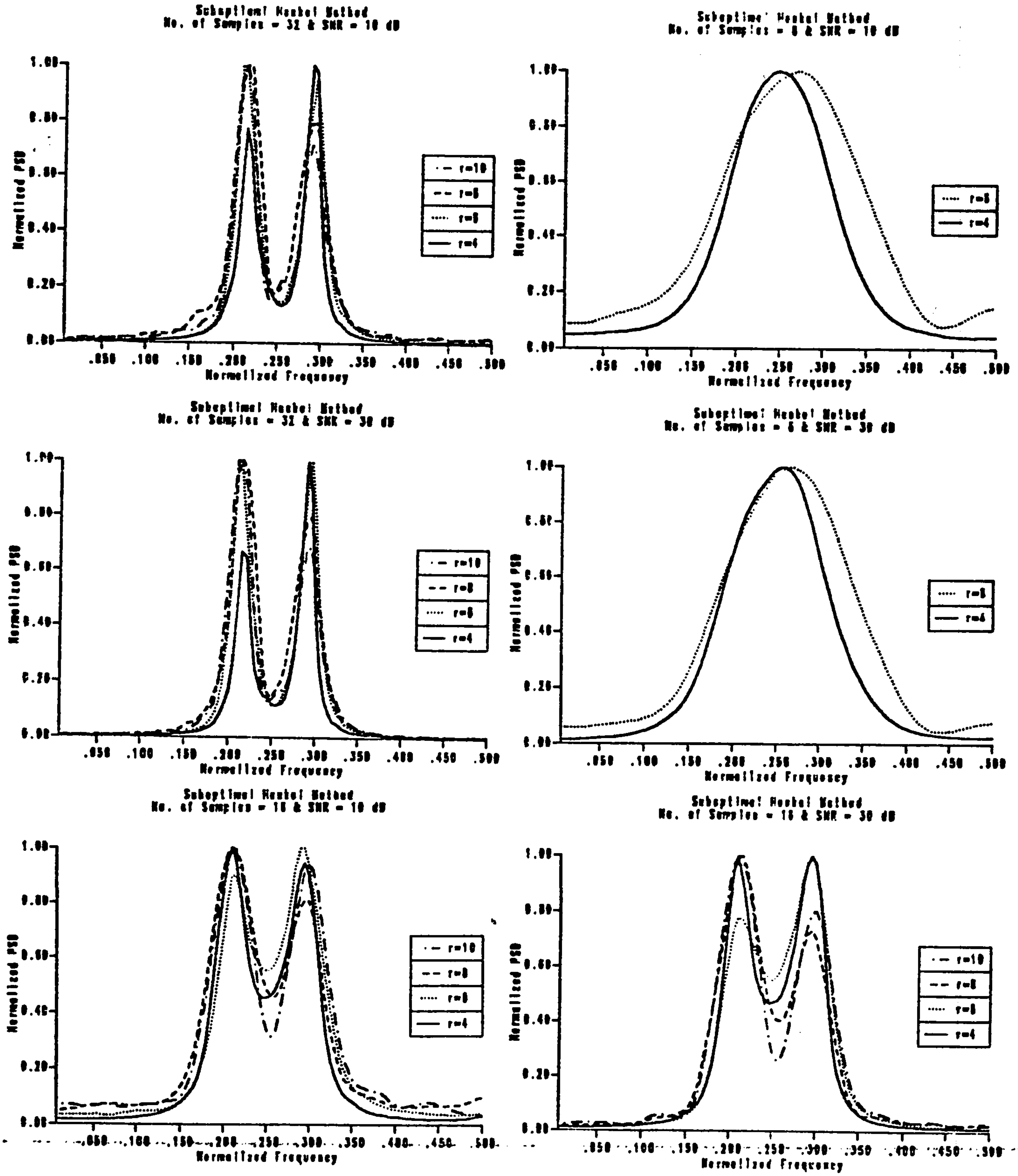

Figure 6: SHM PSD's of two AR models

Fig # 7 shows the estimated PSD's via covariance approximation method . PSD was estimated for different model orders and different number of samples at different SNR's . Most of plots show many peaks at different frequencies but the higher peaks will be at the desired frquencies for high signal to noise ratios and N = 32 and 16 . For N = 8 the method fails to show any response . Notice that , there are negative PSD values which is due to the occurance of spectral density zeros on the unit circle .

MUSIC method of spectral estimation is not very different from the aforementioned methods . The simulation has been done on the basis of different model orders and different number of signal space vectors . As seen from from Fig # 8 , the best estimates are obtained for the largest number of samples and highest SNR's . In addition , increasing the dimension (number) of the signal space vectors improves on the results .

Finally , working with optimal Hankel method (OHM) shows the most reliable and effective results . Fig # 9 shows a clear conclusive advantage of OHM over the previously tested methods . For the case of number of samples , OHM gave the best results of power spectral estimation for N = 32 , and 16 samples . Even for eight data samples , we still have , relatively speaking , acceptable ersults . Similarly , for the case of varying SNR , simulation results of OHM are also considered the most accurate among the others . For the case of model order , the minimum required is r = 4 as previously explained . It is obvious from the figure that the increase in (r) gives a relatively similar performance which is expected . In addition to the previous advantages , OHM transfer function $g^r(z)$ is adequately modeled with the truncated transfer function $g^r_T(z)$ for stable poles .

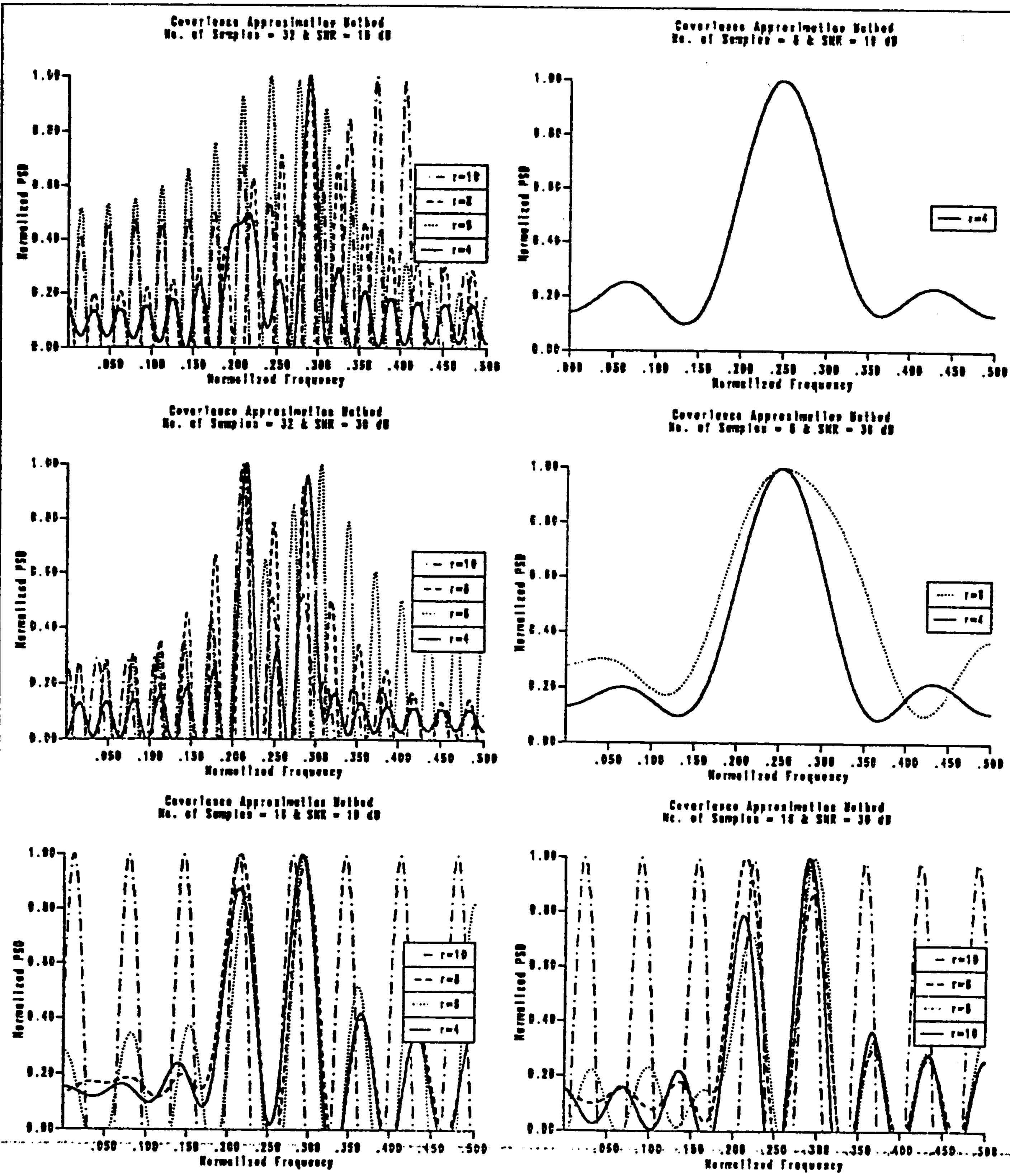

Figure 7: PSD's of two AR models via Covariance Approx. method

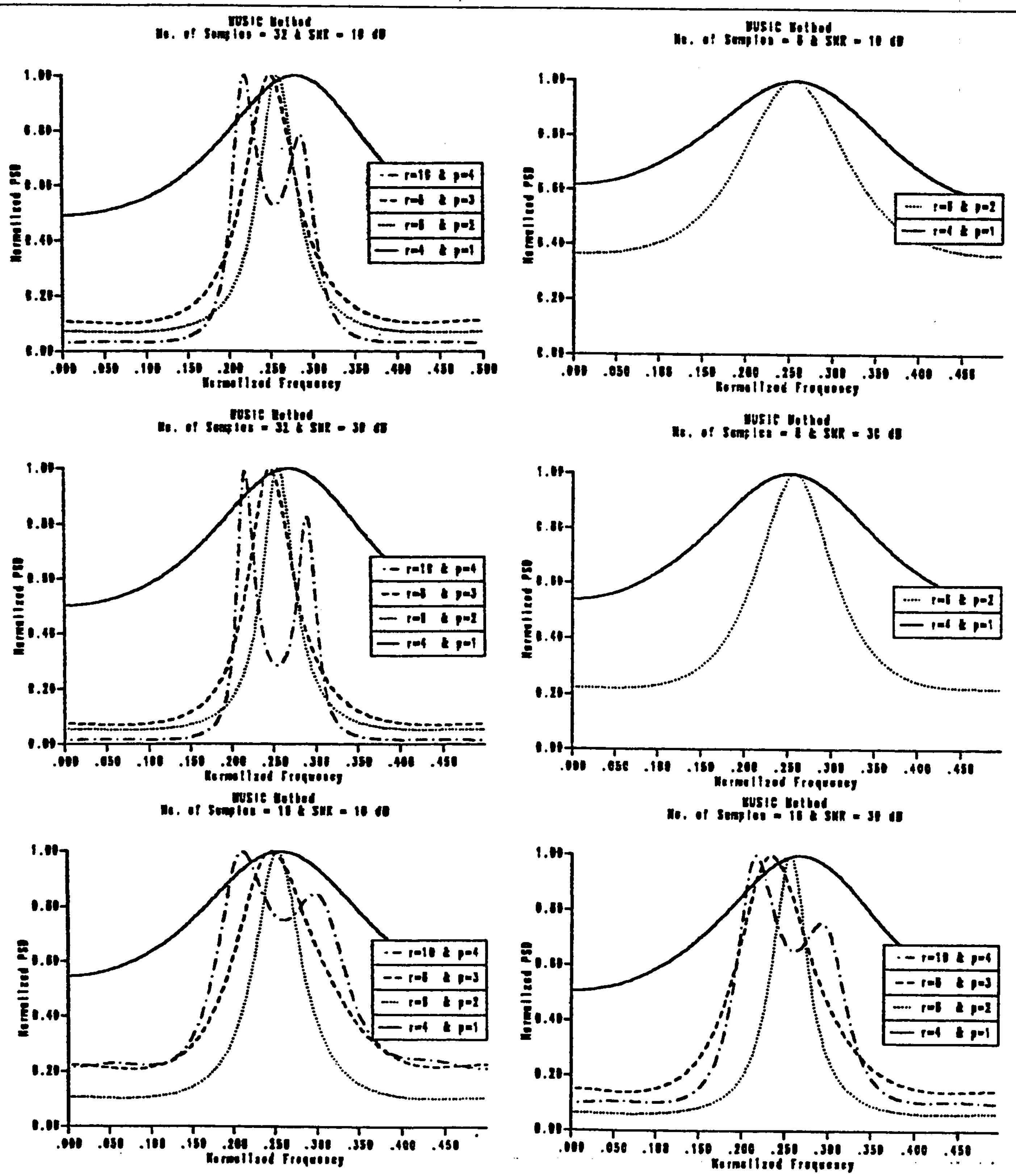

Figure 8: MUSIC estimates of two AR models

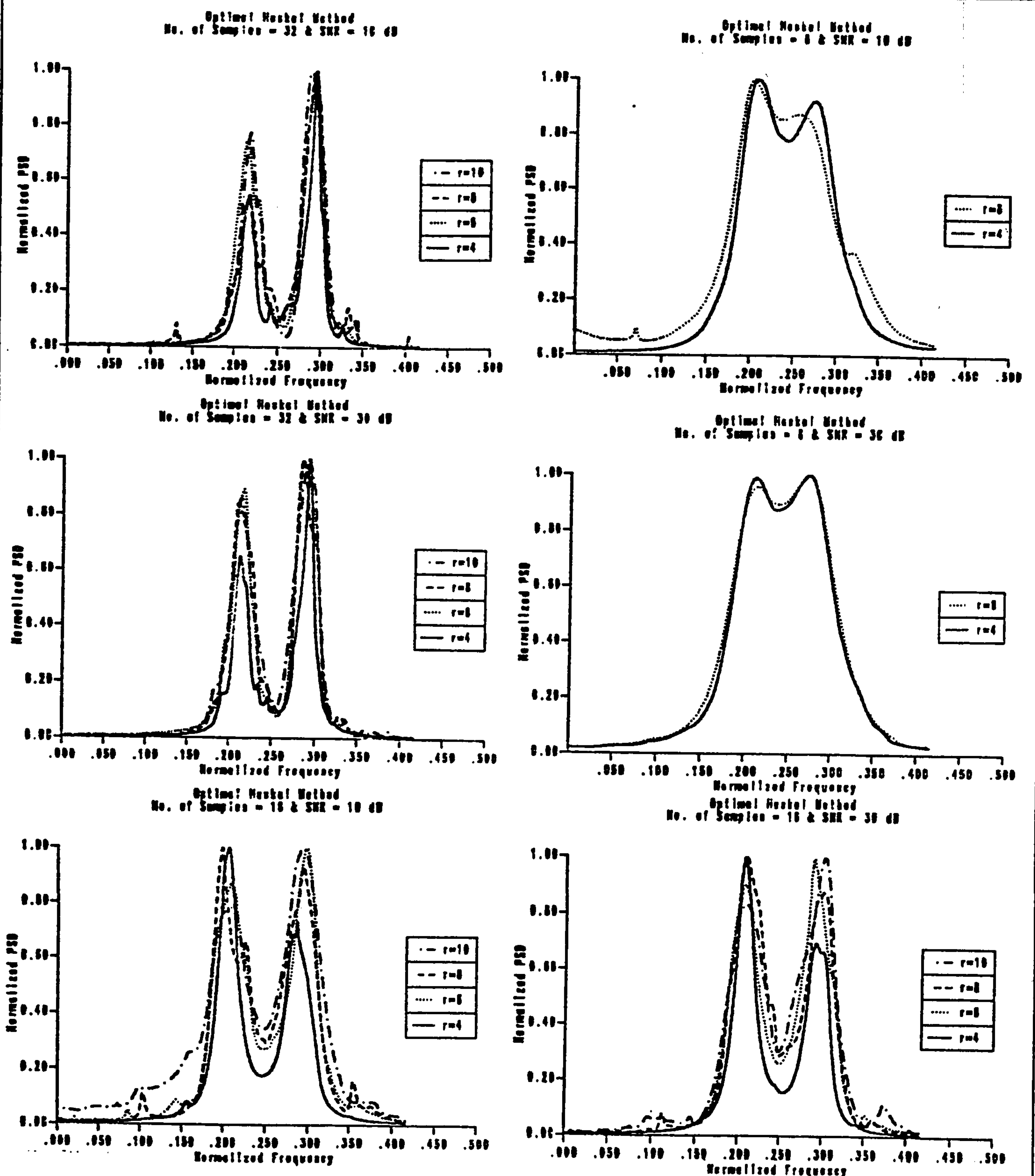

Figure 9: PSD's of two AR models via OHM

Concluding Remmarks

1. When r = 2 , all the methods above failed to estimate PSD
2. When N = 8 , all the methods fail completely , but OHM gives relatively reasonable results
3. $g^r(z)$ without partial fraction expansion gives practically similar results as $g'_T(z)$ considered here .
4. When SNR = 0. dB , all the methods above failed but OHM and SHM gives relatively acceptable results .

As a special case of the first example, a one AR model in white Gaussian noise will be considered.

In this example , we will examine the time series as characterized by [13] :

$$y_n = x_n + \varepsilon_e e_n$$

which is a AR(2) time series generated according to :

$$x_n = 0.4\, x_{n-1} - 0.93\, x_{n-2} + ex_n$$

where e_n , and ex_n are mutually uncorrelated white Gaussian processes with zero mean and variance one . And , ε_e is the scaling constant for varying the signal to noise ratios (SNR) . All simulation results of this example will not be included for lack of space and it is a special case of the first example .

Periodograms with different variants (window length , number of samples , and FFT length) were obtained . They are unable to resolve the right peak except only for large number of samples . Even BT method is not able to show a nice estimate unless the number of samples is large and a reasonably chosen number of autocorrelation lags .

The minimum model order required to estimate the PSD is r = 2 . Regardless of any combination of parameter variations , Cadzow's method fails completely to give any meaningful results unless the number of samples is large enough . Burg's method gives good results for different model orders (r = 2 , 4 , 6 , 8 , and 10) , number of samples (N = 32) , and different signal to noise ratios (SNR = 10 , 20 , and 30 dB) . But , it fails completely for N = 8 samples . For N = 16 , it gives some reasonable results .

The SHM gives very good results at (N = 32 and 16) with the same variants in model orders and SNR's .

Covariance approximation method and MUSIC method give good results only for higher model orders . They show nice results for different number of samples .

The OHM gives better results than all the other methods even for lower number of samples .

4.2 MOVING AVERAGE MODEL PLUS SINUSOIDS

In this example , we will examine the time series plus sinusoids as chacterized by [65] :

$$y_n = x_n + 2\,cos(\frac{11\pi n}{20}) + 2\,cos(\frac{\pi n}{2})$$

where , x_n is of MA(4) time series generated according to :

$$x_n = e_n + 2.\,e_{n-1} - 2\,e_{n-3} - e_{n-4}$$

where , e_n is white Gaussian noise process with zero mean and variance one . The relative exact power spectral density is shown in Fig # 10 .

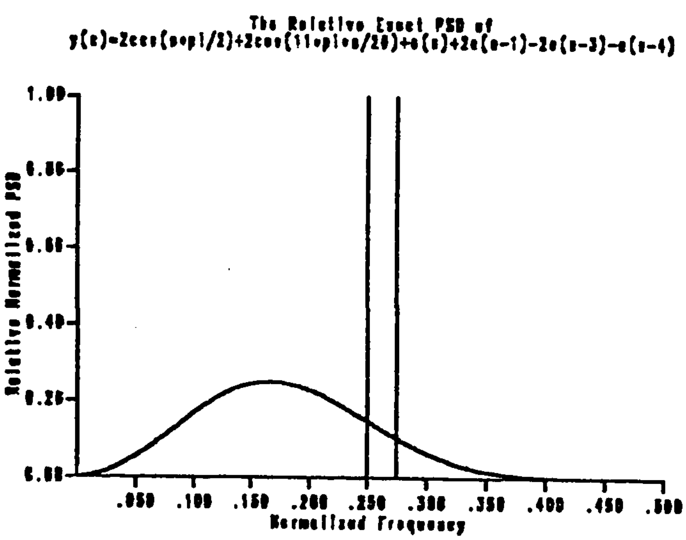

Figure 10: The exact PSD of MA model plus sinusoids

For short data record , the traditional methods fail to give good approximation of PSD . The periodogram with different window lengths , different number of samples , and different FFT lengths are shown in Fig # 11 . The best estimation is given when number of samples = 128 and 256 FFT length . It can resolve two peaks at the desired frequencies . Blackman and Tukey method fails completely to resolve the two peaks even for larger number of both samples and lags as shown in Fig # 12 .

Burg's method gives reasonable results , but it does not show clear peaks at the desired frequencies . When N = 32 samples , it gives a nice estimation only at r = 8 . When N = 16 samples , it gives a reasonable estimation only at r=4. While at N = 8 , it fails completely as shown in Fig # 13 .

Cadzow's method does not give that good results as Burg's method . It fails completely to estimate the right frequencies even at N = 32 samples as shown in Fig # 14 .

Also , SHM is able to resolve the too close frequencies at number of samples N = 32 and different model orders as it is clear from Fig # 15 . But for smaller number of samples SHM does not give results .

Covariance approximation method gives reasonable results . When (N= 32 and 16 samples) , it gives a nice estimation only when r = 4 , while at N=8 it fails completely as shown in Fig # 16 .

MUSIC method does not resolve the too close frequencies even at different number of samples and model orders as clear from Fig # 17 .

OHM gives better results . When N = 32 samples , it gives a nice estimation at (r = 6 , 8 , and 10) . But , when (N = 16 and 8 samples) , it does not resolve the two frequencies as shown in Fig # 18 .

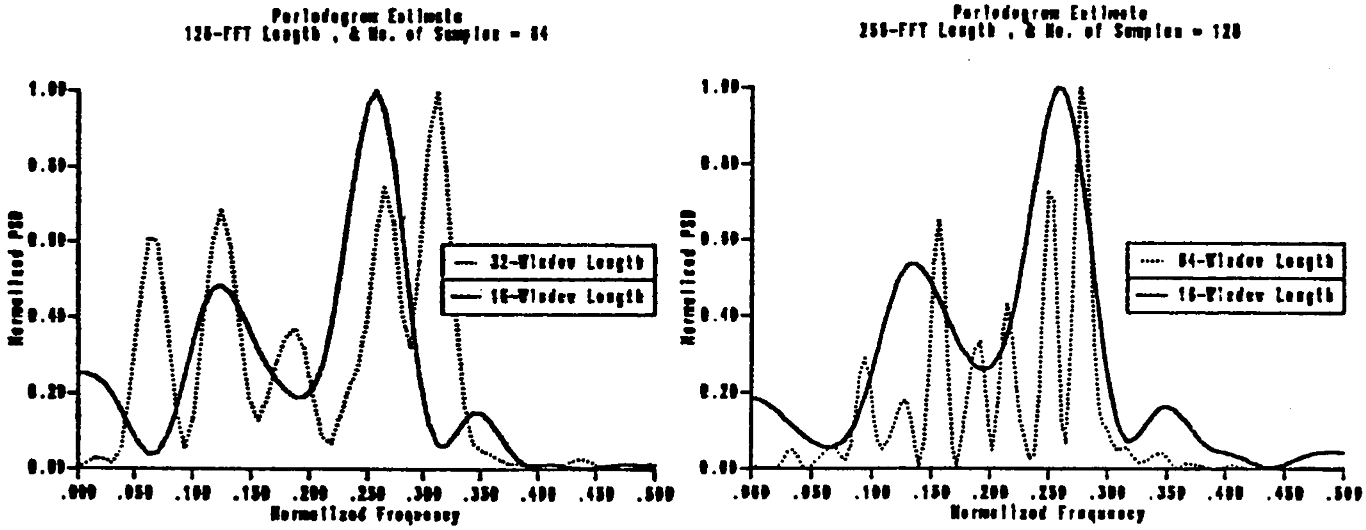

Figure 11: Periodogram PSD's of MA model plus sinusoids

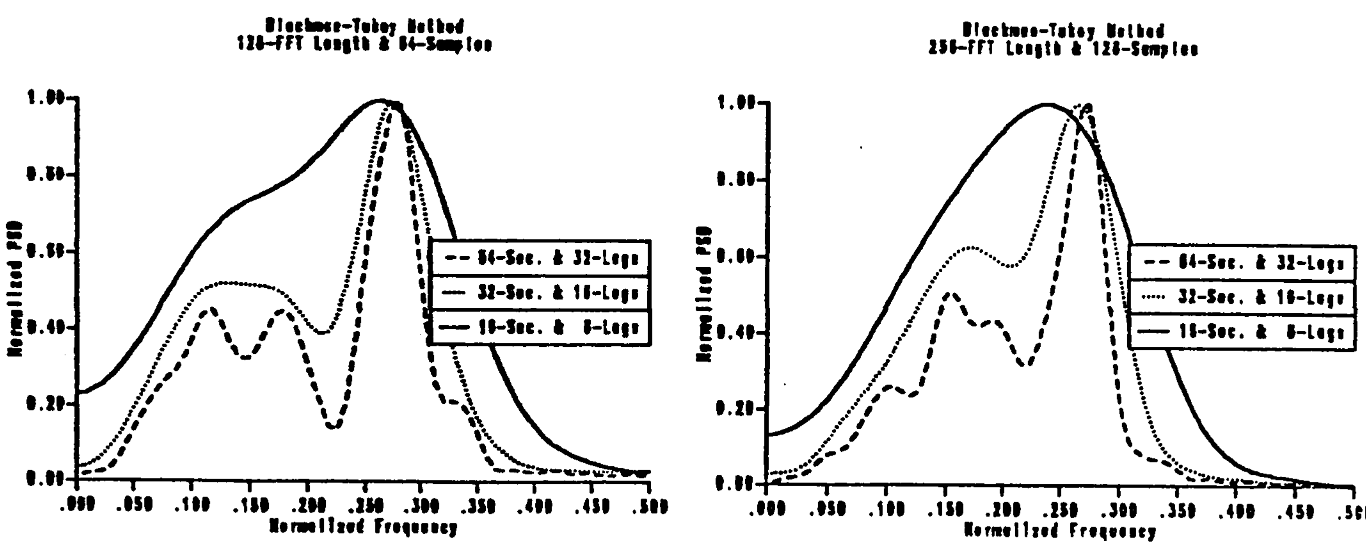

Figure 12: Blackman-Tukey PSD's of MA model plus sinusoids

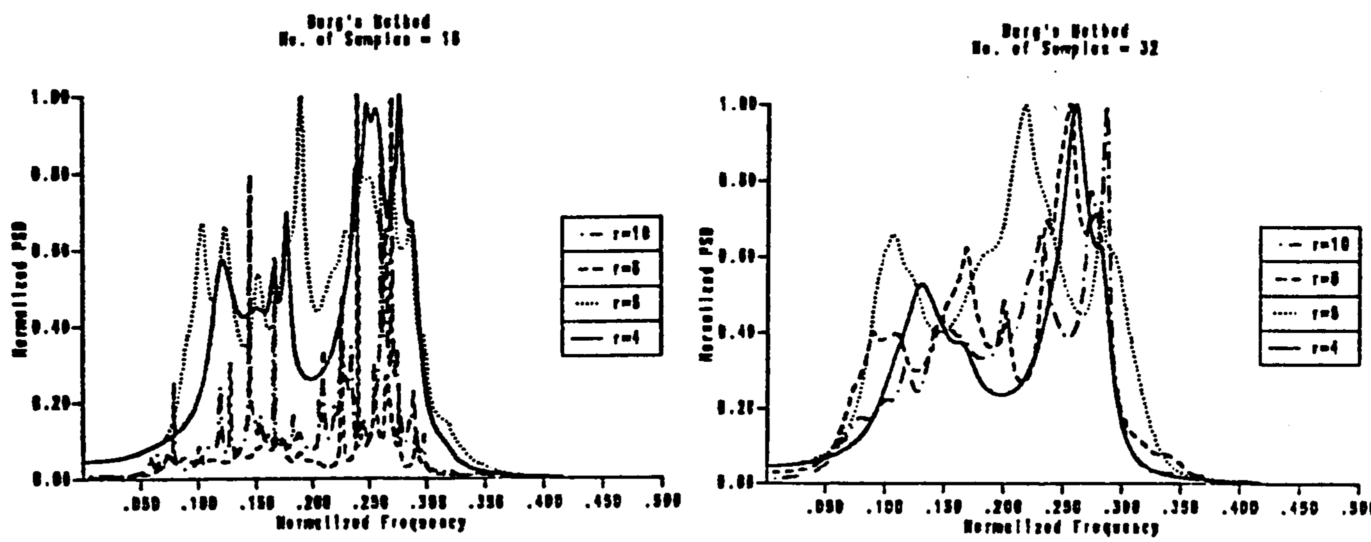

Figure 13: Burg PSD's of MA model plus sinusoids

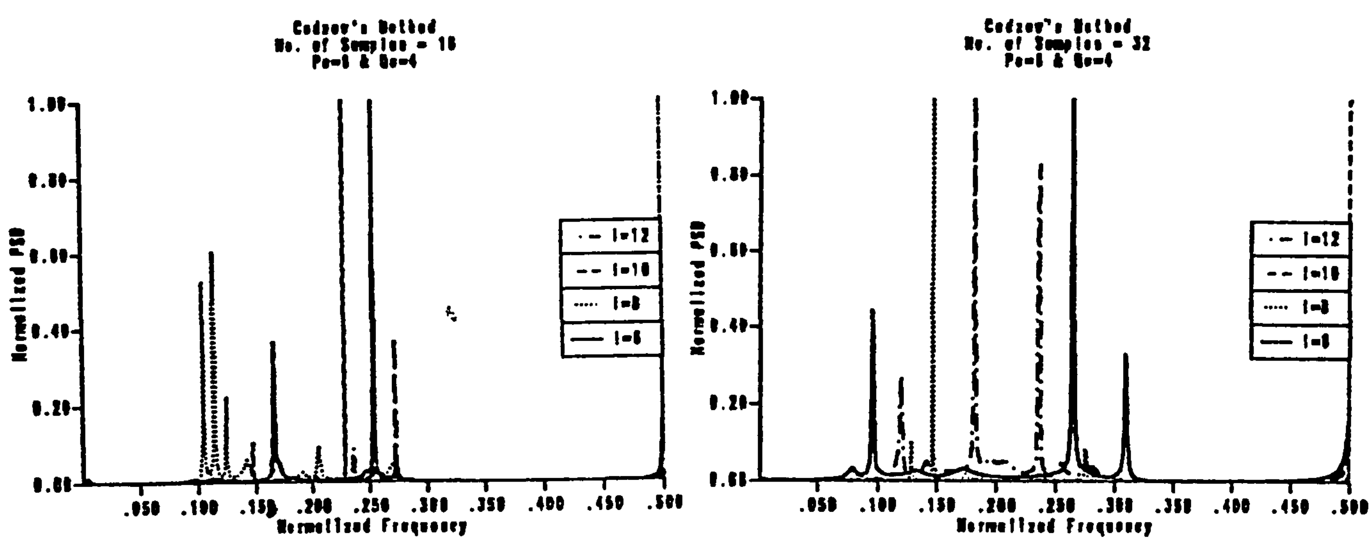

Figure 14: ARMA modeling of MA model plus sinusoids via Cadzow approach

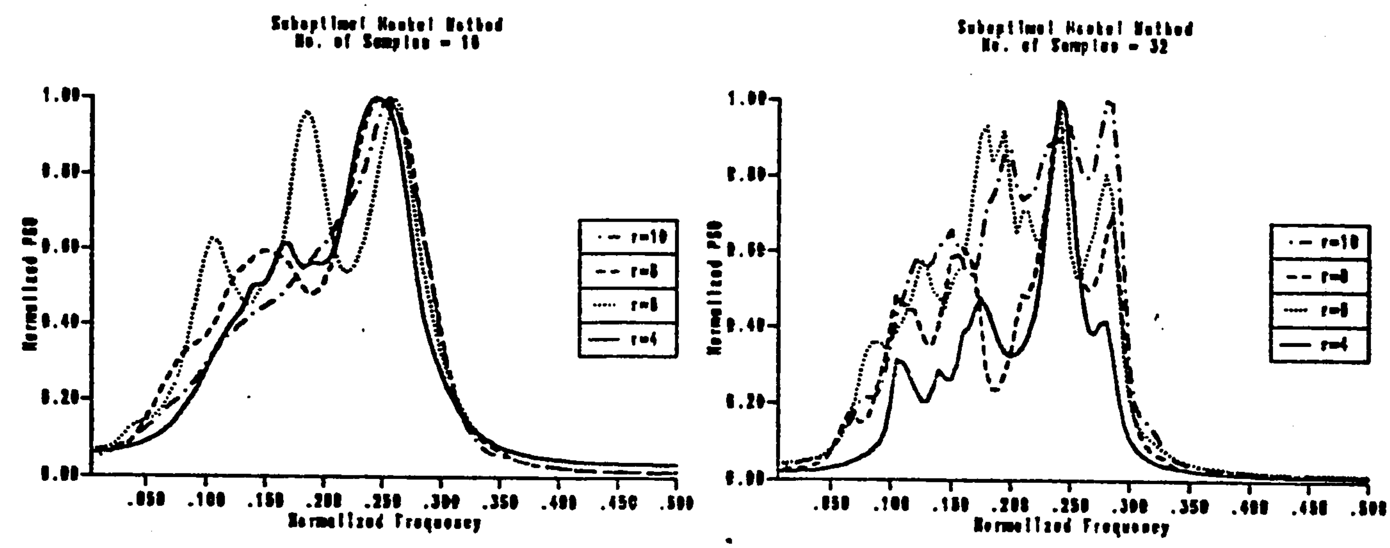

Figure 15: SHM PSD's of MA model plus sinusoids

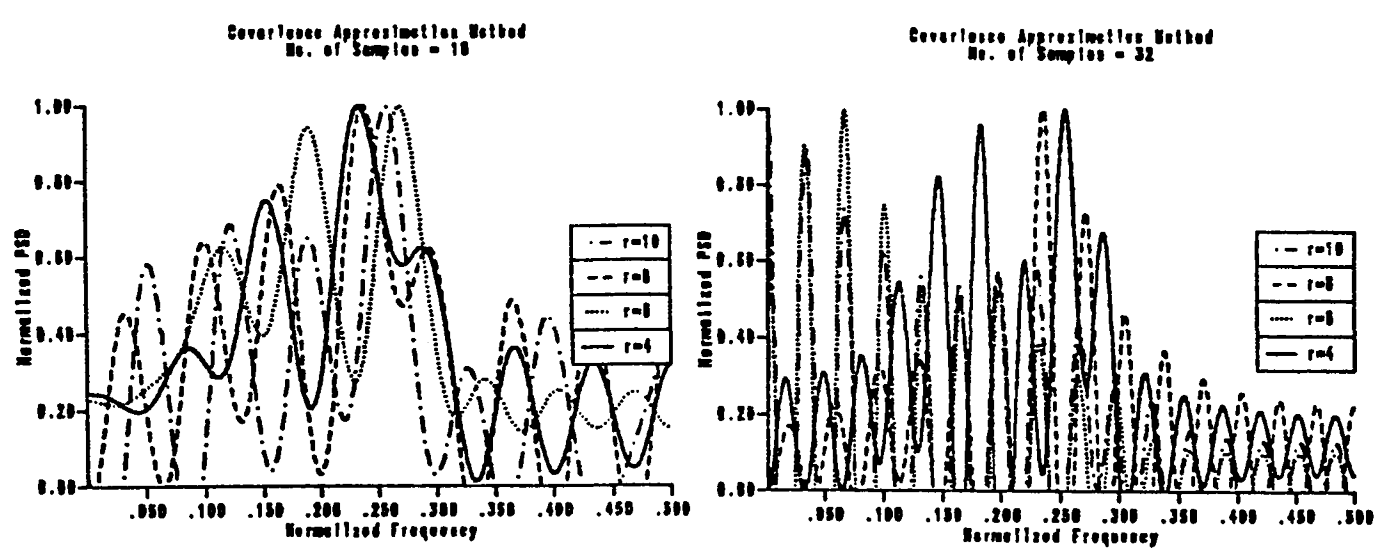

Figure 16: PSD's of MA model plus sinusoids via Covariance Approx. method

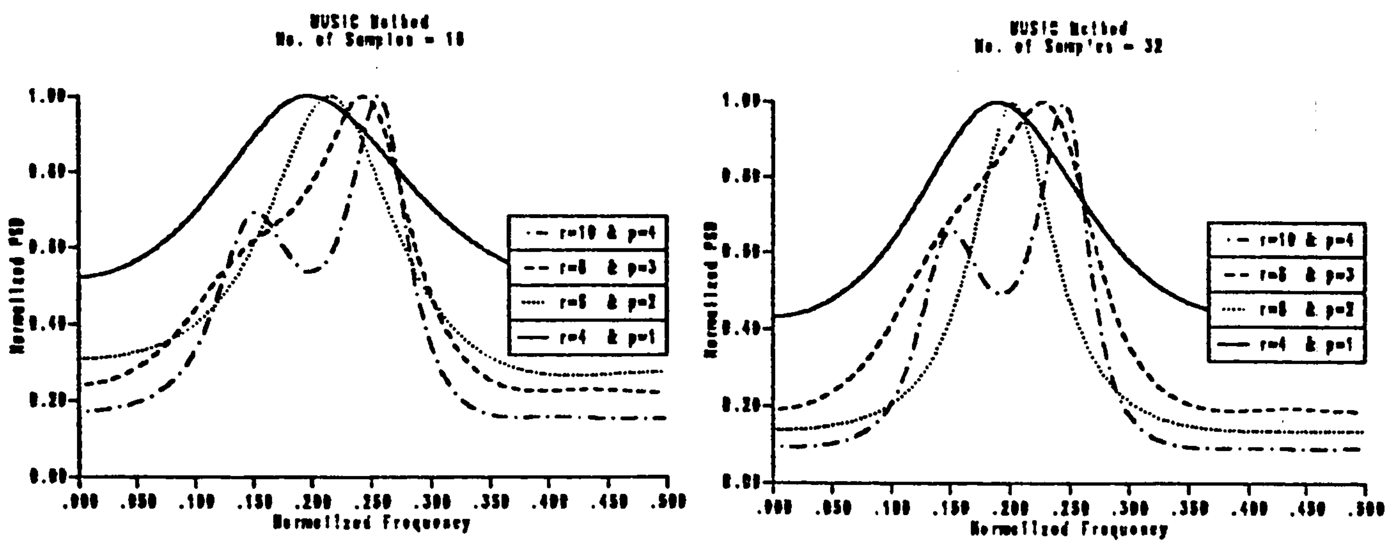

Figure 17: MUSIC PSD's of MA model plus sinusoids

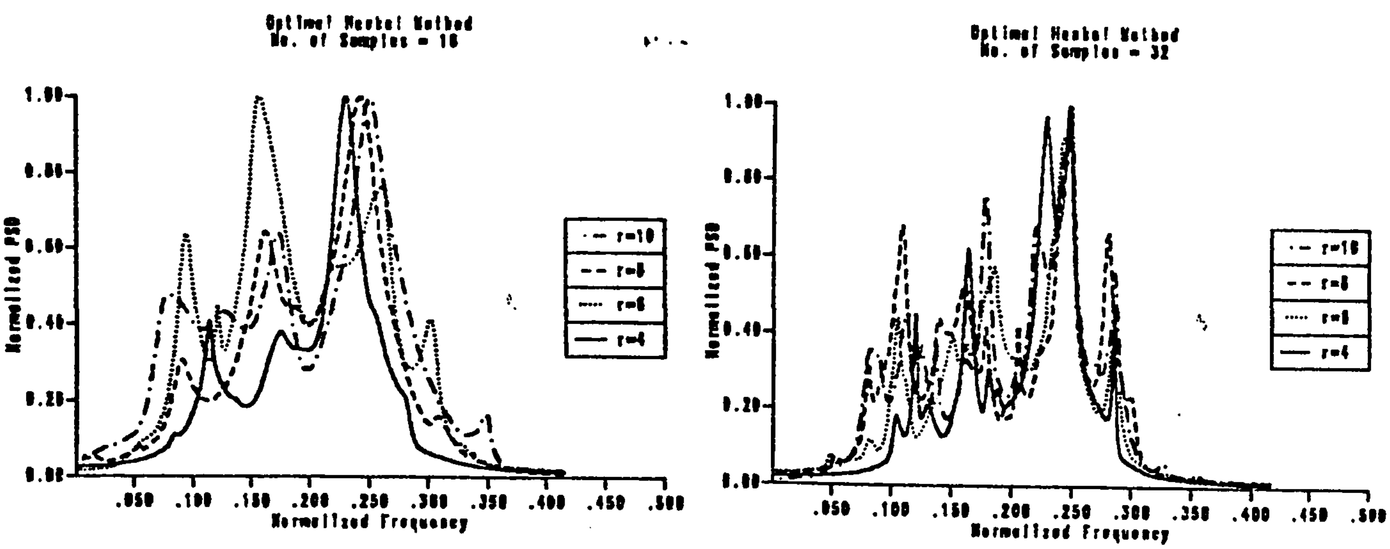

Figure 18: PSD's of MA model plus sinusoids via OHM

As a special case of the second example, a moving average (MA) model will be considered.

In this example , we will examine the time series as chacterized by [65] :

$$y_n = e_n + 2.\, e_{n-1} - 2.\, e_{n-3} - e_{n-4}$$

where , e_n is white Gaussian noise process with zero mean and variance one .

Periodograms with different variants were obtained . They fail to show a nice approximation for the exact PSD . Also , BT method does not show a good approximation .

Burg's method gives reasonable results . When N = 32 samples , it gives a good estimation only at r = 8 . While , at N = 16 , it gives a good approximation at r = 4 . It fails completely at N = 8 .

Cadzow's method does not show any improvement in estimating the exact PSD . Increasing the number of samples $N > 32$ may improve the results .

SHM gives some good approximation for different number of samples . when (N = 32 and 16) , it gives good results only at r = 4 , while it fails for N = 8 .

Covariance approximation method and MUSIC method fail to show good approximation for different model orders and different number of samples .

OHM gives some good results . When N = 32 samples , it gives good estimation only at r = 4 . While , at N = 16 samples , it gives good approximation at r = 8 . It fails completely at N = 8 which is the problem with all the aforementioned methods .

4.3 TWO DAMPED SINUSOIDS IN WHITE GAUSSIAN NOISE

The power spectral density of a process consisting of two exponentially decayed sinusoids with different frequencies in white Gaussian noise will be estimated via most of the methods described before .

$$y_n = e^{-0.1n} cos(2\pi f_1 n) + e^{-0.2n} cos(2\pi f_2 n) + e_n$$

where , e_n is white Gaussian noise process with zero mean and variance ε^2 .

Three different cases will be studied :

4.3.1 Very Close Frequencies

In this case the two frequencies ($f_1 = 0.2\ Hz$ and $f_2 = 0.23125\ Hz$) are very close , where $\Delta f < \frac{1}{N}$ and N = 32 , 16 , and 8 . The exact power specrum is shown in Fig # 19 .

Periodograms with different variants are shown in Fig # 20 . They fail to show good approximation . BT method shows good approximation for the exact PSD only at large number of both samples and lags as shown in Fig # 21 . Burg's method fails completely as depicted in Fig # 22 .

Fig # 23 shows the estimated PSD's via Cadzow method . PSD was estimated for different number of samples and SNR's . Cadzow's method fails to give approximation . Increasing the model order the results are improved .

SHM gives a very good results in PSD estimation at high SNR's and large number of samples as shown in Fig # 24 .

Covariance approximation method and MUSIC method fail to give a good approximation even for different model orders as shown in Fig # (25 and 26) respectively .

OHM is reliable as SHM as depicted in Fig # 27 , they look better for low SNR's.

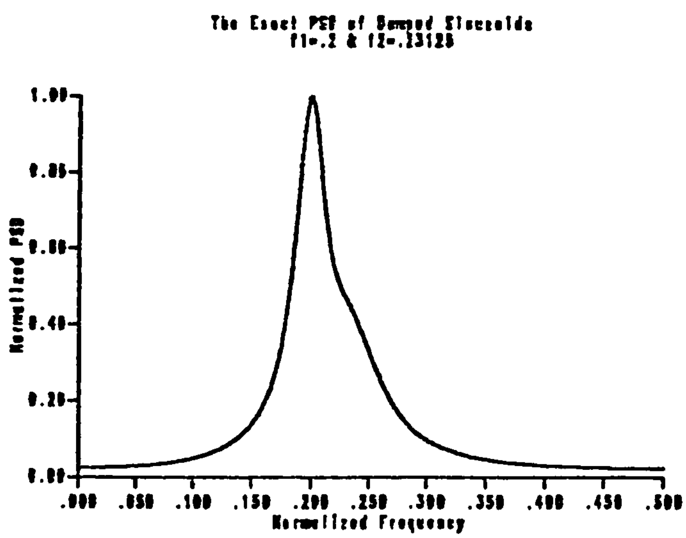

Figure 19: The exact PSD of two damped sinusoids with very close frequencies

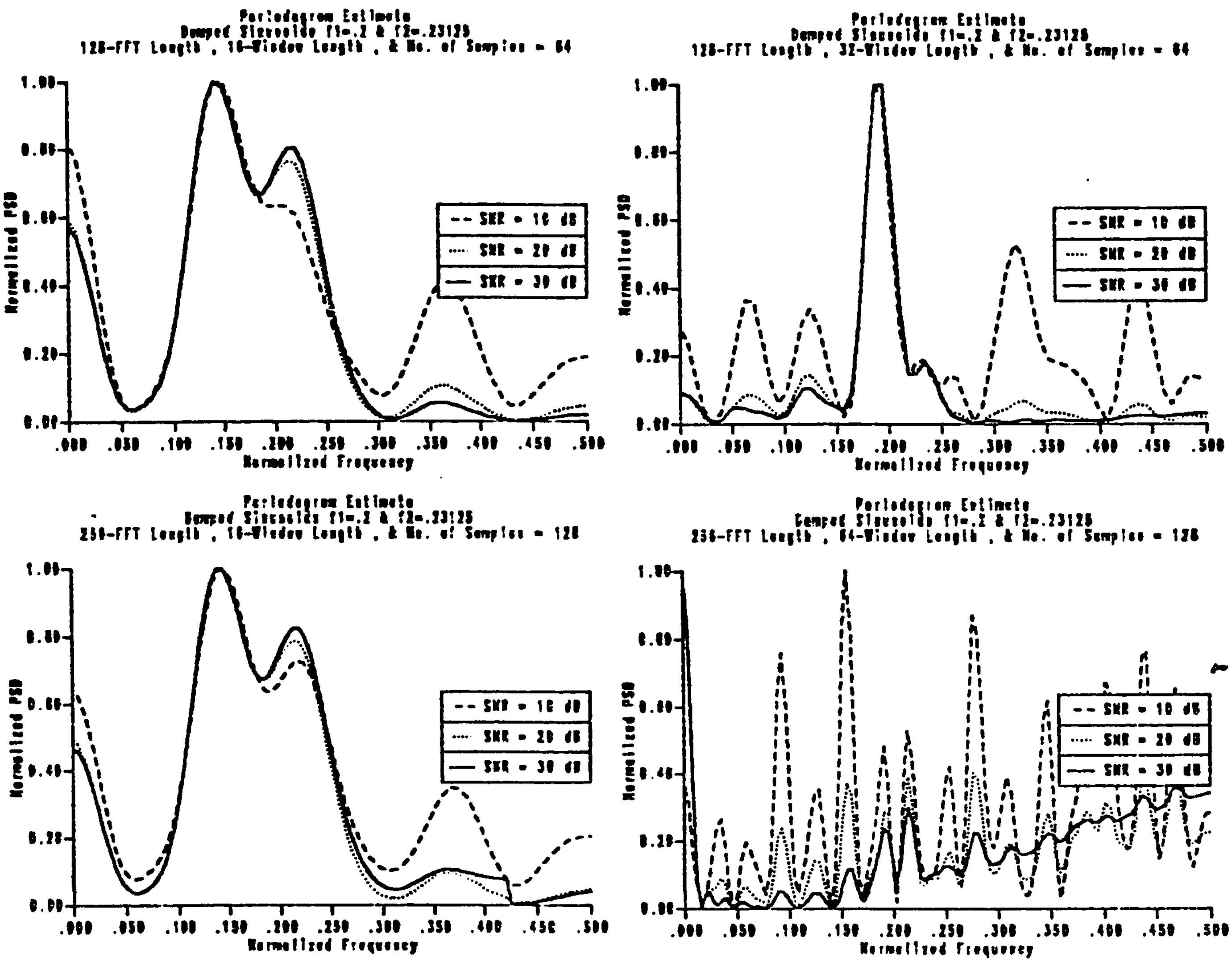

Figure 20: Periodogram PSD's of damped sinusoids (very close freq.)

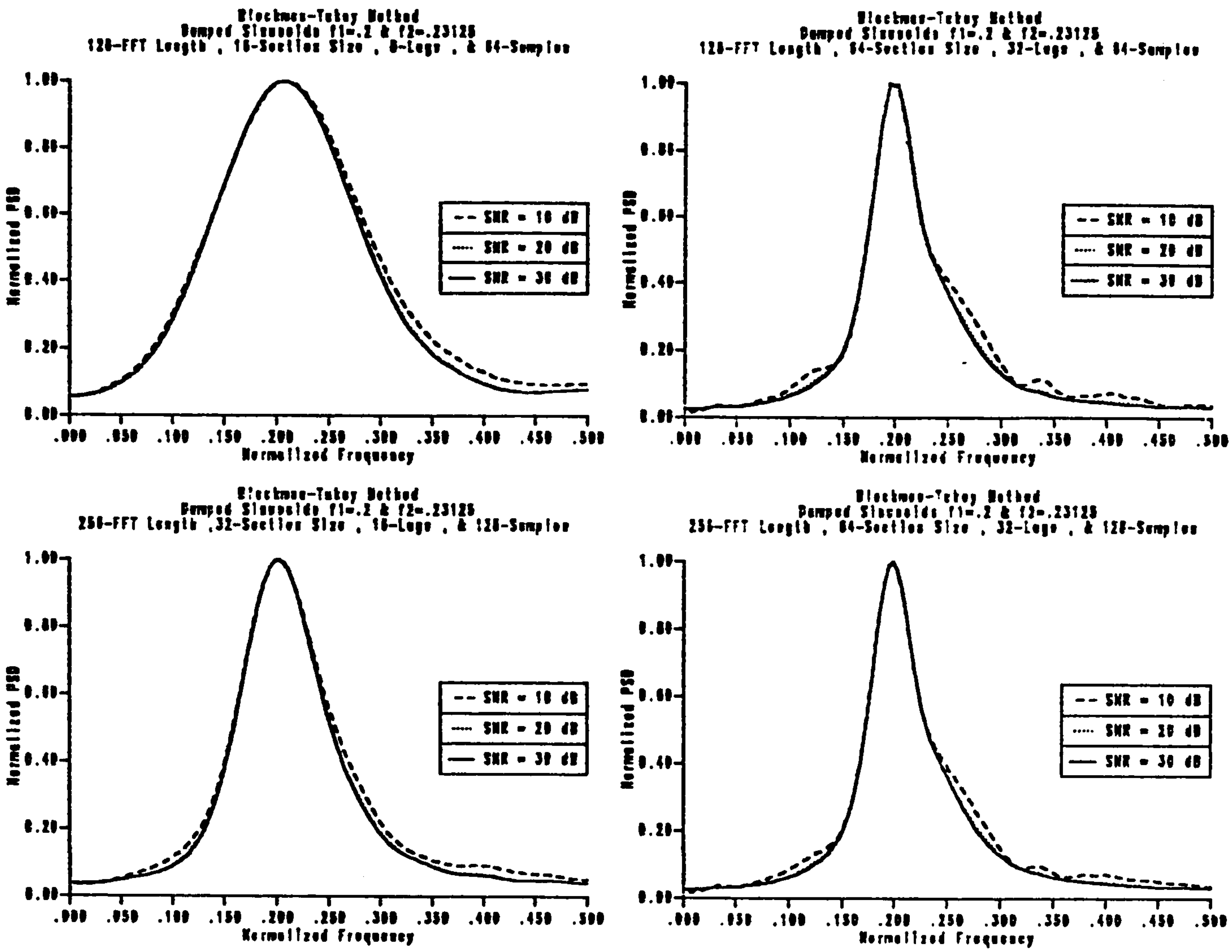

Figure 21: BT PSD's of damped sinusoids (very close freq.)

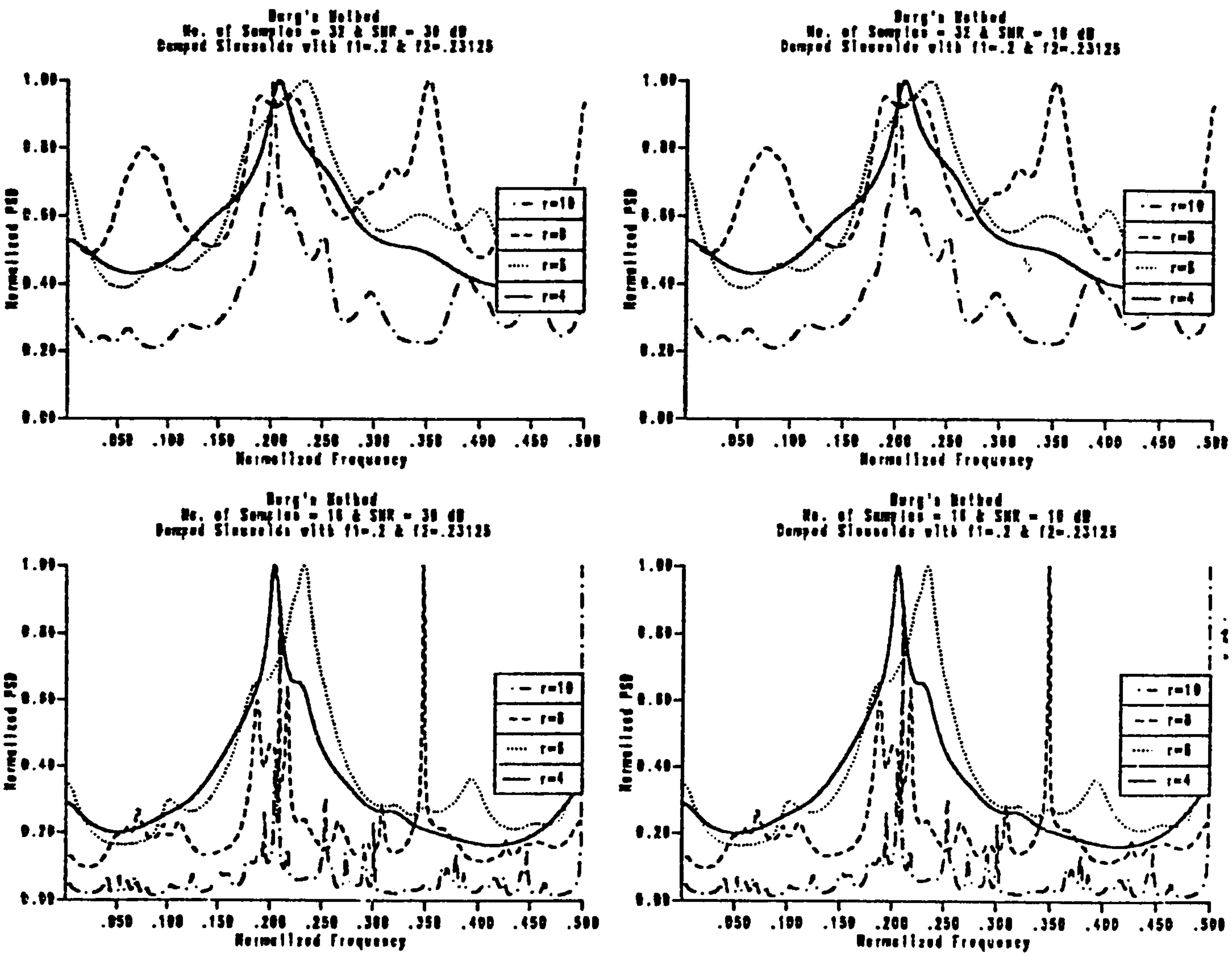

Figure 22: Autoregressive PSD's via Burg algorithm (damped with very close freq.)

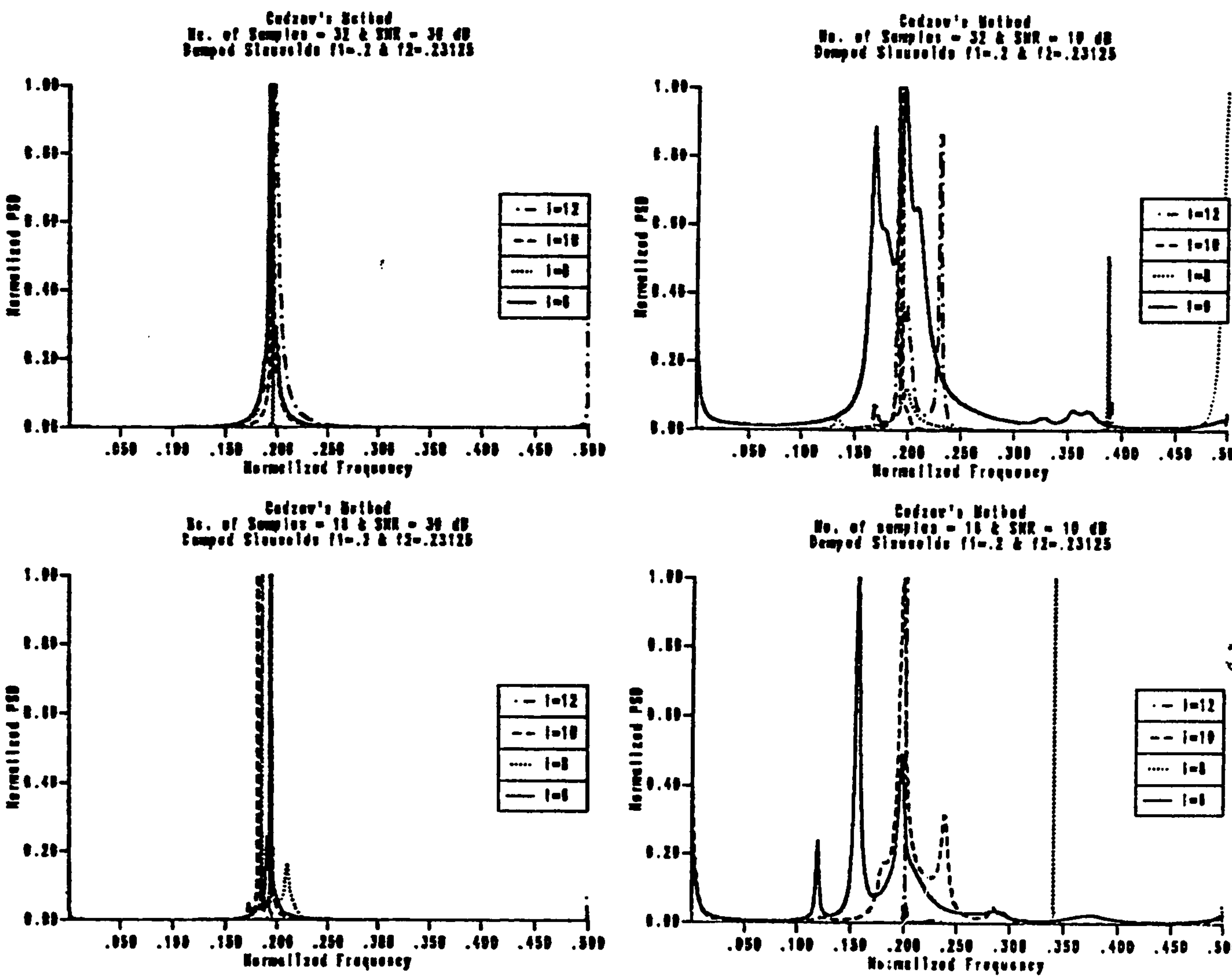

Figure 23: ARMA PSD's of damped with very close frcq. via Cadzow approach

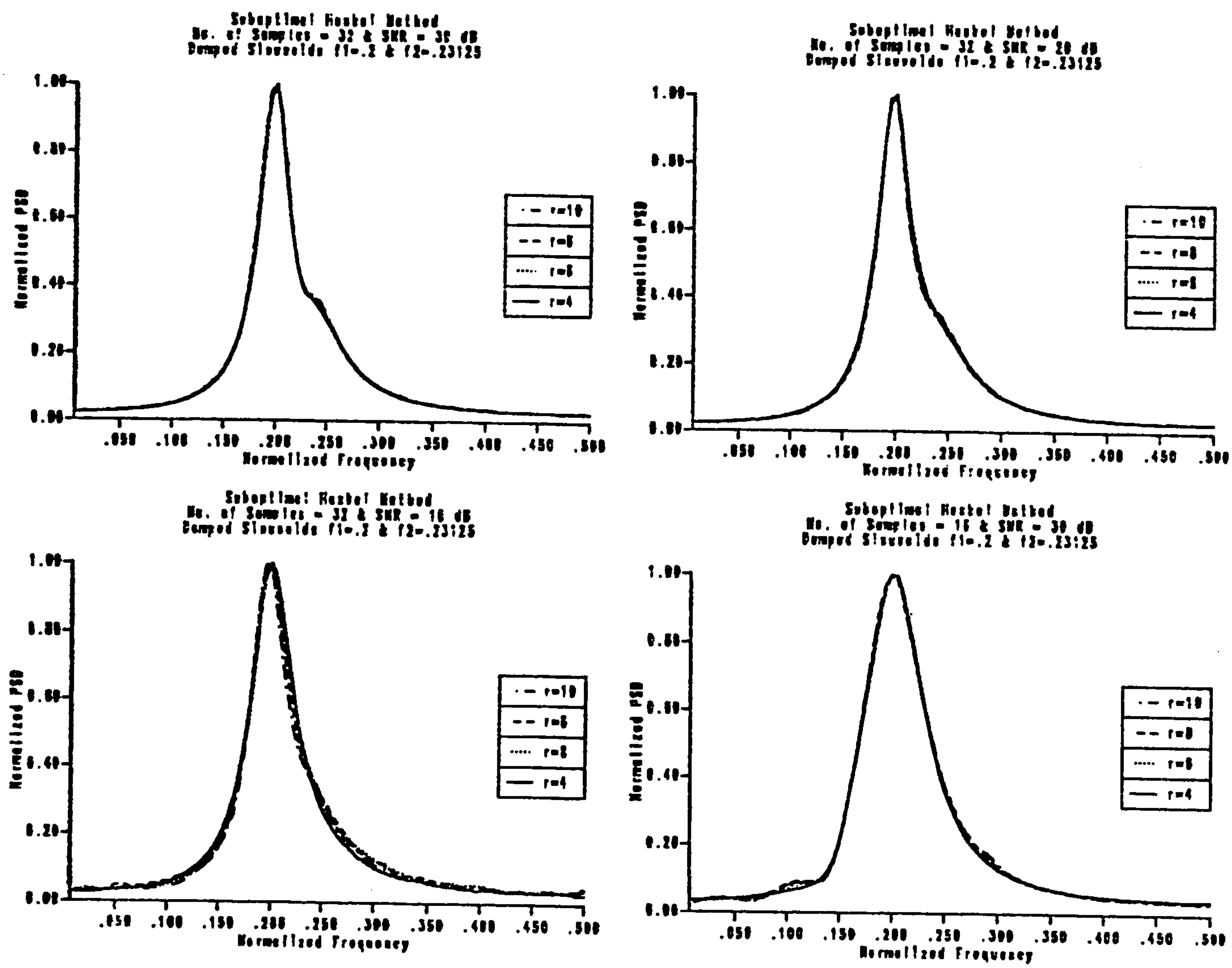

Figure 24: PSD's of damped sinusoids with very close freq. via SHM

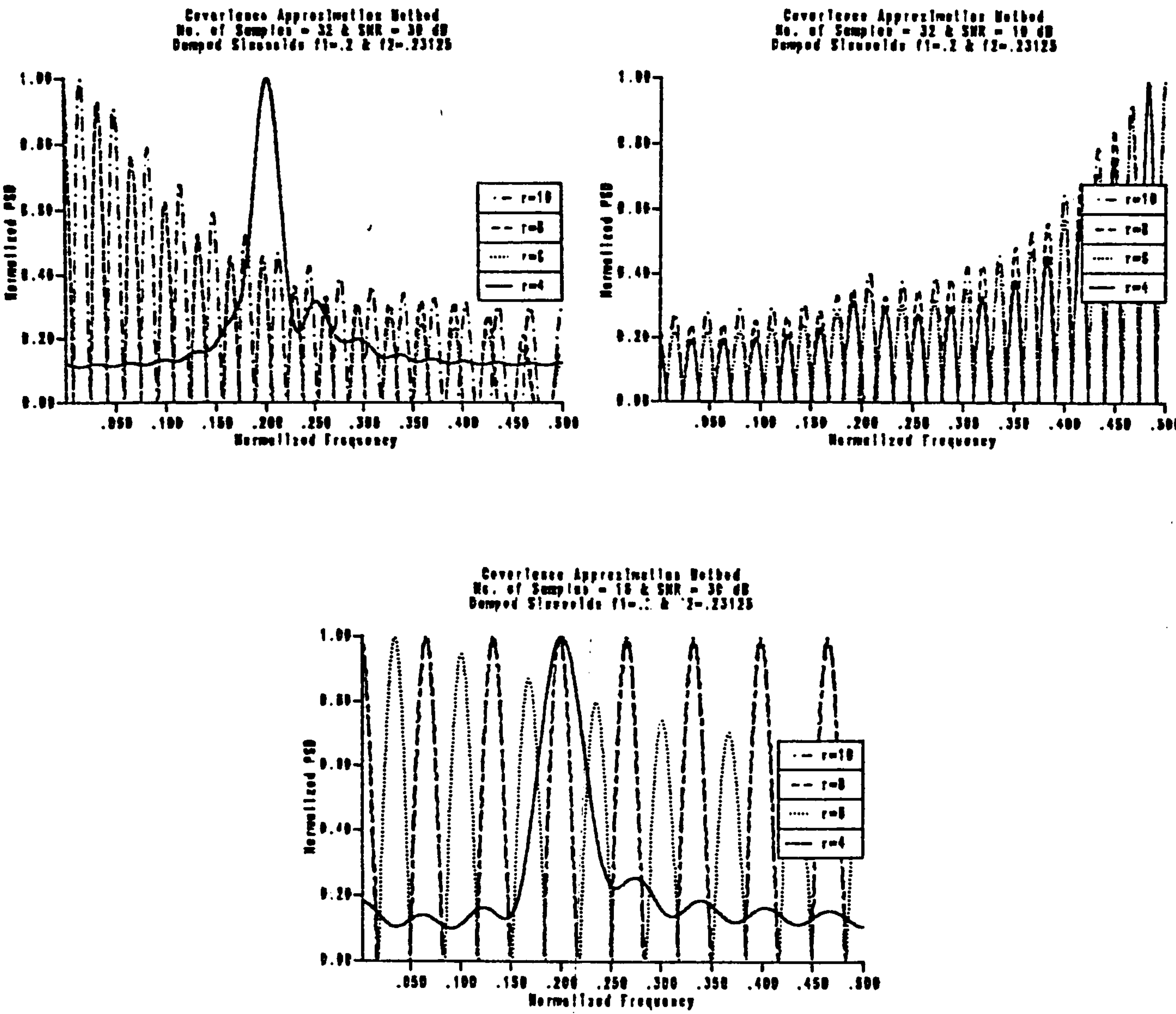

Figure 25: Covariance Approx. method of damped with very close freq.

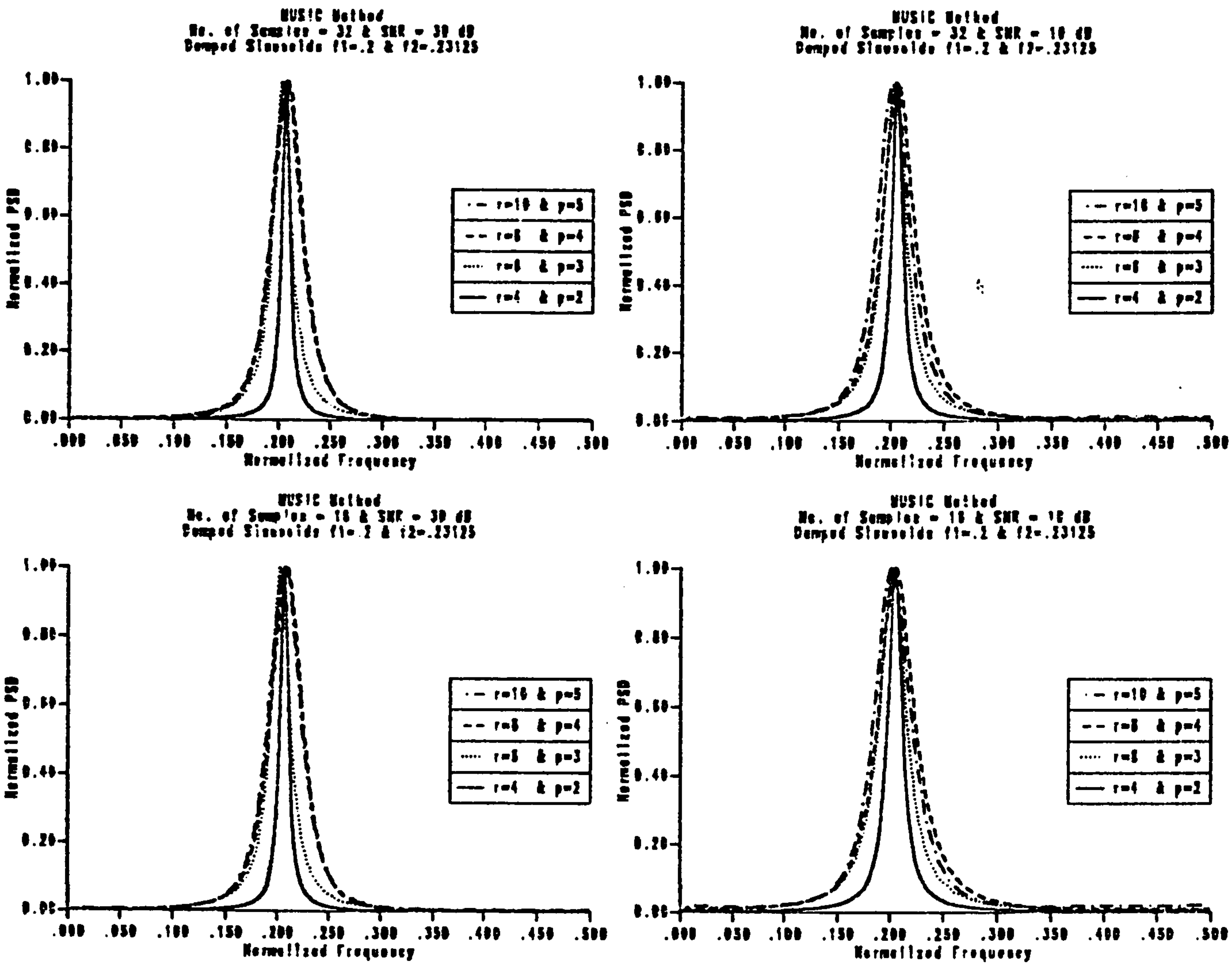

Figure 26: MUSIC estimates of damped sinusoids with very close freq.

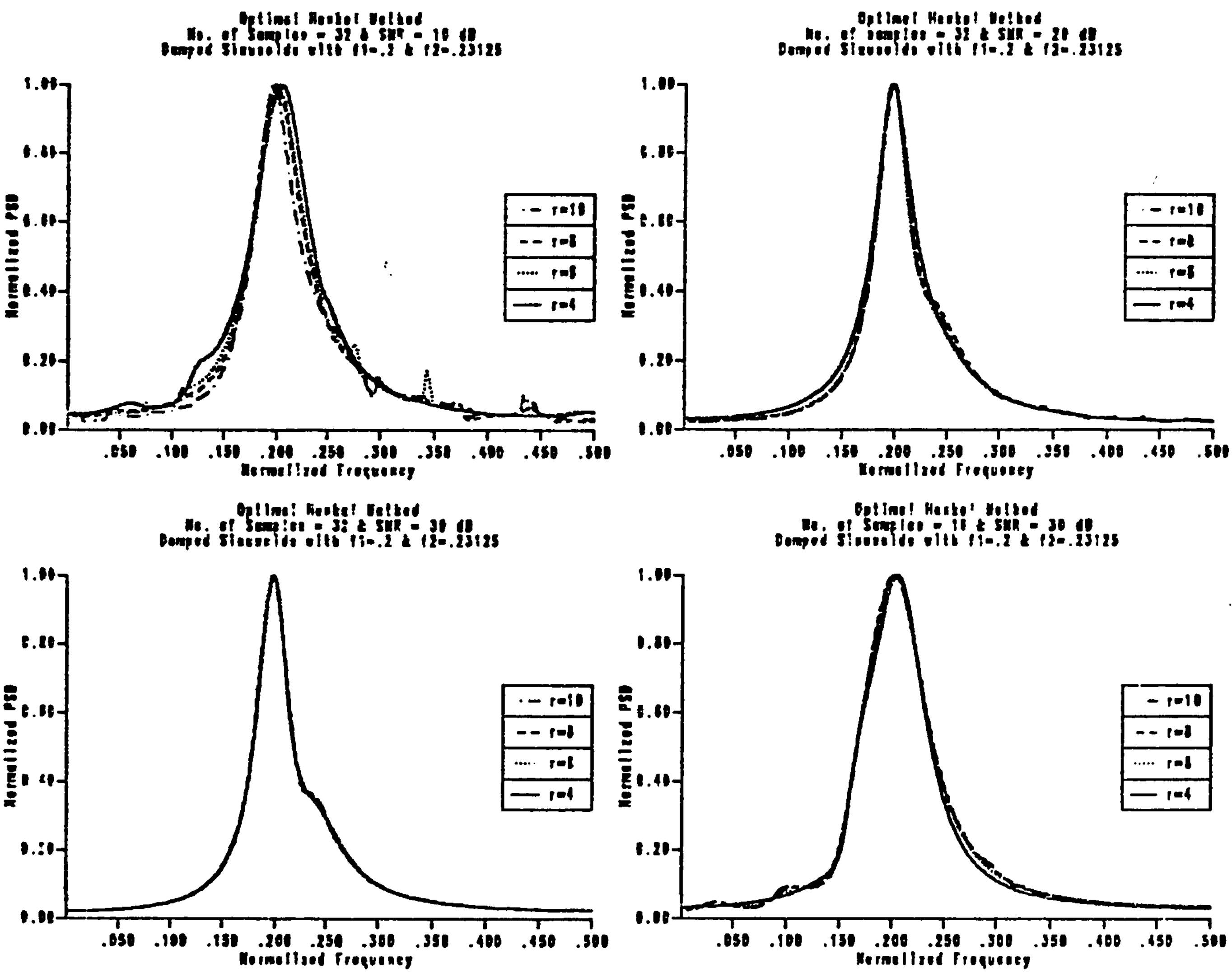

Figure 27: OHM estimates of damped sinusoids with very close freq.

4.3.2 Close Frequencies

In this case the two frequencies ($f_1 = 0.2\ Hz$ and $f_2 = 0.2625\ Hz$) are close , where $\Delta f \leq \frac{1}{N}$ and N = 32 , 16 , and 8 . The exact power specrum is shown in Fig # 28 .

Peridograms with different variants fail to estimate the right PSD as shown in Fig # 29 . BT method shows good improvement at large number of both samples and autocorrleation lags as depicted in Fig # 30 .

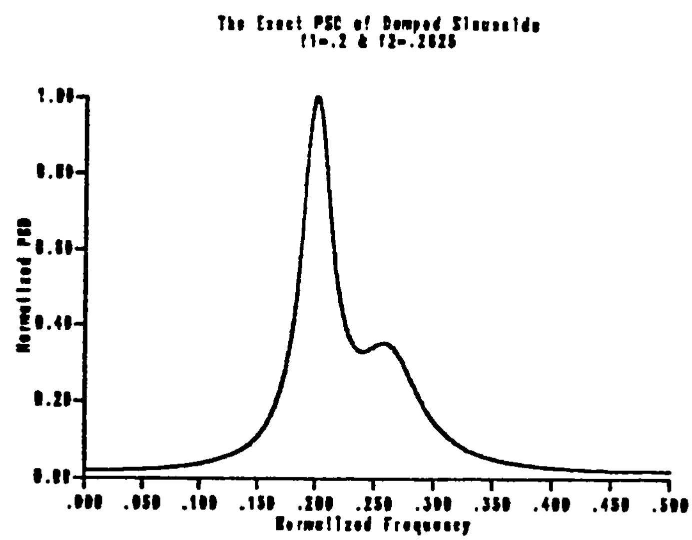

Figure 28: The exact PSD of damped sinusoids with close frequencies

Burg's method fails completely except for larger model orders and higher SNR's as shown in Fig#31 . Cadzow's method fails completely as shown in Fig#32 .

The estimated PSD's via SHM are shown in Fig # 33 . They give good results at different signal to noise ratios (SNR = 30 , 20 , and 10 dB) and different number of samples (N = 32 and 16) .

Covariance approximation method gives reasonable results only when the exact model order is chosen. The plots for different varints are given in Fig#34.

MUSIC method fails completely to estimate the exact PSD even for higher model orders . The graphs are shown in Fig # 35 .

The results obtained by OHM are as good as SHM results . They are shown in Fig # 36 for different number of samples , different number of SNR's , and different model orders .

4.3.3 Far Frequencies

In this case the two frequencies ($f_1 = 0.2\ Hz$ and $f_2 = 0.45\ Hz$) are widely seperated , where $\Delta f \geq \frac{1}{N}$ and N = 32 , 16 , and 8 . Periodogram estimate fails to catch the exact two peaks while BT method gives better results but with large number of samples and lags .

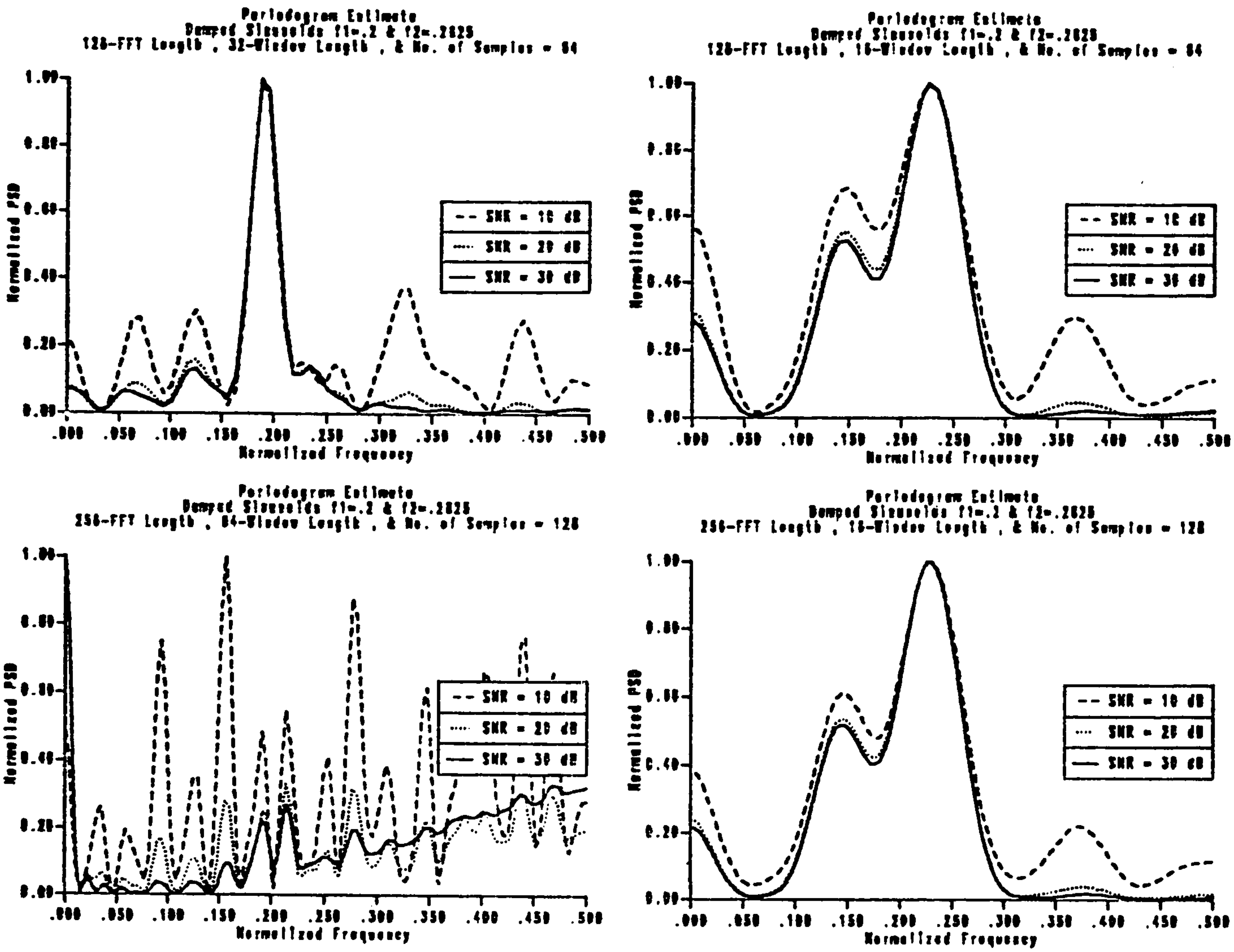

Figure 29: Periodogram PSD's of damped sinusoids (close frequencies)

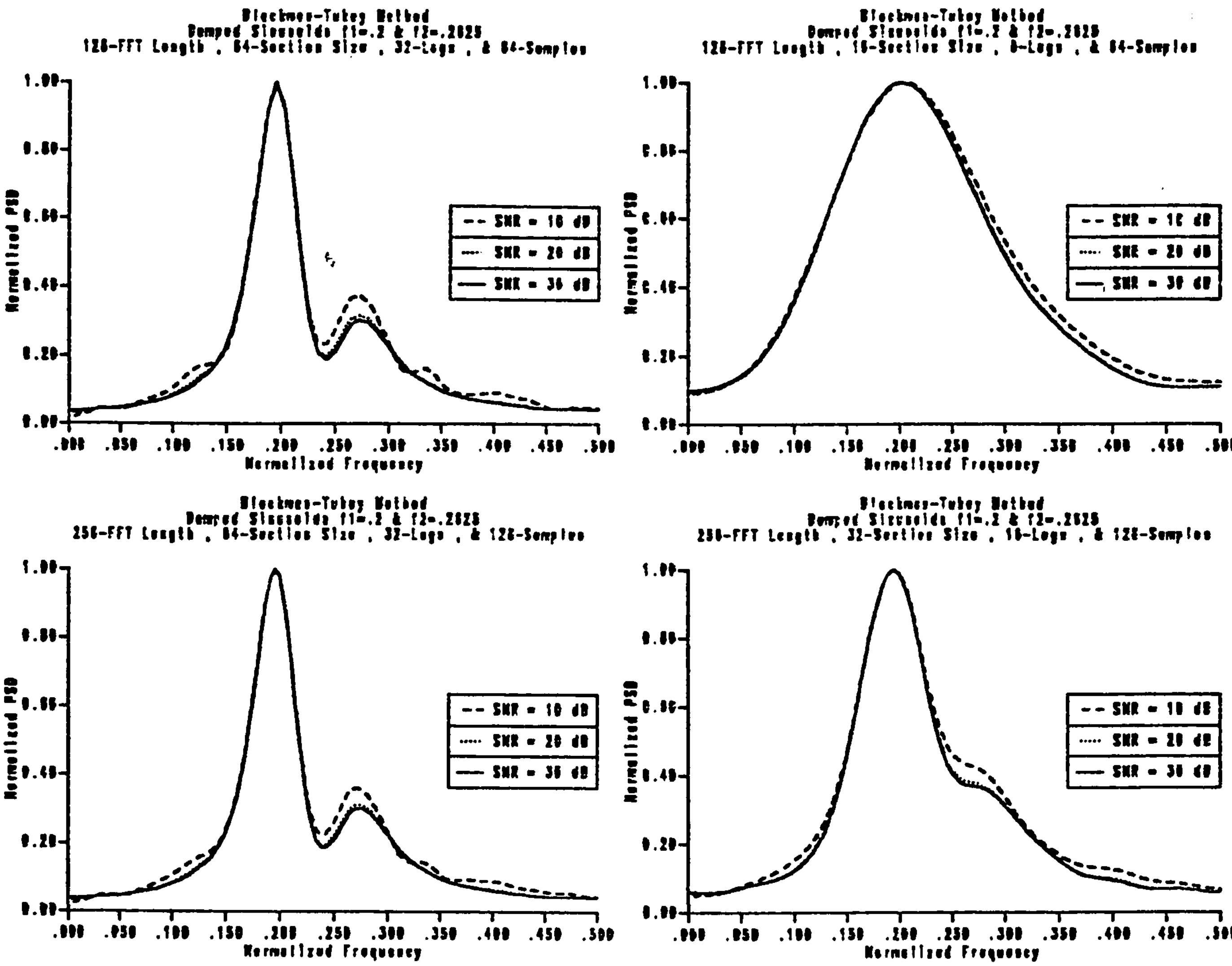

Figure 30: BT PSD's of damped sinusoids (close frequencies)

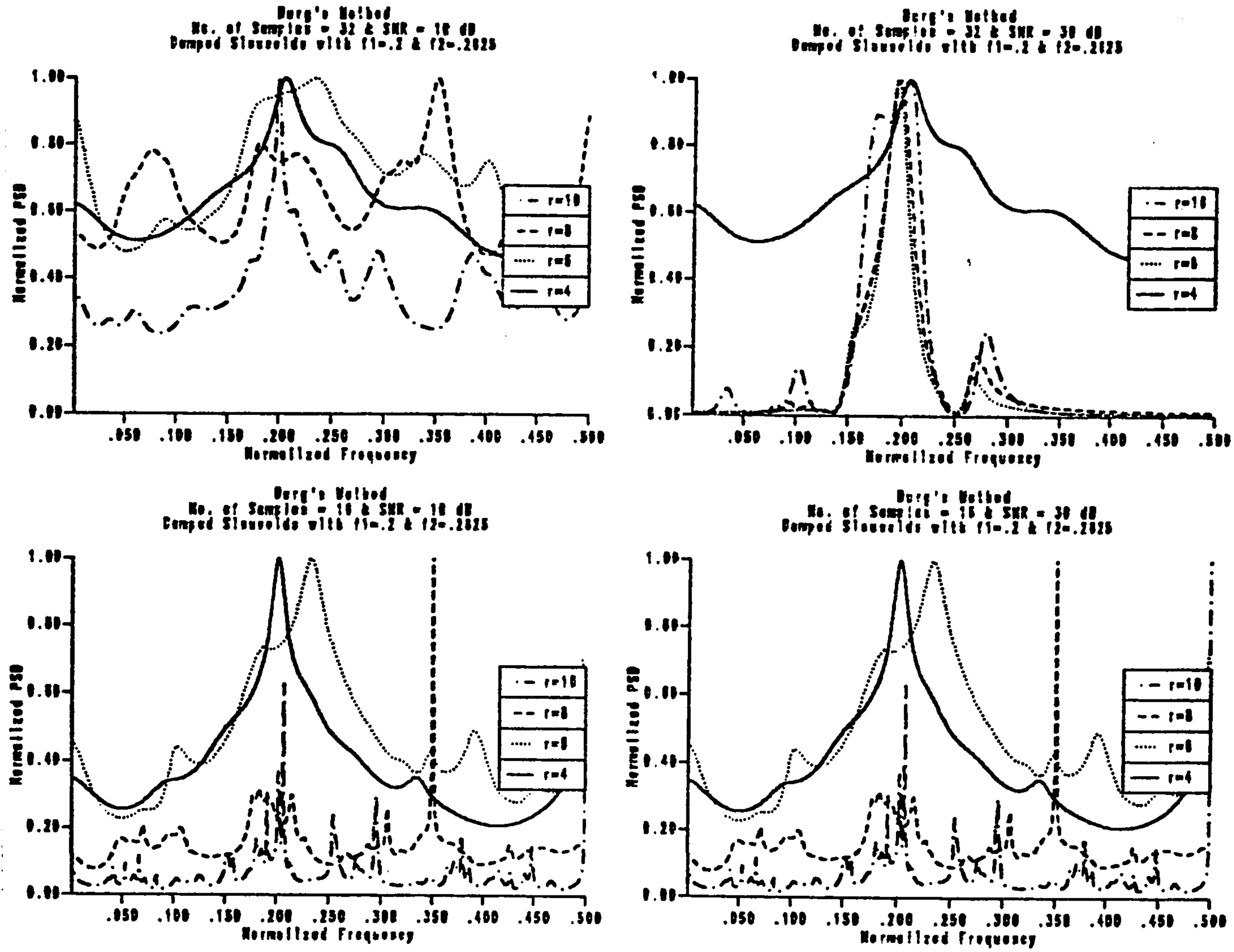

Figure 31: Autoregressive PSD's via Burg algorithm (damped with close frequencies)

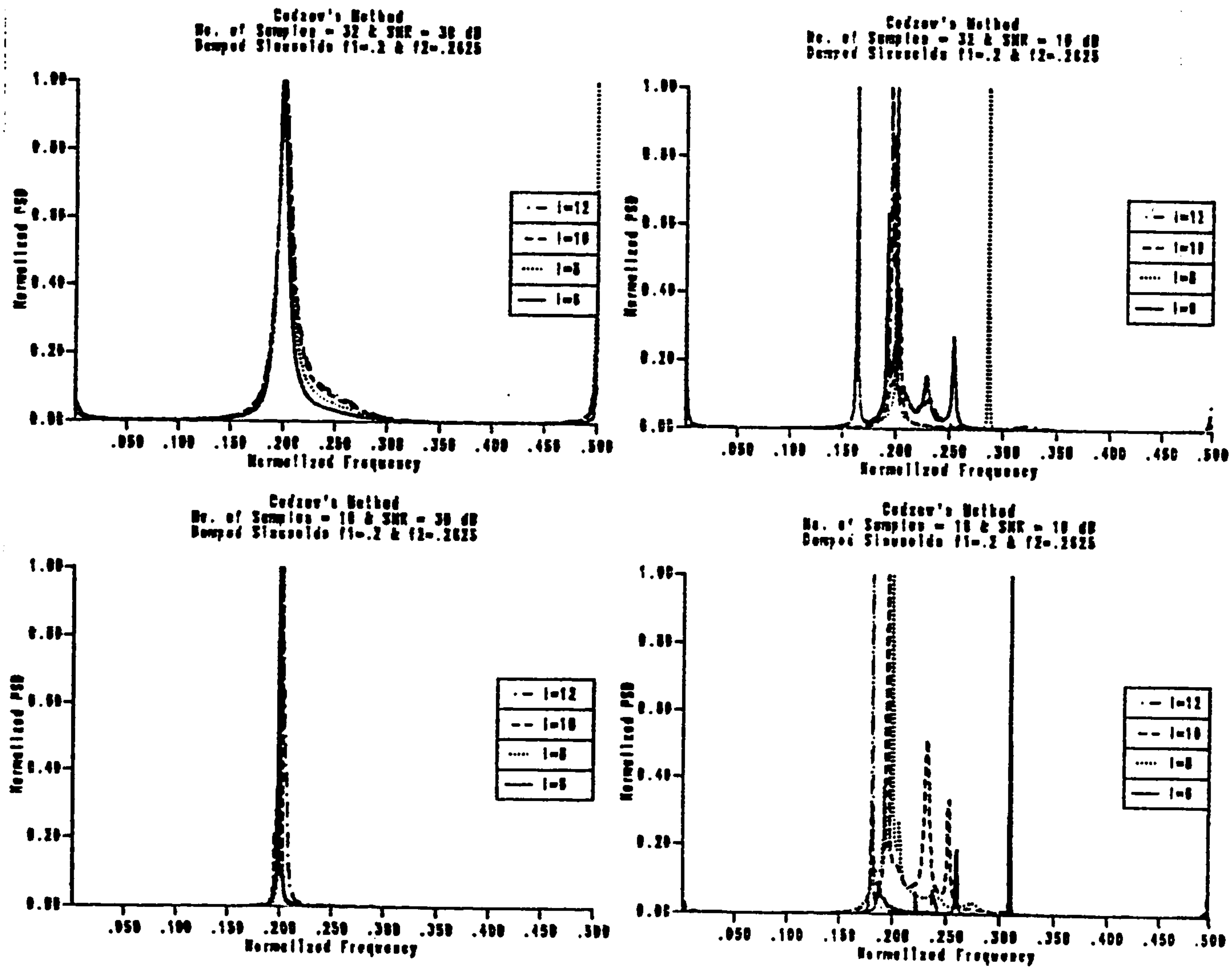

Figure 32: ARMA modeling of damped with close freq. via Cadzow approach

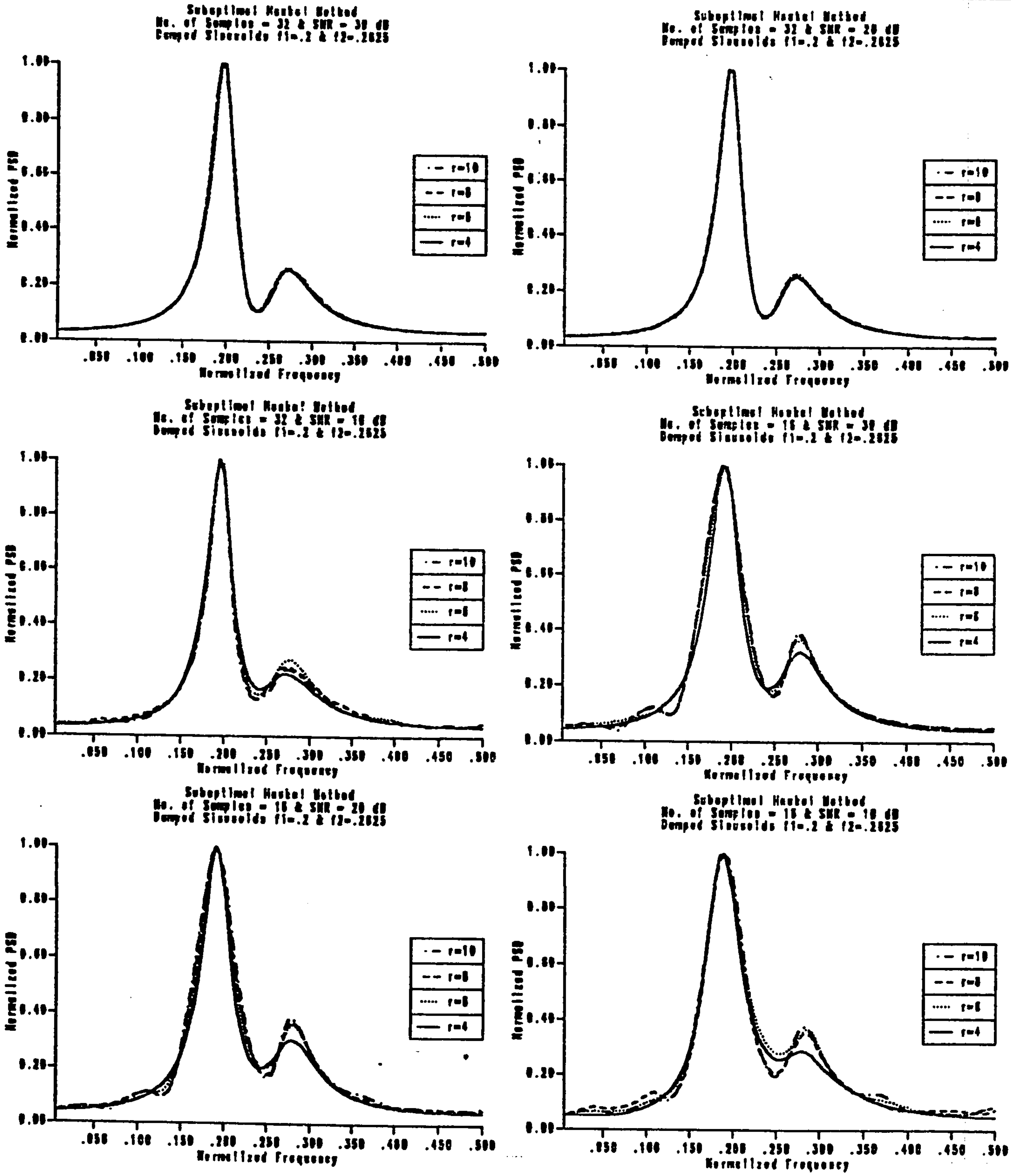

Figure 33: PSD's of damped sinusoids with close freq. via SHM

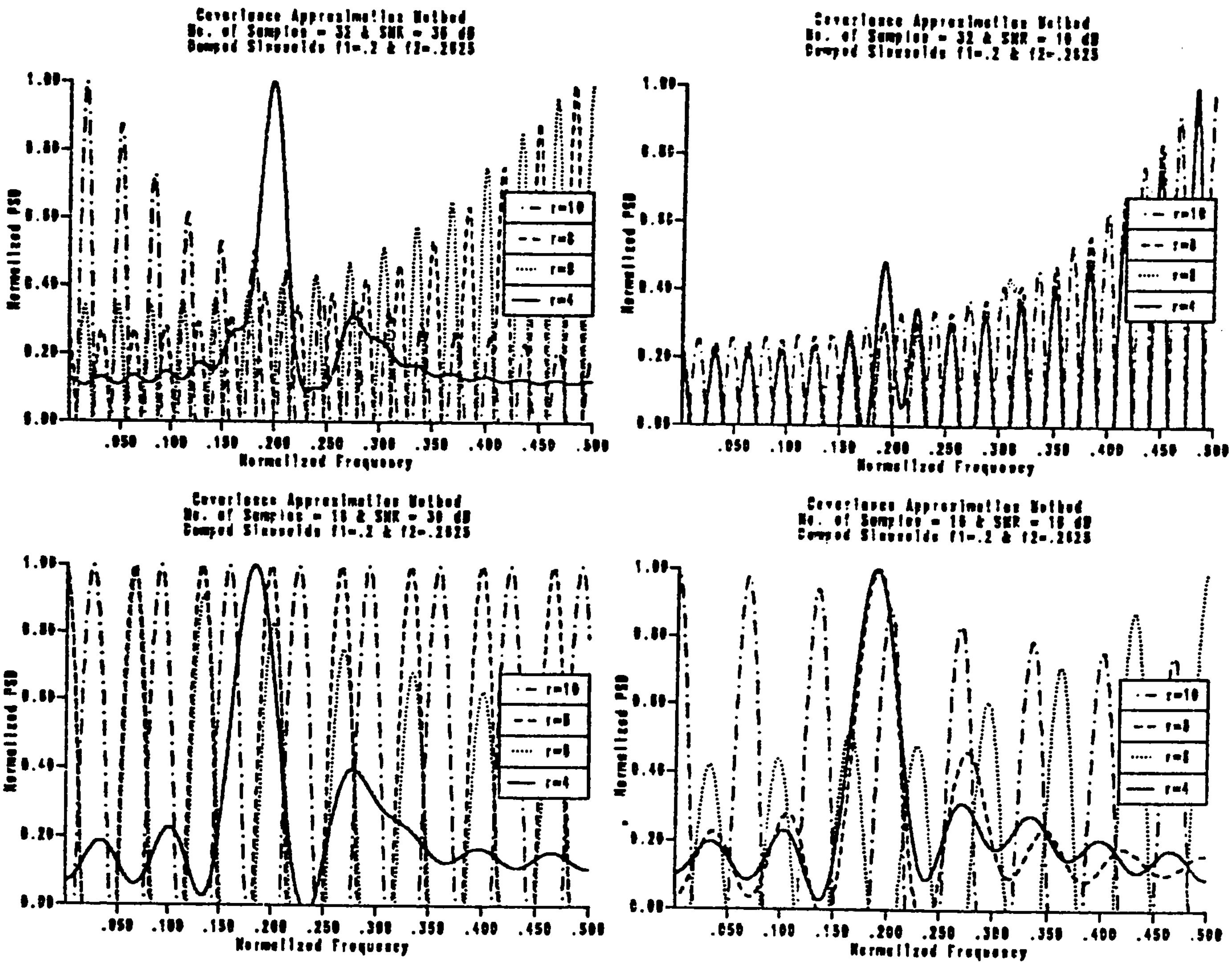

Figure 34: Covariance Approx. method of damped with close freq.

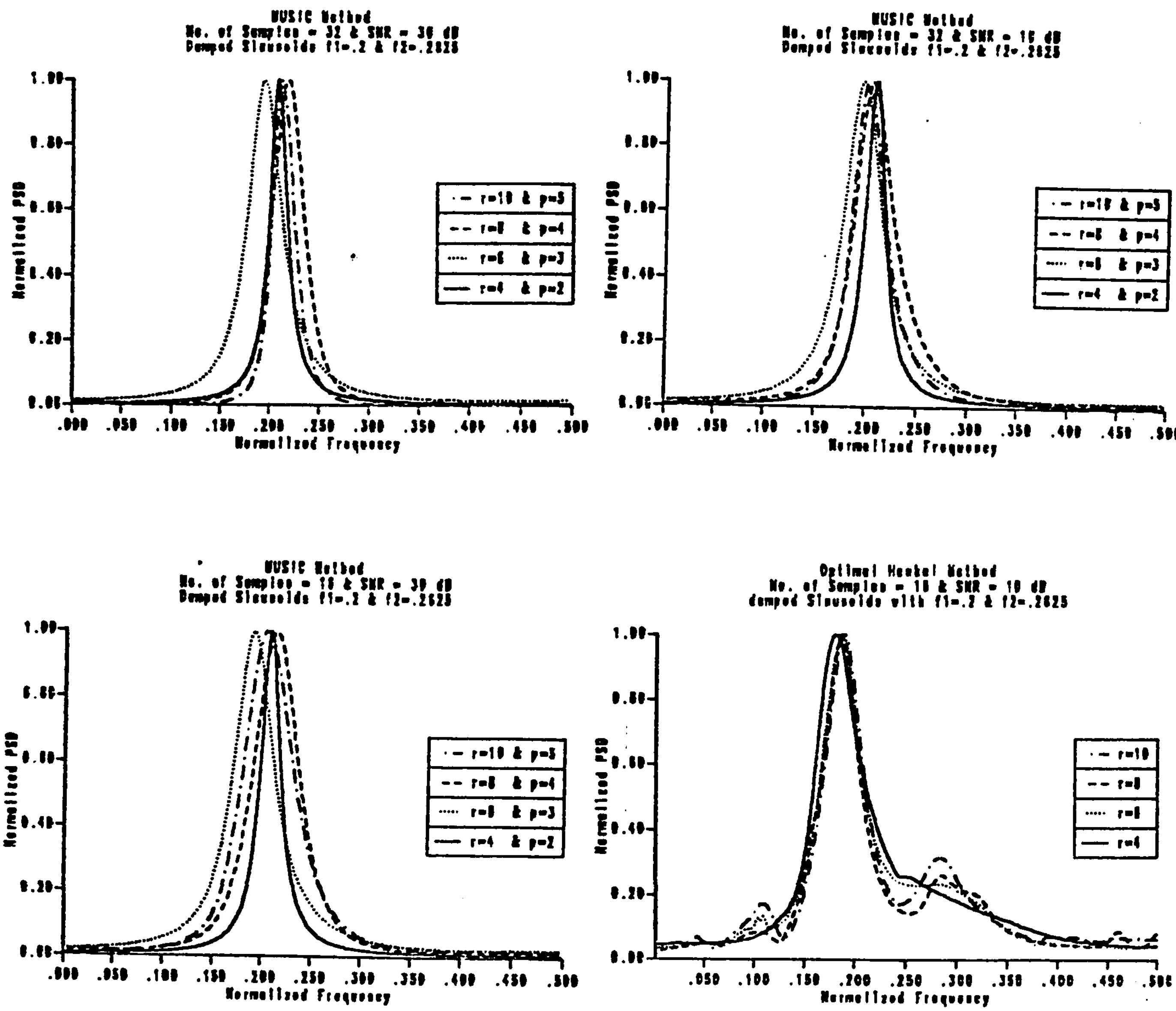

Figure 35: MUSIC estimates of damped sinusoids with close freq.

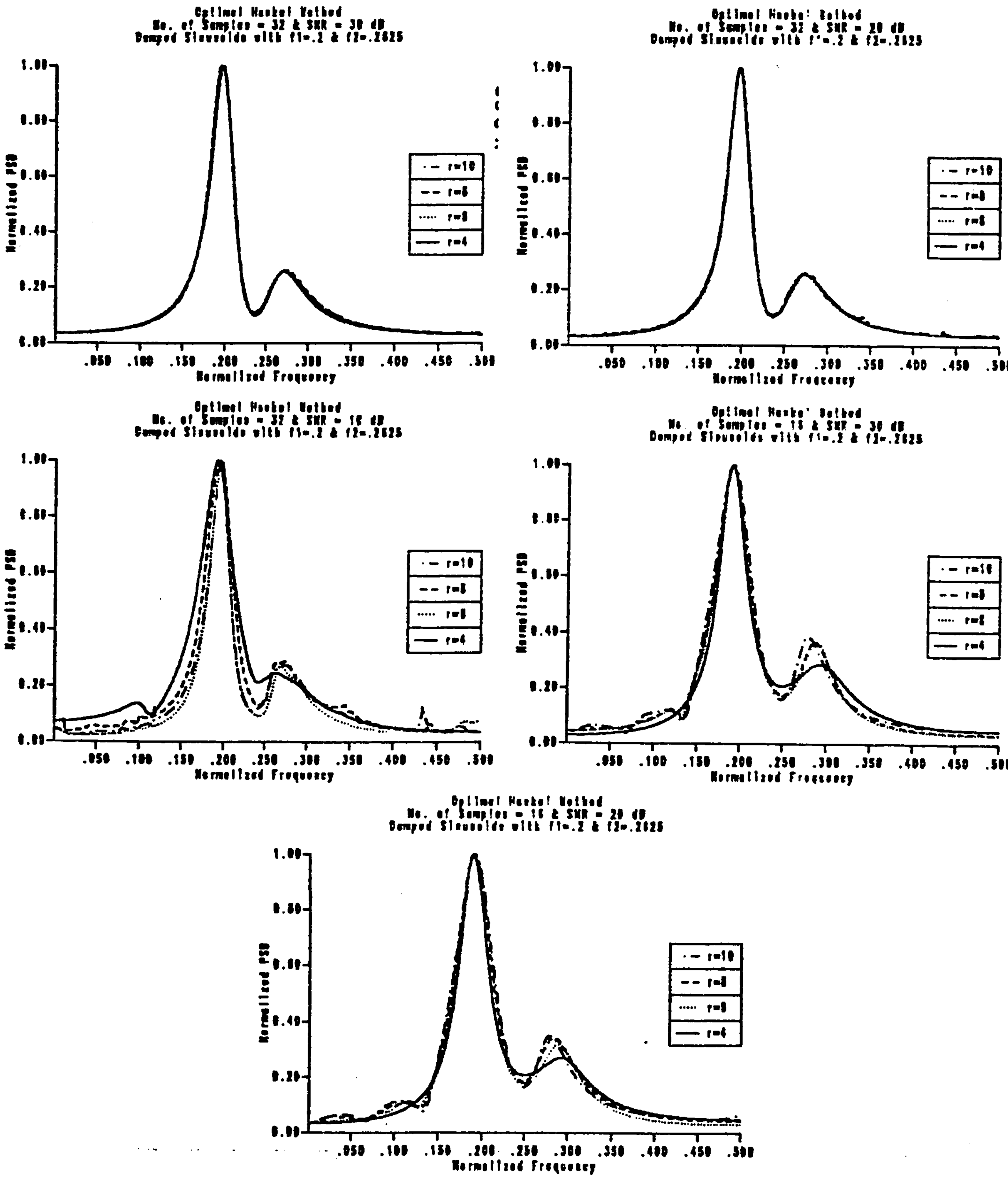

Figure 36: OHM estimates of damped sinusoids with close freq.

Burg's method does not show any good approximation even with different variants . While , Cadzow's method gives some reasonable results at (N = 16) and (SNR = 30 dB) .

By choosing the exact model order , Covariance approximation method will give good results even with low SNR's . MUSIC method works only with very large model orders and number of samples .

OHM and SHM give better and excellant results . They are identical in most of their results . But , OHM gives better only when the number of available samples is small (N = 8) .

4.4 TWO UNDAMPED SINUSOIDS IN WHITE GAUSSIAN NOISE

Estimation of frequencies via spectral estimation is a very important concept in signal processing and could be cosidered as an appliction of power spectral estimation . The power spectral density of a process consisting of two undamped sinusoids with different frequencies in white Gaussian noise will be estimated via most of the methods explained before . Considering the process given below , three different cases will be studied :

$$y_n = \sqrt{20}\, cos(2\pi f_1 n) + \sqrt{2}\, cos(2\pi f_2 n) + e_n$$

where , e_n is white Gaussian noise process with zero mean and variance ε^2 .

Three different cases will be studied :

4.4.1 Very Close Frequencies

In this case the two frequencies ($f_1 = 0.2\ Hz$ and $f_2 = 0.23125\ Hz$) are very close , where $\Delta f < \frac{1}{N}$ and N = 32 , 16 , and 8 . The exact power specrum is shown in Fig # 37 .

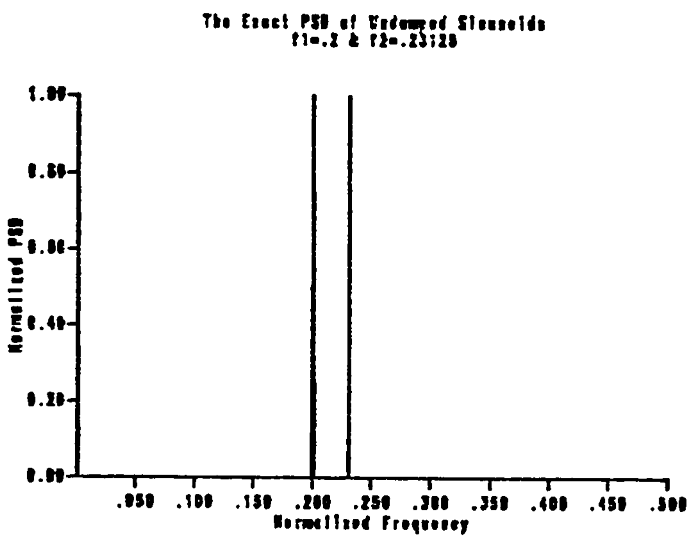

Figure 37: The exact PSD of undamped sinusoids with very close frequencies

Periodograms with different variants are shown in Fig # 38 . They give good approximation at large number of both samples and FFT length . Fig # 39 illustrates the BT spectrum . When the number of autocorrelation lags is small the spectrum will not resolve the two frequencies . Increasing the number of both lags and samples will not improve the results .

The AR PSD's based on Burg's algorithm are shown in Fig # 40 . Burg's method fails completely to determine the two frequencies . Also , Cadzow's method does not resolve the two frequencies which is due to the short data length as shown in Fig # 41 .

SHM shows very good results at different signal to noise ratios (SNR = 30 , 20 , and 10 dB) at number of samples (N = 32) . The results are given in Fig # 42 .

Covariance approximation method is not able to resolve the desired frequencies irrespective of any combination of its parameters . It was simulated and the results are depicted in Fig # 43 .

MUSIC method gives good results and it resolves the two frequencies only at high signal to noise ratios (SNR = 30 dB) at different number of samples (N = 32 and 16) . The simulated results are shown in Fig # 44 .

OHM gives the same results obtained by SHM . These results are shown in Fig # 45 .

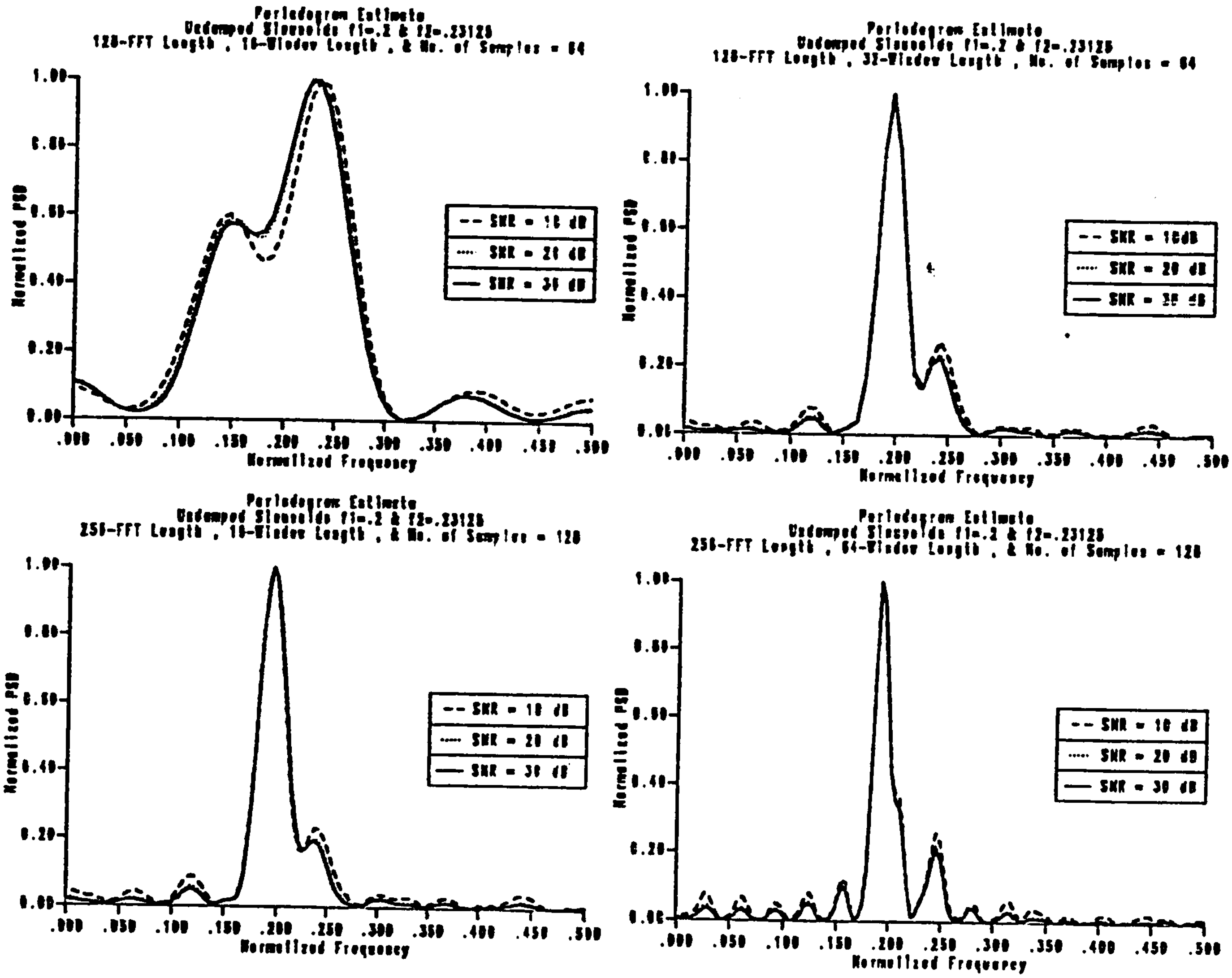

Figure 38: Periodogram PSD's of undamped sinusoids (very close frequencies)

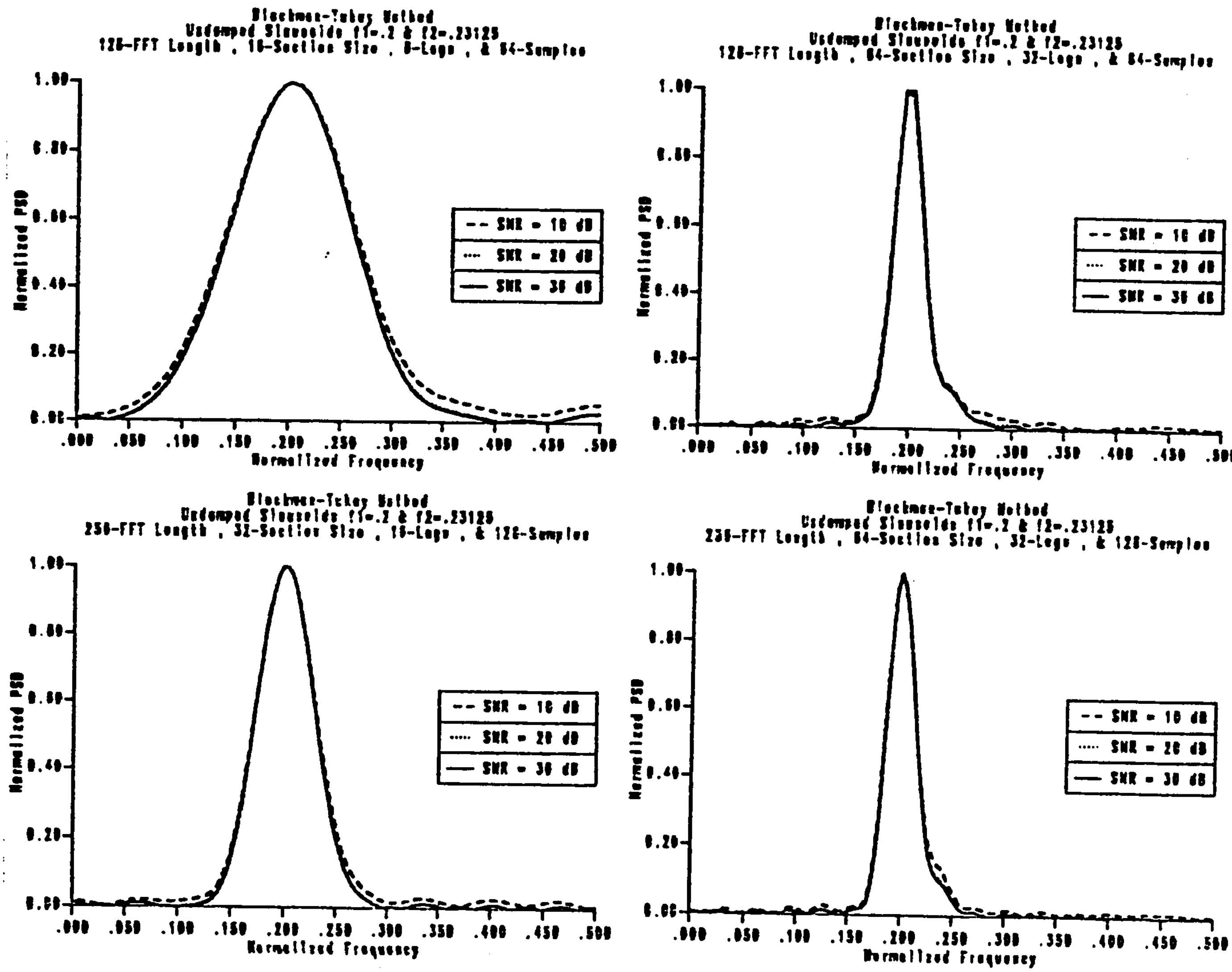

Figure 39: BT PSD's of undamped sinusoids (very close frequencies)

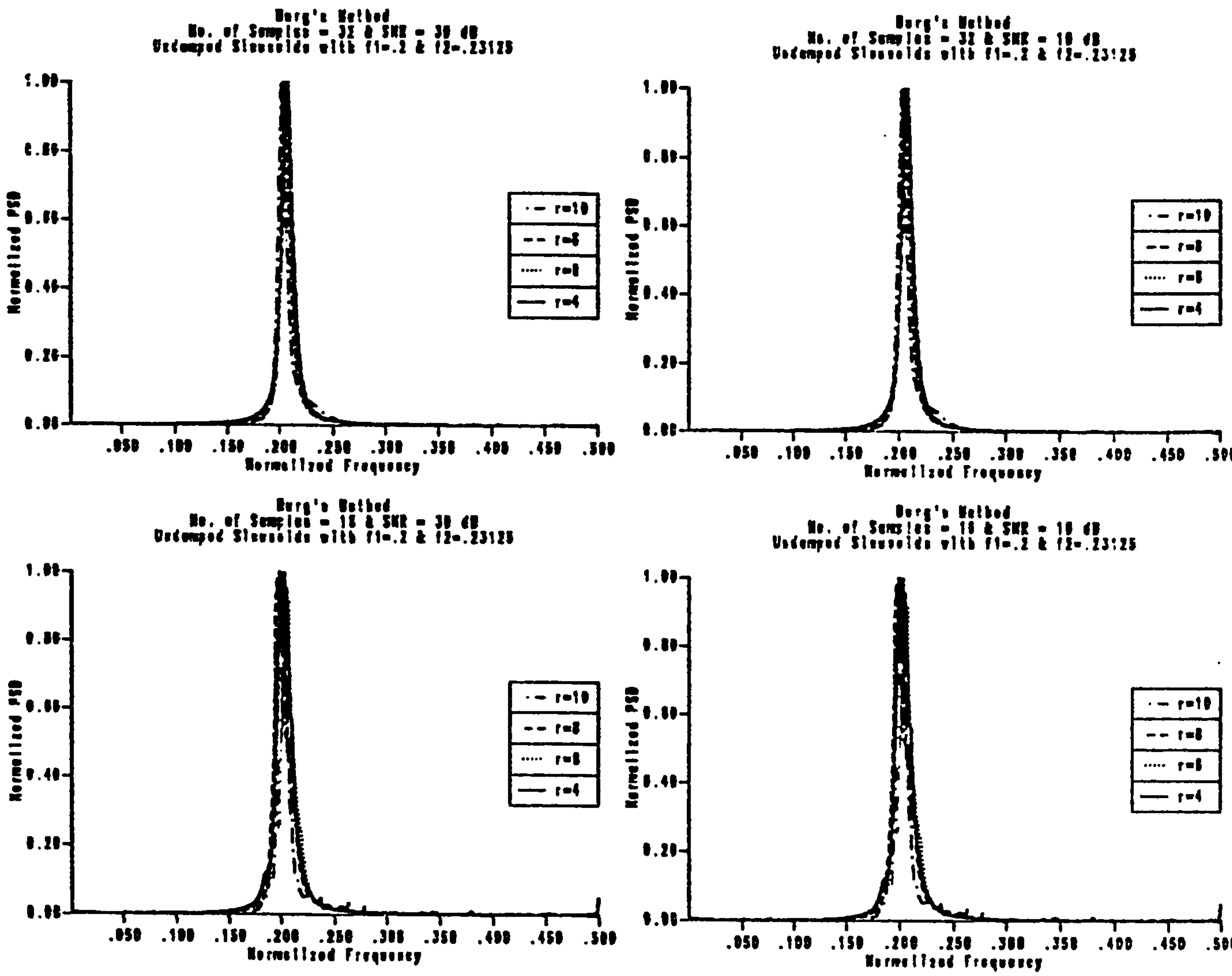

Figure 40: Autoregressive PSD's via Burg algorithm (undamped with very close frequencies)

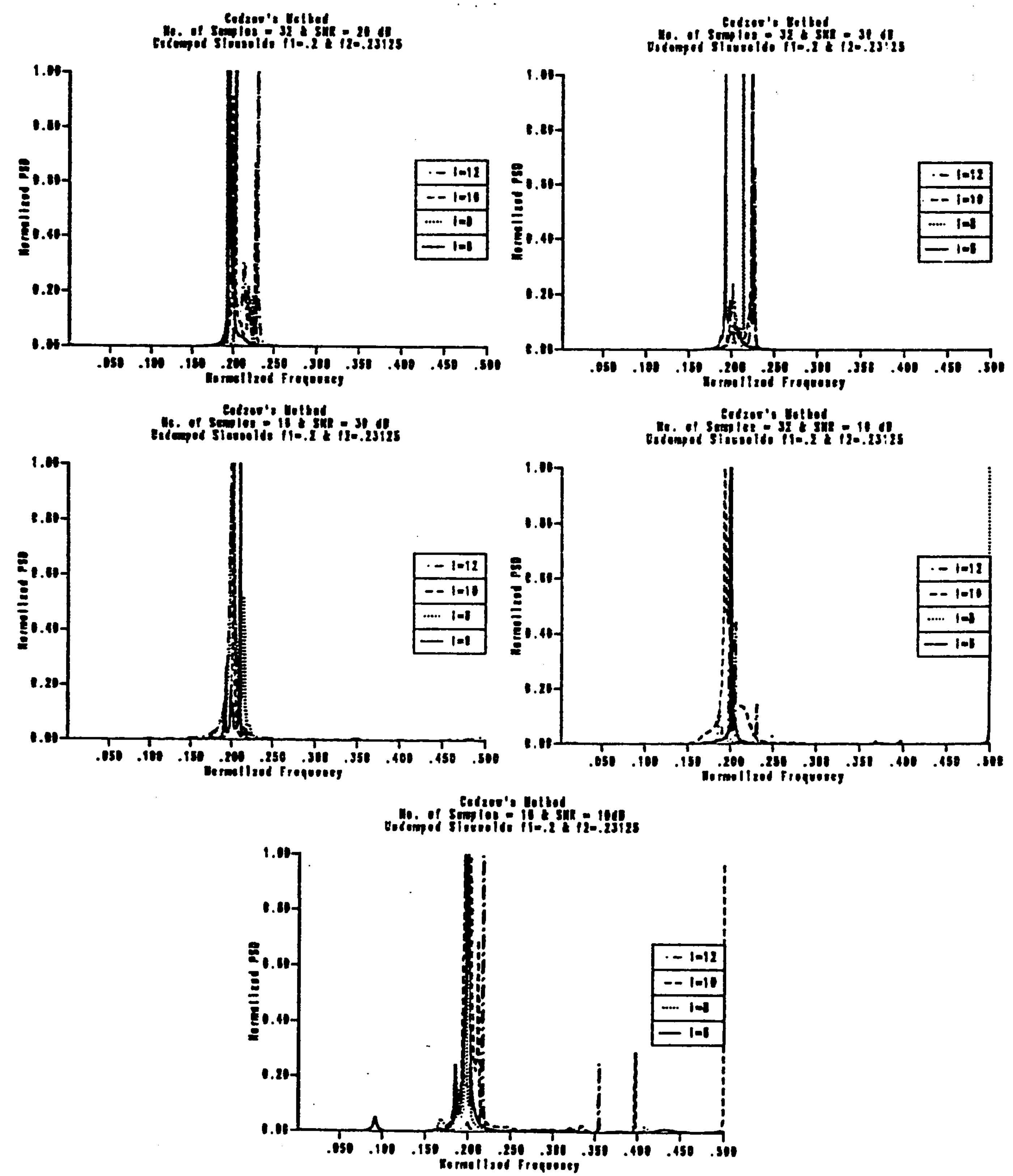

Figure 41: ARMA modeling of undamped(very close freq.) via Cadzow approach

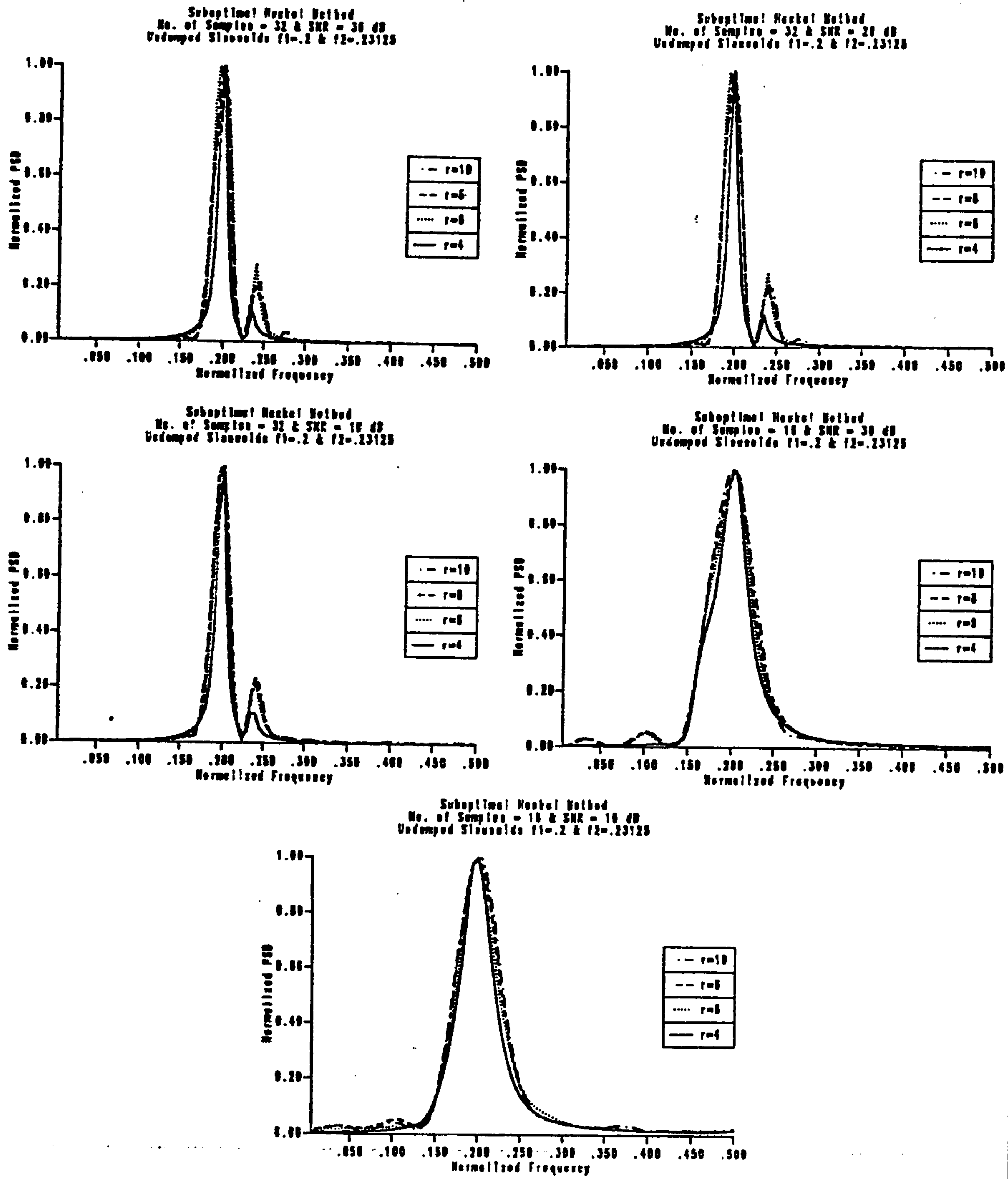

Figure 42: PSD's of undamped sinusoids with very close freq. via SHM

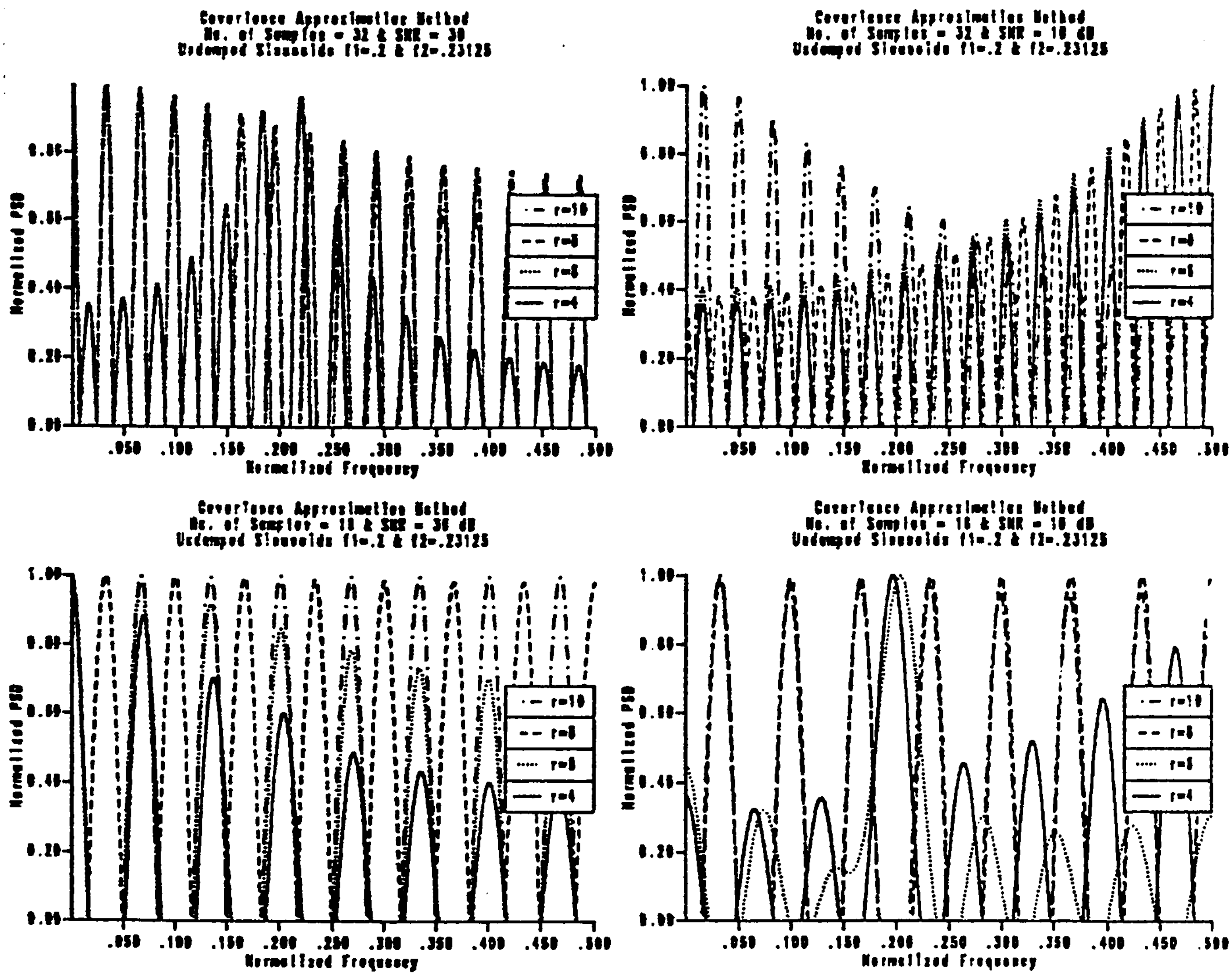

Figure 43: Covariance Approx. method of undamped(very close freq.)

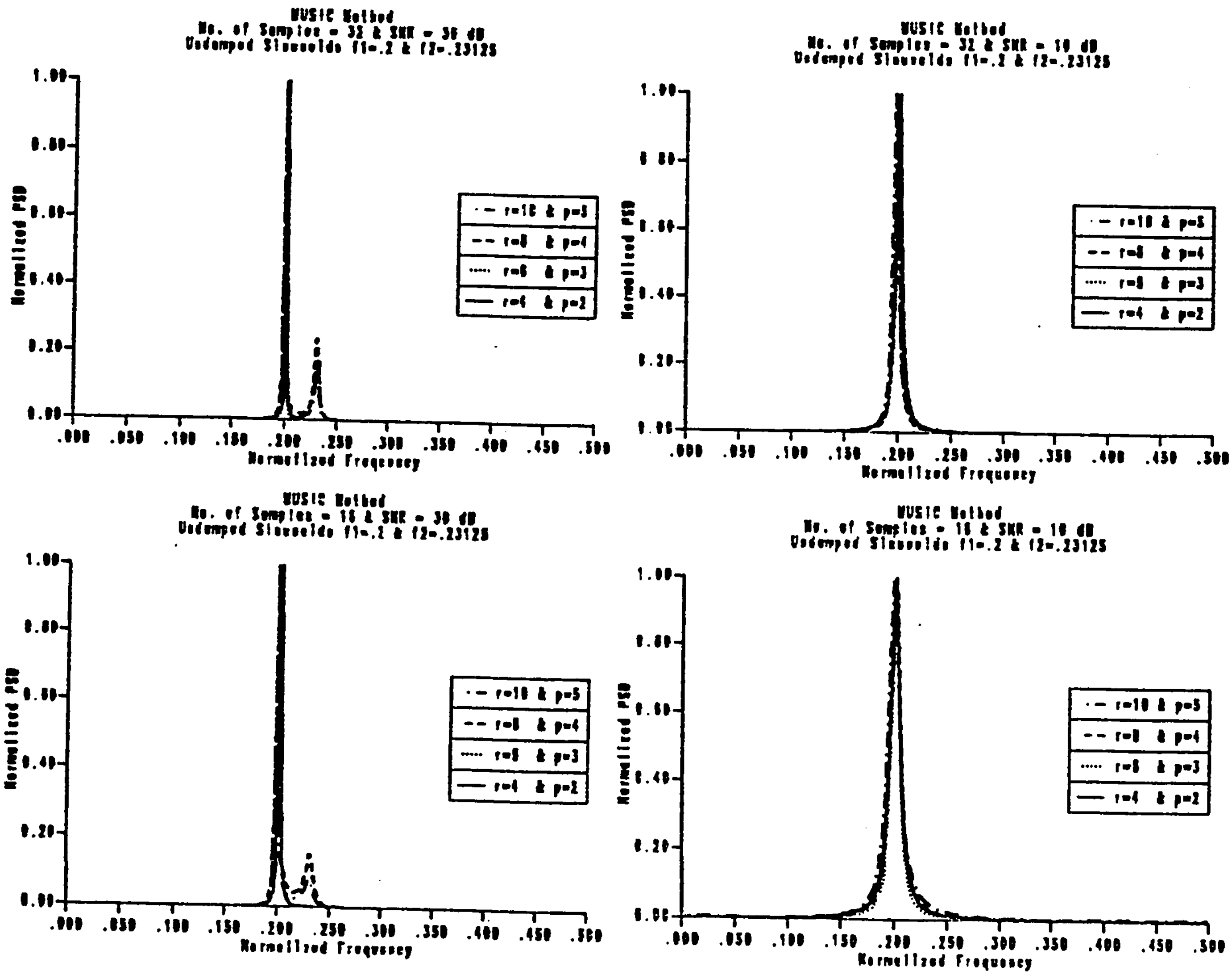

Figure 44: MUSIC estimates of undamped sinusoids with very close freq.

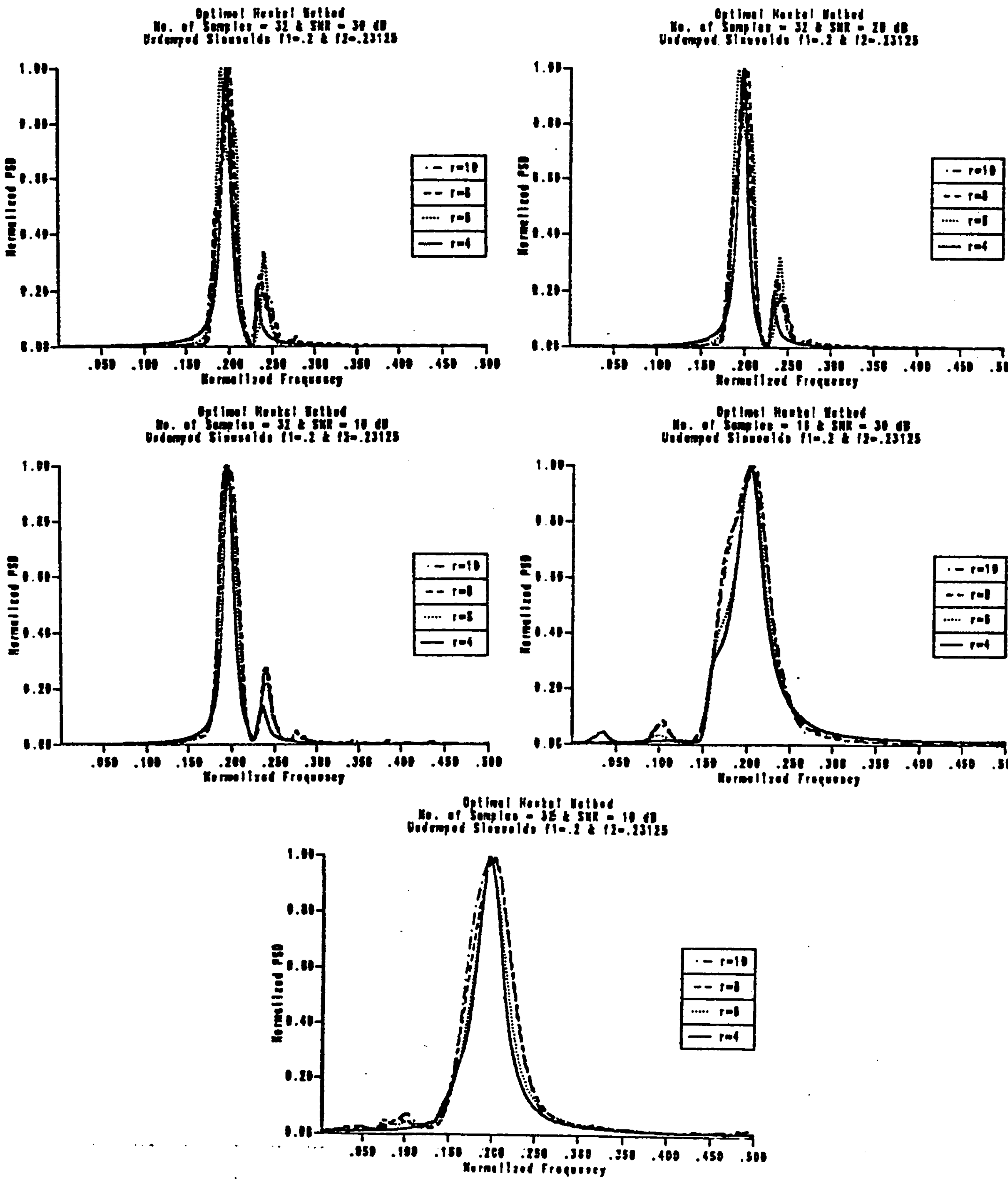

Figure 45: OHM estimates of undamped sinusoids with very close freq.

4.4.2 Close Frequencies

In this case the two frequencies (f_1 = 0.2 Hz and f_2 =0.2625 Hz) are close , where $\Delta f \leq \frac{1}{N}$ and N = 32 , 16 , and 8 . The exact power specrum is shown in Fig # 46 .

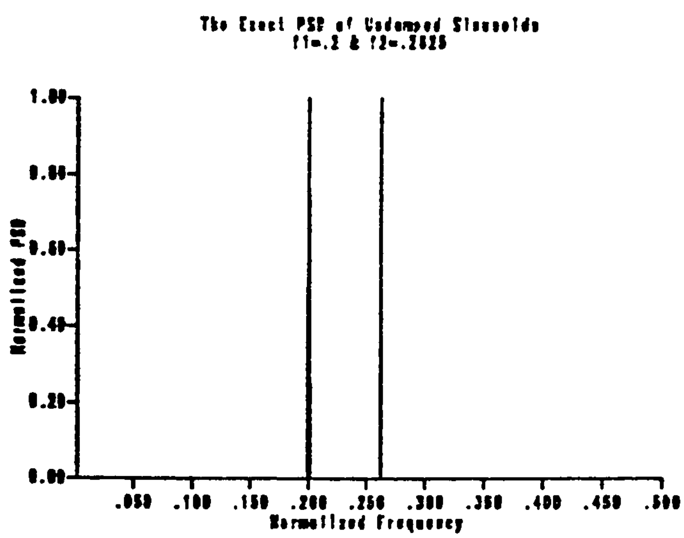

Figure 46: The exact PSD of undamped sinusoids with close frequencies

For short data record , traditional methods failed to show two different peaks at the desired frequencies . the periodograms with different variants , shown in Fig # 47 , are unable to resolve the two frequencies except at 64 - window length and 128 samples . Fig # 48 illustrates BT spectrum . The results will be improved only at large number of autocorrelation lags .

Burg's method fails completely irrespective of model order as shown in Fig#49. Cadzow's method gives some reasonable results at different SNR's and different number of samples . The results are given in Fig # 50 .

The estimated PSD's via SHM are shown in Fig # 51 . They give good results at different signal to noise ratios (SNR = 30 , 20 , and 10 dB) and different number of samples (N = 32 and 16) . When N = 8 , the method fails completely .

Fig # 52 shows the estimated PSD's via Covariance approximation method . PSD was estimated for different model orders . This method does not resolve the two frequencies .

MUSIC method gives good results and resolves the two frequencies only at (SNR = 30 dB) and different number of samples (N = 32 and 16) . The simulated results are shown in Fig # 53 .

The results obtained by OHM are as good as SHM results . They are shown in Fig # 54 for different number of samples , different number of SNR's , and different model orders .

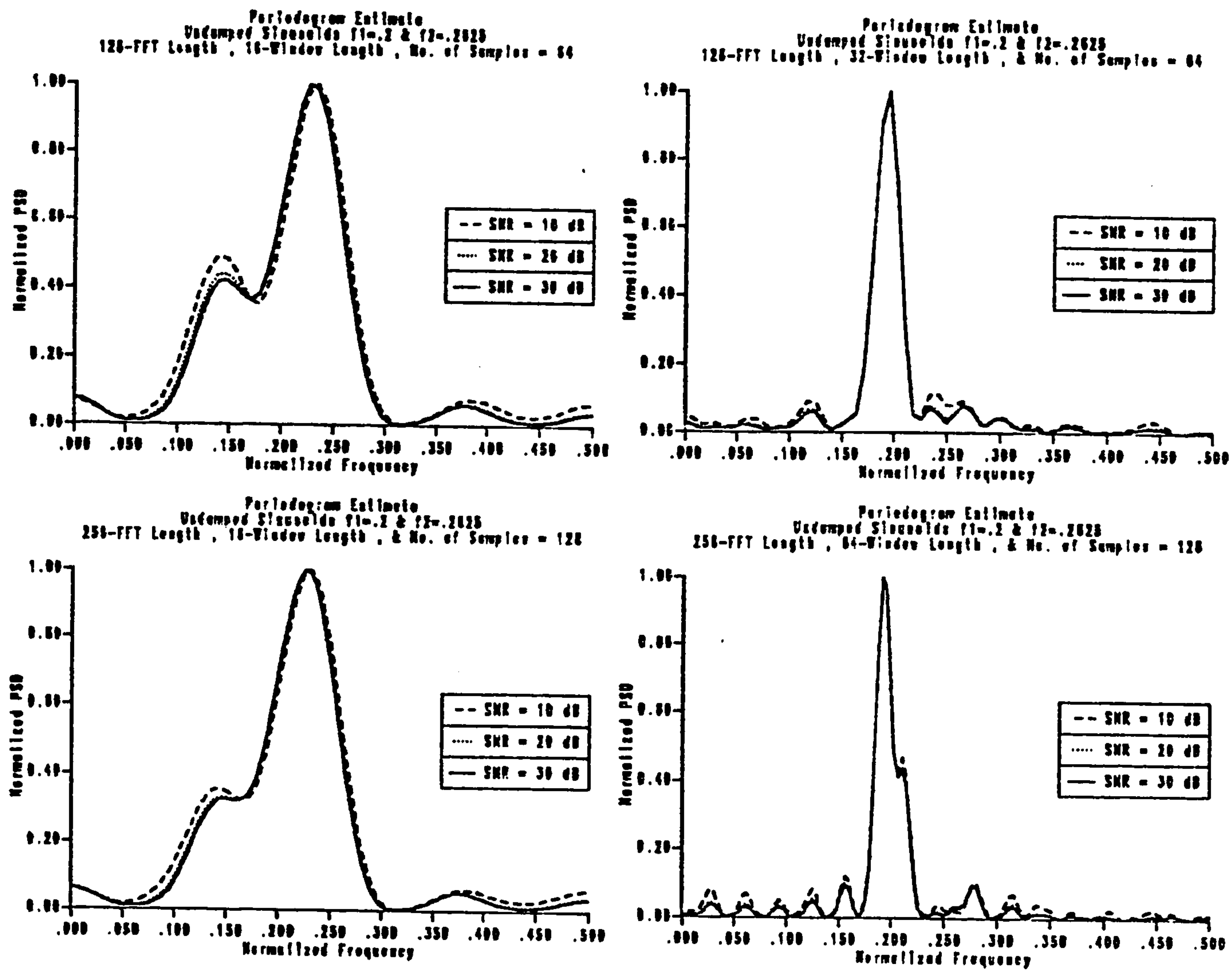

Figure 47: Periodogram PSD's of undamped sinusoids (close frequencies)

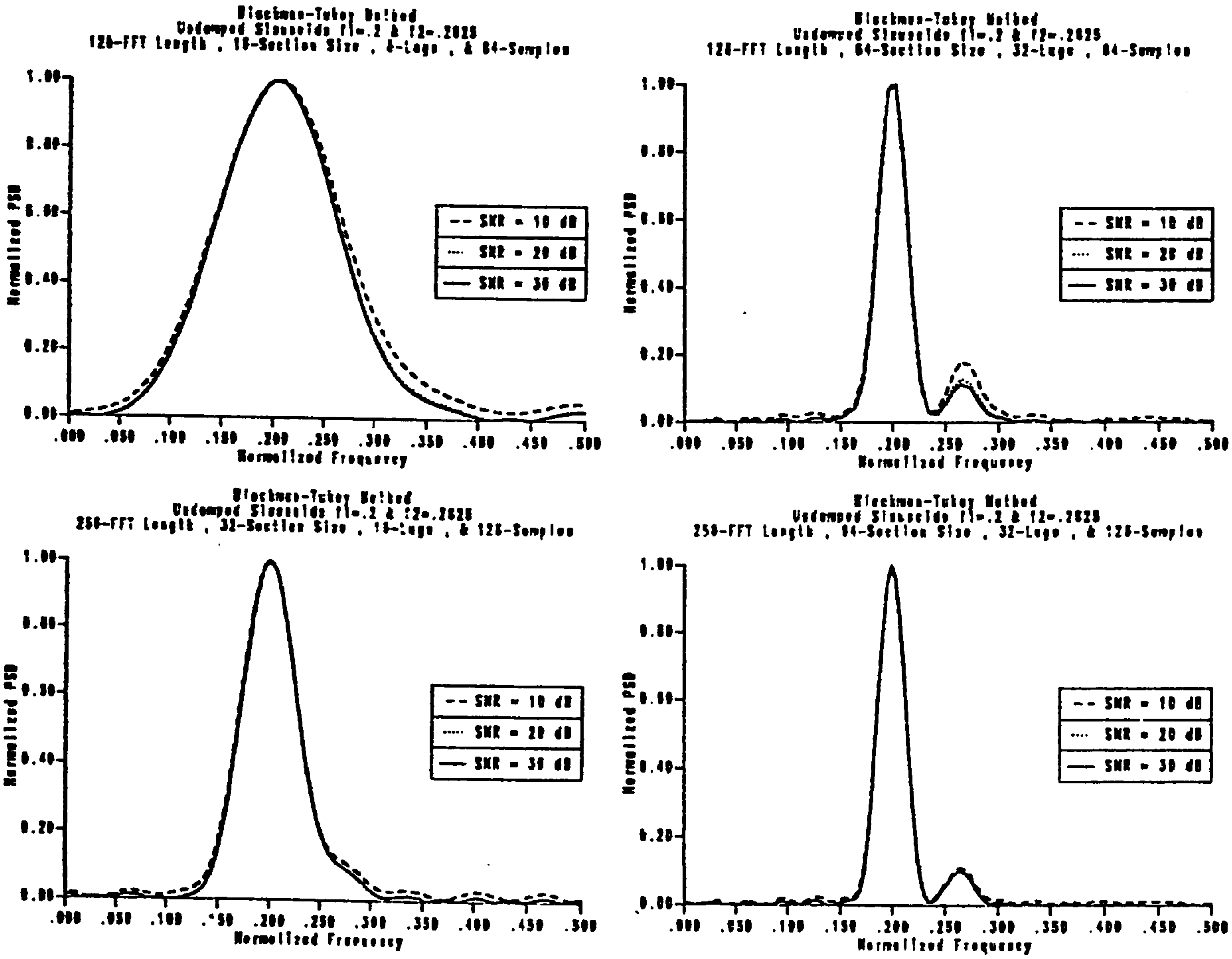

Figure 48: BT PSD's of undamped sinusoids (close frequencies)

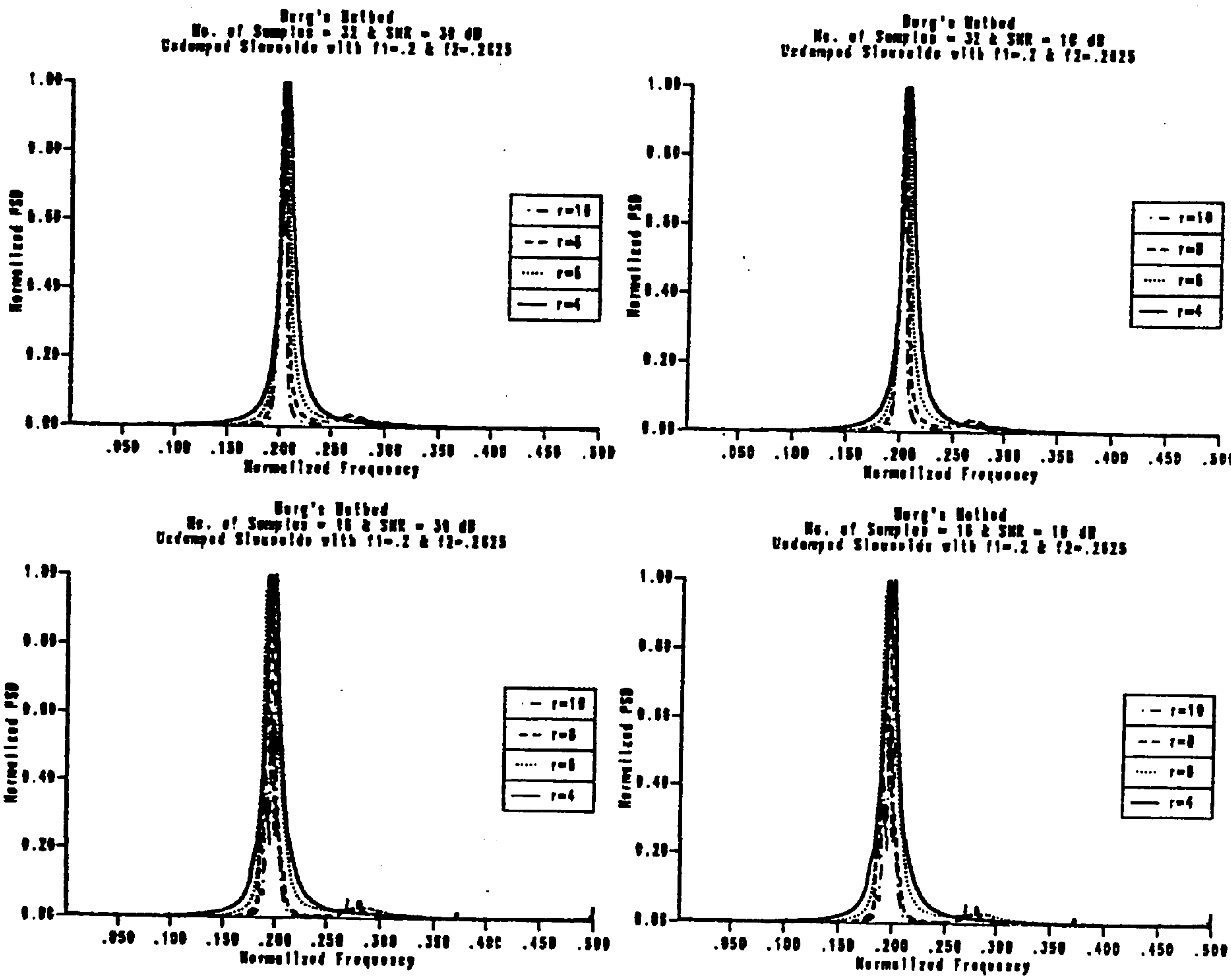

Figure 49: Autoregressive PSD's via Burg algorithm (undamped with close frequencies)

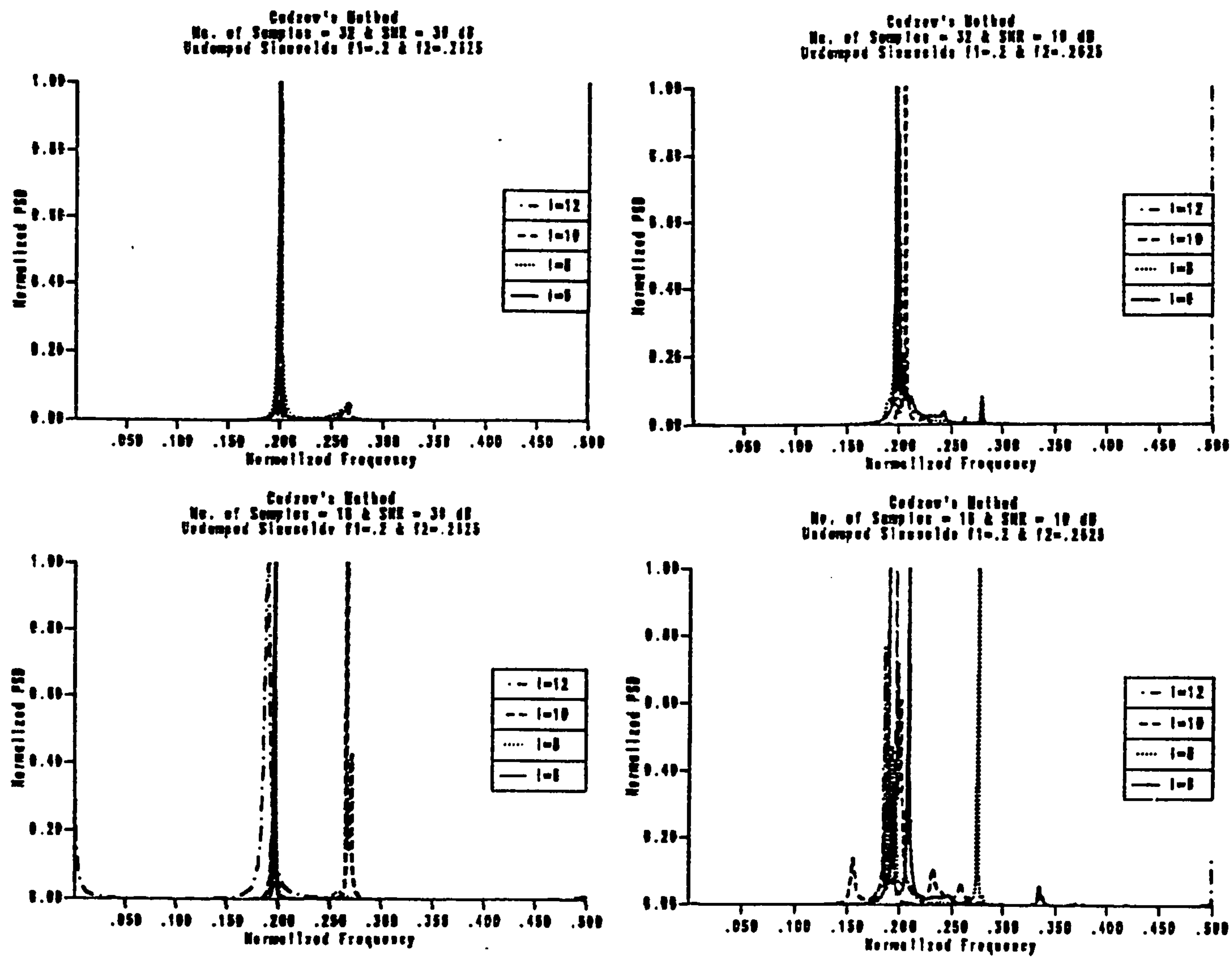

Figure 50: ARMA modeling of undamped(close freq.) via Cadzow approach

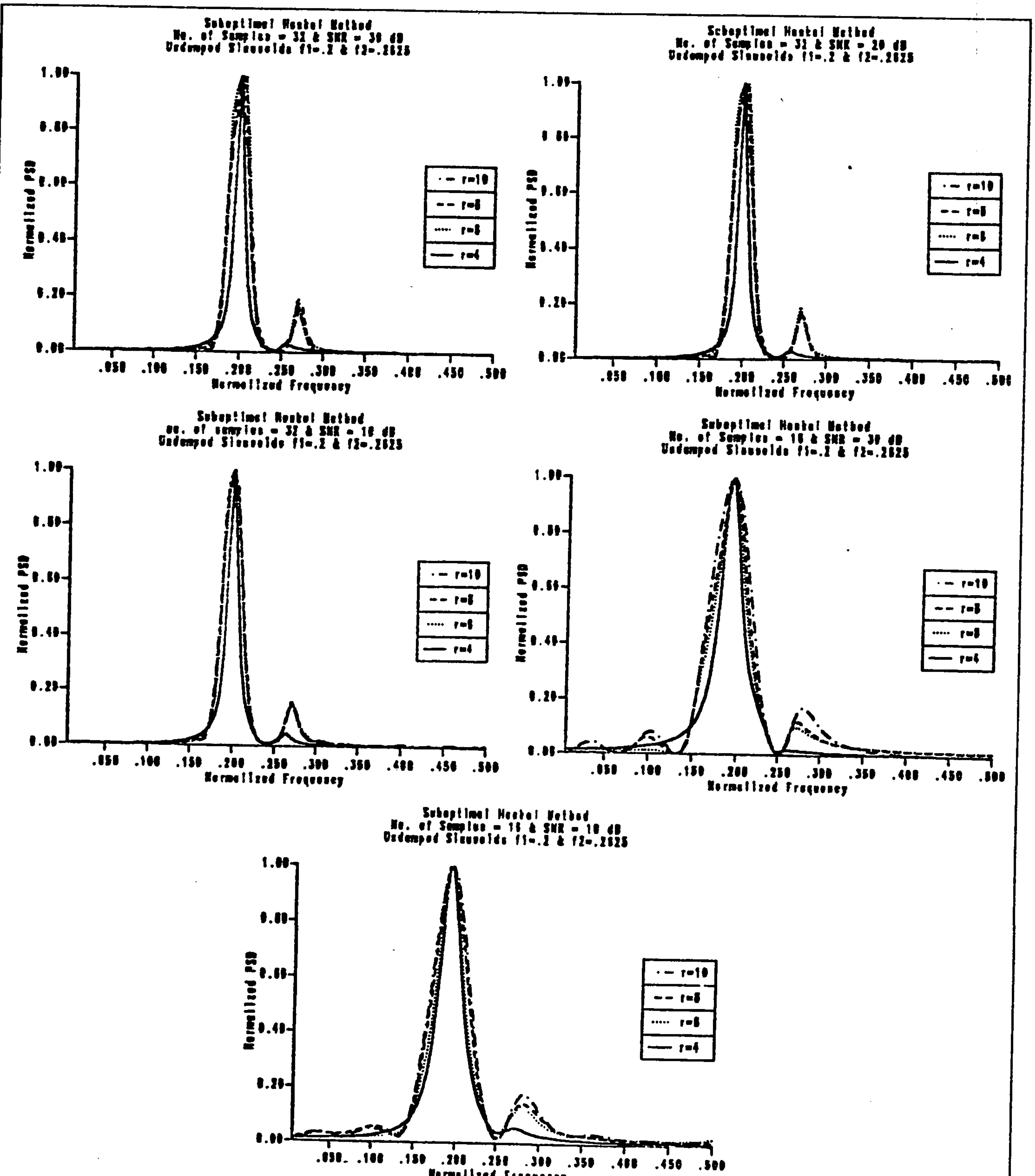

Figure 51: PSD's of undamped sinusoids with close freq. via SHM

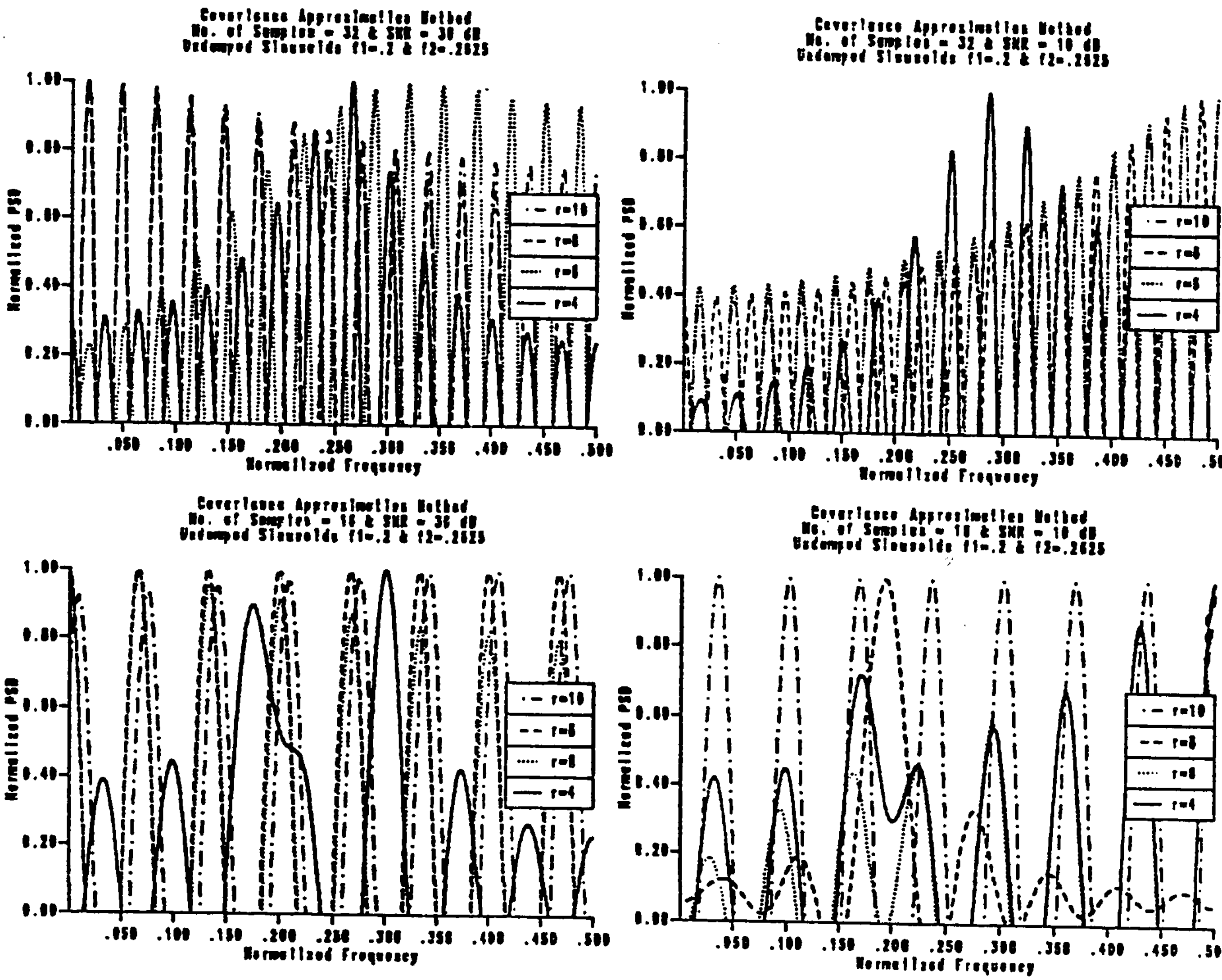

Figure 52: Covariance Approx. method of undamped (close freq.)

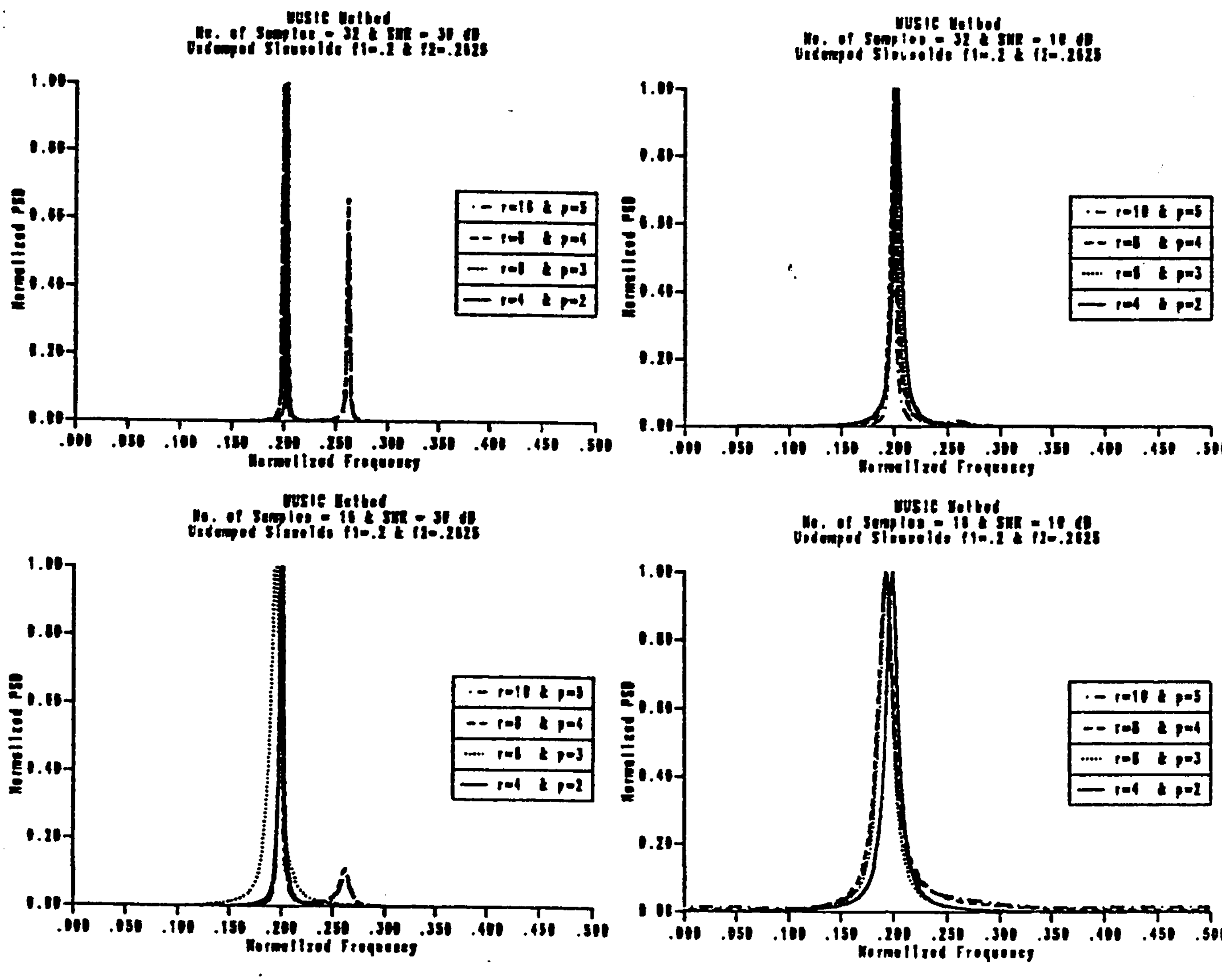

Figure 53: MUSIC estimates of undamped sinusoids with close freq.

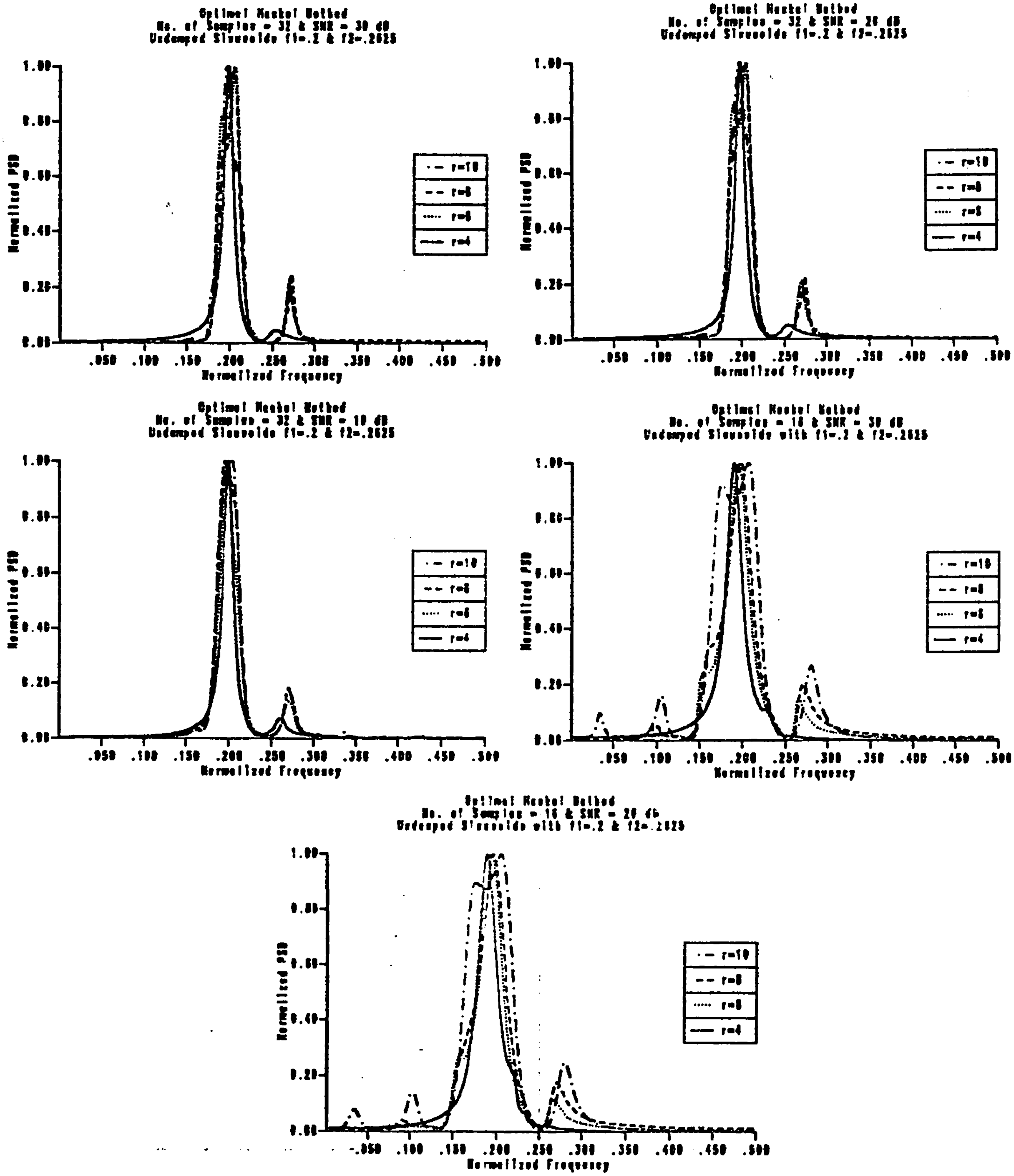

Figure 54: OHM estimates of undamped sinusoids with close freq.

4.4.3 Far Frequencies

In this case the two frequencies ($f_1 = 0.2\, Hz$ and $f_2 = 0.45\, Hz$) are widely seperated , where $\Delta f \geq \frac{1}{N}$ and N = 32 , 16 , and 8 .

For short data length , the traditional methods fail to resolve the two frequencies . Increasing the number of samples , both of which will resolve these two frequencies . Burg's method does not show sharp peak at f = 0.45 Hz .

Cadzow's method gives some very sharp peaks at different variants . The obtained results are very good even for low SNR's , but it fails for smaller number of samples (N = 16) .

SHM shows good results at different signal to noise ratios (SNR = 30 , 20 , 10 dB) different number of samples (N = 32 and 16) , and different model orders (r = 4 , 6 , 8 , 10) .

Covariance approximation method does not resolve the right frequencies . MUSIC method is very sensetive to noise . It gives good estimation only at SNR = 30 and different number of samples (N = 32 and 16)

OHM shows good results at different signal to noise ratios (SNR = 30 , 20 , 10 dB) , different number of samples (N = 32 and 16) , and different model orders (r = 4 , 6 , 8 , 10) .

CHAPTER 5

FREQUENCY ESTIMATION

In this chapter three different methods will be presented . These three methods will estimate directly the frequencies without going through spectral estimation .

5.1 FREQUENCY ESTIMATION TECHNIQUES

5.1.1 TK Method Based on Backward Linear Prediction (BLP) Technique:

Tufts and Kumaresan (TK) proposed the parametric model to estimate the frequencies of the received samples . This method models the system by ARMA model . It has the advantage of high accuracy and good estimator for number of sinusoids . There are two primary assumptions to this model :

1. The original signals are exponentially damped sinusoidals
2. The noise associated with the signal is additive white Gaussian noise with zero mean

The Tufts and *Kumaresan* (*TK*) *algorithm* can be summarized as follows :

1. Form the data matrix as follows :

$$A = \begin{bmatrix} y_2 & y_3 & \cdots & y_{N-L+1} \\ y_3 & y_4 & & \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_{L+1} & y_{L+2} & \cdots & y_N \end{bmatrix}$$

where y_n is the sample of the noisy signal, N is the number of received samples . The dimension of the matrix is $(N-L)\times L$. The number L represents the number of rows of A and it ranges between $M \le L \le (N-M)$. M is twice number of damped sinusoids .

Define

$$X = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_{N-L} \end{bmatrix}$$

2. Compute the singular values of A

$$A = U \Sigma V^T$$

where U and V are orthogonal square matrices and the matrix Σ is a diagonal matrix containing the singular values of A .

3. From the SVD, we can estimate accurately the number of exponentially damped sinusoids . Taking the largest M values and solving for b in :

$$A b = -X$$

where b can be found as

$$b = - [\sum_{i=1}^{M} \sigma_i^{-1} V_i U_i^T] X$$

where V_i represents the i^{th} column of V and U_i represents the i^{th} row of U .

4. The entries vector b are used as coefficients of the polynomial:

$$a(z) = z^L + b_1 z^{L-1} + \ldots + b_L$$

Tufts and Kumaresan have shown that in the absence of noise, the polynomial a(z) has M roots outside the unit circle and the extra roots are inside the unit circle . From the roots that are outside the unit circle we can estimate the frequencies of the noisy signal .

Also, they have used the Forward-Backward Linear Prediction Technique (FBLP) for estimating the frequencies of undamped sinusoids . This algorithm has been tested extensively by Kumaresan [69] and it will not be included .

5.1.2 Pisarenko Method:

Pisarenko proposed the parametric model to estimate the frequencies and line spectrum of the received samples . There are two primary assumptions to this model :

1. The original signal are undamped sinusoids

2. The noise associated with the signal is additive white Gaussian noise with zero mean

Pisarenko algorithm can be summarized as follows :

1. Form the autocorrelation matrix B as follows :

$$B = \begin{bmatrix} r(0) & r(1) & \dots r(m) \\ r(-1) & r(0) & \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r(-m) & r(-m+1) & \dots r(0) \end{bmatrix}$$

where r_i's are the autocorrelation lags defined in equation (2.3) . Then, find the rank of matrix B and it will be the number of the sinusoids in the signal .

2. Find the minimum eigenvalue and denote by λ_o

3. Find the eigenvector associated with λ_o and denote the components by :

$a_0 , a_1 , \dots , a_M$

4. Find the zeros of the characteristic polynomial generated by the minimum eigenvalue λ_o :

$$a_M z^M + \dots + a_1 z + a_0 = 0$$

where M is twice number of sinusoids .

5. Determine the frequencies of the sinusoids

Pisarenko's algorithm has some difficulties associted with it . His algorithm does not utilize the data effectively to obtain accurate estimates . Also, the algorithm is only applicable for undamped sinusoids which is because Pisarenko forced the autocorrelation matrix to be Toeplitz . Furthermore, the estimated frequencies are biased which is due to the same reason mentioned . To over come the aforementioned problems, Tufts and Kumaresan [69] proposed a new improved Pisarenko algorithm . This improved algorithm will be applicable for both damped and undamped sinusoids .

5.1.3 Suboptimal Hankel Method:

The algorithm of this method is given in chapter two and it will be repeated here in brief :

1. Form the finite Hankel matrix given in equation (2.44)
2. Perform SVD of the $N \times N$ Hankel matrix
3. Find A given in equation (2.47b)
4. Find the eigenvalues of A
5. Determine the frequencies from the computed eigenvalues

5.2 COMPUTER SIMULATION RESULTS

The results were tested by varying different parameters of each model . Also, the frequency resolution was tested in the three algorithms .

5.2.1 TK & Suboptimal Hankel Method Simulation Results:

The testing is given by :

$$y_n = e^{-0.1n}\cos 2\pi f_1 n + e^{-0.2n}\cos 2\pi f_2 n + e_n$$

where, e_n is white Gaussian noise with zero mean and variance σ^2 . Three different cases have been studied :

1. $f_1 = 0.2\ Hz$ and $f_2 = 0.23125\ Hz$

The two frequencies are very close $\Delta f \le \frac{1}{N}$. The TK method with different number of samples, different SNR's, and different values of (L) fails completely to estimate the second frequency as shown in tables 2 and 3 . While, The Suboptimal Hankel Method (SHM) gives reasonable results in propotional to TK method only at N = 32 and SNR = 30 dB as shown in table 1 .

2. $f_1 = 0.2\ Hz$ and $f_2 = 0.2625\ Hz$

The two frequencies are close . In this case, both of which give good results at only high SNR's and N = 32 . TK method gives better estimation as the number of samples increases . The value of L should be between $M \le L \le N-M$, and L is in this range then we would get acceptable results as shown in tables 5 and 6 . SHM gives better estimation than TK as shown in table 4 .

3. $f_1 = 0.2\ Hz$ and $f_2 = 0.45\ Hz$

In this case, TK method gives good results only at (SNR = 30, and 20 dB), (N = 32, and 16) and a proper choice of L (L = 8). While the SHM gives better results for different signal to noise ratios and (N = 32, and 16) .

Table 1: SHM estimation for very close frequencies

N	SNR (dB)	f1=0.2	f2=.23125
32	30	0.199662	0.235439
	20	0.072440	0.191132
	10	0.025154	0.141342
16	30	0.151242	0.202124
	20	0.128506	0.197957
	10	0.025689	0.101871

Table 2: TK estimation for very close frequencies (N = 32)

N	SNR(dB)	L	f1=0.2	f2=.23125
	30	6	0.206799	-
	30	10	0.205030	-
	30	14	0.204372	-
	30	18	0.204372	-
	20	6	0.206718	-
32	20	10	0.204910	-
32	20	14	0.204192	-
	20	18	0.161772	-
	10	6	0.206377	-
	10	10	0.176954	-
	10	14	0.181607	-
	10	18	0.055147	-

Table 3: TK estimation for very close frequencies (N = 16)

N	SNR(dB)	L	f1=0.2	f2=.23125
	30	6	0.207509	-
	30	8	0.201745	-
	30	10	0.194114	-
	30	12	0.221537	-
	20	6	0.207412	-
16	20	8	0.182421	-
16	20	10	0.147886	-
	20	12	0.159408	-
	10	6	0.206824	-
	10	8	0.186405	-

Table 4: SHM estimation for close frequencies

N	SNR (dB)	f1=0.2	f2=.2625
	30	0.200073	0.262306
32	20	0.199768	0.262944
	10	0.185112	0.404123
	30	0.193846	0.273953
16	20	0.192857	0.272816
	10	0.160226	0.375523
8	30	0.200501	0.412333

Table 5: TK estimation for close frequencies (N = 32)

N	SNR(dB)	L	f1=0.2	f2=.2625
	30	6	0.214560	0.248493
	30	10	0.199889	0.262698
	30	14	0.199902	0.262608
	30	18	0.181992	0.253208
	20	6	0.205009	-
32	20	10	0.173041	0.255857
32	20	14	0.199085	0.246004
	20	18	0.171666	0.225704
	10	6	0.205731	-
	10	10	0.174946	-
	10	14	0.210995	-
	10	18	0.232276	-

Table 6: TK estimation for close frequencies (N = 16)

N	SNR(dB)	L	f1=0.2	f2=.2625
	30	6	0.228444	0.234276
	30	8	0.229526	0.233042
	30	10	0.166070	0.227644
	30	12	0.226771	0.235732
	20	6	0.141677	0.212890
16	20	8	0.156484	0.224636
16	20	10	0.170555	0.243021
	20	12	0.210342	0.228341
	10	6	0.189882	-
	10	8	0.195688	-
	10	10	0.208800	-
	10	12	0.200684	-

5.2.1 Pisarenko & Suboptimal Hankel Method Simulation Results:

The testing is given by :

$$y_n = \cos 2\pi f_1 n + \cos 2\pi f_2 n + e_n$$

where, e_n is white Gaussian noise with zero mean and variance σ^2 . Three different cases have been studied :

1. $f_1 = 0.2\ Hz$ and $f_2 = 0.23125\ Hz$

The two frequencies are very close $\Delta f \leq \frac{1}{N}$. In this case, SHM gives good results only at N = 32 as shown in table 7 . The estimation of frequencies gets better for higher SNR's which is expected until the noise dominates the frequency components of the signal then, the estimation is no longer valid . Pisarenko method gives better estimation for smaller number of samples (N = 16) as shown in table 8 .

2. $f_1 = 0.2\ Hz$ and $f_2 = 0.2625\ Hz$

In this case, Pisarenko method gives better estimation than SHM estimation . This due to the assumption of pending with zeros is no longer valid . The results are shown in tables 9 & 10 .

3. $f_1 = 0.2\ Hz$ and $f_2 = 0.45\ Hz$

In this case, Pisarenko method gives very good results and at different signal to noise ratios (SNR = 30, 20, and 10 dB) and different number of samples (N = 32, 16, and 8) . While SHM gives acceptable results only at N = 32 and different SNR's .

Table 7: SHM estimation for very close frequencies (undamped)

N	SNR (dB)	f1=0.2	f2=.23125
32	30 20 10	0.198988 0.198950 0.198459	0.231836 0.231640 0.230588
16 16	30 10	0.157754 0.132544	0.203414 0.195055
8	30	0.102527	0.198579

Table 8: Pisarenko estimation for very close frequencies (undamped)

N	SNR (dB)	f1=0.2	f2=.23125
32	30 20 10	0.207694 0.207377 0.201207	0.284588 0.286099 0.291294
16 16	30 10	0.187221 0.181676	0.248737 0.260524
8	30	0.161880	0.251385

Table 9: SHM estimation for close frequencies (undamped)

N	SNR (dB)	f1=0.2	f2=.2625
32	30 20 10	0.200055 0.200017 0.199629	0.318348 0.451018 0.486117
16 16	30 10	0.192747 0.163607	0.292260 0.421766
8	30	0.190698	0.419741

Table 10: Pisarenko estimation for close frequencies (undamped)

N	SNR (dB)	f1=0.2	f2=.2625
32	30 20 10	0.209163 0.209070 0.205802	0.292338 0.292930 0.294400
16 16	30 10	0.159201 0.160050	0.267185 0.268910
8	30	0.181715	0.281364

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

6.1 CONCLUSION

The most popular traditional and modern power spectrum estimation techniques have been briefly reviewed . These methods are optimal and suboptimal with respect to different criteria . Also, it has been shown how signal processing has developed recently to model the data with either all-zeros (MA), all-pole (AR), or a combination of both poles and zeros (ARMA) . A new algorithm developed in the context of model reduction by Bettayeb [5,53] was applied to the problem of power spectral estimation . The implementation of the above algorithms on typical simulation examples showed the performance of the various techniques . All algorithms have been systematically tested on various different examples of signals with colored noise containing pure or decayed sinusoids . The simulation studies were done for different data length, different signal to noise ratios, and for varied frequency resolution .

Both of the traditional methods are computationally efficient if the Fast Fourier Transform (FFT) is used in the direct approach or if only a few lags are needed in Blackman-Tukey approach . In both of them , the estimated power spectral density (PSD) is directly propotional to the power for sinusoid processes . Also, they can be cosidered as good models for some applications . On the other hand, the weak signal main- lobe responses are supperessed by strong signal side lobes . Also, the sidelobe leakage will introduce some distortion in the spectrum . Furthermore, some negative PSD values appear with the BT approach . The frequency resolution is limited by the available data record duration . Two main factors affect the spectral resolution of the spectral estimate . The first is the record length of the data and the autocorrelation lags . The second is the windowing process applied to autocorrelation function and the data . The spectral resolution usually increases as the record length increases . But, this is not always, e.g., the periodogram of white noise becomes oscillatory [67] . Also, the spectral resolution is decreased if the window is designed to reduce the sidelobe or leakage in spectral estimate domain . The increase in spectral resolution can be achieved by pending a sequence of zeros of the windowed autocorrelation function and data prior to transforming to the frequency domain .

Most of the modern methods do not suffer from the problems associated with the traditional methods . They can achieve increased spectral resolution, especially, for short data records and low signal to noise ratios . Also, their spectral estimates do not have strong sidelobes . The autoregressive models based on Yule-Walker equations are related to linear prediction analysis and adaptive filtering . They give better results than the conventional methods, but not as good as the other AR methods . However, the implied windowing distorts the spectrum and the line splitting frequently occurs . In Burg's maximum entropy method, spectrum splitting and frequency shifting frequently occur, especially for sinusoidal signals . In general, MEM enhances the peaky components of the spectrum . However , while an AR model may give a good spectrum estimate, the coefficients may not predict the data well beyond a few points . Thus , the coefficients are needed to be more frequently updated for some kinds of data and less frequently for other kinds of data . Generally , all AR methods suffer from selecting the proper model order .

Although the ARMA model might give better resolution and performance in spectral estimation for different applications, many practitioners prefer to use either AR or MA models . This is because it is more computationally involved . It is required to compute the order of the model and the coefficients in both the denominator and numerator . Pisarenko's algorithm and MUSIC algorithm are known to give superior resolution, yielding estimates that are asymptotically unbiased and efficient . Pisarenko's method is computationally inefficient and it does not work well in high noise levels .

Since the available methods are not applicable for all types of signals, our developed algorithm has a wider scope of applications as it treats indeferently signals derived from AR or ARMA processes . The proposed optimal Hankel method gave accurate power spectral estimates, and compared favorably against well known methods . Also, the algorithm has good robustness properties against noise and computational errors as it is based on the robust singular value decomposition (eigenvalue decomposition of a symmetric matrix) [29] .

6.2 SUGGESTIONS FOR FURTHER WORK

The following can be considered as good points for further research:

1. Direct optimal spectrum estimation.

   In our work, minimization was done for the infinite norm of the difference between the exact transfer function and the transfer function obtained using optimal Hankel approach. Instead, in the minimization...

of the infinite norm we could consider the difference between the exact spectral density and the spectral density obtained by an estimated model.

2. Spectrum error bounds.

For the optimal Hankel approach model, derived tight bounds on error spectral estimates will be useful to evaluate the degree of the suboptimality of the approach considered.

3. Order selection.

The order selection plays a significant role in the clarity of the contents of the spectrum. Hence, we suggest for future work that a criterion for proper selection of order of our model be developed. The singular values tools look promissing in this regard.

4. Computational reduction.

This work was mainly concerned with the degree of accuarcy of estimation, rather than the computational complexity . Although computational complexity was not investigated, but it is expected that this method has a reasonable one. Therefore, it is suggested to further investigate this point in future work. For example, one can save in computations if only a partial number of eigenvalues of the Hankel matrix is computed as only the $(r+1)^{st}$ eigenvalue is needed.

5. Multichannel optimal Hankel spectrum estimation.

Our work was restricted to single input single output system. For certain applications, such as image processing, our work could be extended to include multichannel systems.

APPENDIX A

APPLICATIONS

Several applications of spectral estimation methods will be presented covering radar, sonar, speech processing, and others .

In radar applications, it is a common practice that a transmitted signal is always reflected by a target and other surrounding objects like the ground, migrating bird flocks, and weather disturbances . The traditional objective here is to measure range to a target . This is done by measuring time difference between transmission and receiption of the signal (Δt) and converting the analysis from time processing to frequency processing . Consequently, over a certain time overlap (Δt), there will be a constant frequency due to a certain object . This constant frequency is known as the beat frequency (f_R) . Power spectral estimation is employed here to find the beat frequencies and then by using :

$$range\,(R) = f_R \frac{TC}{2\,BW}$$

the range will be computed, where T is the pulse width and BW is the radar pulse bandwidth, and C is the speed of light .

Another radar application involves investigating target's velocity . The range rate of the target can be located by the main lobe of each PSD during a given processing interval .

One last radar application involves targert identification . The spectral content of the reflected signals helps in determining physical characteristics of the targert . Since the unwanted objects usually share some previously expected power spectral distribution, they could be filtered out.

In speech processing, the estimation of the power spectrum will lead to the determination of the pitch and formant frequencies, which have found extensive use in speech encoding, synthesis, and recognition

Also, bandwidth compression is an important problem in speech processing . Reducing the redundancy of speech, more speech signals can be sent through a fixed bandwidth . The basis of differential pulse code modulation (DPCM) is to send only information that can not be predicted . If the speech waveform were completely predictable from past samples, then the receiver, once it had these samples, would be able to reconstruct the complete waveform . In practice, linear prediction coding is used for bandwidth reduction, in which all- pole modelling is used for representing the speech waveform [29] . If speech can be accurately modeled as the output of an all-pole filter driven by an impulse train for voiced speech, or driven by white noise for unvoiced speech, then the speech waveform could be reduced to a set of parameters .

In speech encoding, vocal tract simulation and other applications, an autoregressive model appears to be a good model of the waveform (Atal and Hanauer 1971) . This model is good for modeling strong resonances . Sinusoids can be considered as strong resonances . As a result at high signal to noise ratio (SNR), they can be modeled by all pole model . While, at low SNR, the all pole model of the signal containing both sinusoids and additive noise will give poorer resolution .

In sonar, where one has to determine the presence and location of the objects in the sea by using underwater sound, the frequency of the received signals is found by the estimation of the power spectrum . This will lead to the determination of the position of the target .

Also, power spectrum estimation is important in determining the bandwidth of random signals and in the subsequent design of filters to pass or reject those signals . Furthermore, it has played an important role in geophysics especially in seismic studies such as petroleum exploration, nuclear detection, earthquake research, and marine seismic studies . In some of the above applications, only a few samples are available but a high resolution spectral analysis is required . Therefore, the proper choice of seismic signal processing techniques is important. The ARMA modelling is considered as the most general linear seismic signal model that provides simple parametric representation of the signals . In theory, the spectrum of any physical signal can be fitted perfectly by an AR model with a proper choice of a model order . But, if the ARMA model is used, a better spectral matching can be achived . The power spectrum estimation is useful in differentiating between a man-made explosion and an earthquake . Also, the parameters of the model are very useful in classifying the teleseismic events . Furthermore, good spectral estimation gives accurate computation of spectral ratio, which is useful in seismic discrimination .

High resolution spatial spectrum analysis using spectral methods has been applied to sonic well logging. The sonic waves are received by the array of receivers in a sonic device. The velocities of the various components of the received waveforms at each receiver can determine properties of the rock through which the sonic wave travels. Wave components which propagate at lower velocities reach later in time and often overlap earlier wave arrivals. Since the spatial extent of the array aperture is only with limited number of receivers, high resolution spectral estimation is required for the spatial dimension of signal processing.

The procedure is to estimate the spatial frequencies present in the complex-valued sequences $S(f,n)$ using any spectral estimation method, from which wave velocities may be deduced. $S(f,n)$ is obtained by calculating, at a single temporal frequency f, the DFT of the n^{th} received waveform, $s(t,n)$, after a windowing procedure has been excited to attenuate interfering signals. The estimation of the spatial frequency leads to the estimation of the velocity. The estimates of temporal and spatial frequencies for various times used in clustering procedure to identify individual wave components.

BIBLIOGRAPHY

1. K.S. Arun and Bhaskar Rao, "An improved Toeplitz approximation method", IEEE. ICASSP , pp.2352-2355, 1988.
2. K.S. Arun,"Principal components algorithms for ARMA spectrum estimation", IEEE Trans. ASSP ,Vol.37, No. 4, pp.566-572, Apr 1989.
3. A.A. Beex and L. Scharf ,"Covariance sequence approximation for parametric modeling", IEEE Trans. ASSP ,Vol.29, No. 5, pp.1042-1052, Oct 1981.
4. --,"Recursive digital filter design via covariance sequence approximation", IEEE Trans. ASSP ,Vol.29, No. 1, pp.51-57, Feb 1981.
5. M. Bettayeb,"Approximation of linear systems: new approaches based on singular value decomposition", Ph.D. dissertation, Dept. E.E., Southern California University, 1981.
6. D.V. Bhaskar and S. Y. Kung ,"Adaptive notch filtering for the retrieval of sinusoids in noise", IEEE Trans. ASSP ,Vol.32, No. 4 ,pp.791-802,Aug 1984.
7. D.V. Bhaskar ,"Analysis of coefficient quantization errors in state -space digital filters", IEEE Trans. ASSP ,Vol.34, No. 1 ,pp.131-139,Feb 1986.
8. --,"An analysis of the Prony-Kumaresan method" , IEEE. ICASSP ,pp.467-471, 1987.
9. --,"Sensitivity analysis of state-space methods in spectrum estimation", IEEE. ICASSP ,pp.1517-1520, Apr. 1987.
10. J.P.Burg,"Maximum entropy spectral analysis",in Proc. 37th Meeting Society of Exploration Geophysicists, Oct. 31,1967.

11. --,"A new anaylsis technique for time series data",NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, pp.12-23, Aug. 1968.

12. --,"The relationship between maximum entropy and maximum likelihood spectra", Geophys. Vol. 37, pp.375-376, Apr. 1972.

13. J.A. Cadzow,"Spectral estimation: an over-determined rational model equation approach",Proc. IEEE, Vol. 70 , No. 9, pp.907-938 , Sep. 1982.

14. J. Capon,"High-resolution frequency-wavenumber spectrum analysis", Proc. IEEE, Vol. 57 , No. 8, pp.1408-1419 , Aug. 1969.

15. C. Chatterjee,R.L. Kashyab, and G. Boray,"Estimation of close sinusoids in colored noise and model descrimination",IEEE Trans. ASSP, Vol. 35, No. 3, pp.328-336 , Mar. 1987.

16. E.R. Dowski,C.A. Whitemore, and S.K. Avery,"Estimation of randomly sampled sinusoids in additive noise ",IEEE Trans. ASSP, Vol. 36, pp.1906-1908 , 1988.

17. B. Friedlander,"Lattice methods for spectral estimation " Proc. IEEE, Vol. 70, No. 9, pp.990-1017,Sep. 1982.

18. J.J. Fuchs,"Estimating the number of sinusoids in additive white noise",IEEE Trans. ASSP, Vol. 36,No. 12, pp.1846-1853 , Dec.1988.

19. G.H. Golub and C. Reinsch,"Singular value decomposition and least squares solutions",Numer. Math. 14,pp.403-420, 1970.

20. S. Haykin,"Communication Systems", Jhon Wiley & Sons Inc., New York, 1983.

21. Y. Hua and T. Sarkar,"Perturbation analysis of TK method for harmonic retrieval problems",IEEE Trans. ASSP, Vol. 36,No. 2, pp.228-240 , Feb.1988.

32. A.C. Kot, S. Parthasarthy ,D.W. Tufts ,and R.J. Vaccaro,"the statistical performanceof state-variable balancing and Prony's method in parameter estimation", IEEE ICASSP, pp.1549-1552,1987.

33. R. Kumaresan and D.W. Tufts,"Data-adaptive principal component signal processing", in Proc. 19th IEEE Int. Conf. Decision and Contr. ,pp.949-954,1980.

34. ,"Estimating the parameters of exponentially damped sinusoids and pole-zero modelling in noise",IEEE Trans. ASSP, Vol. 30, No. 6, ,pp.833-840,Dec. 1982.

35. ,"Improved spectral resolution II", IEEE ICASSP, pp.592-597,1980.

36. ,"Improved spectral resolution III: efficient realization",Proc. IEEE, Vol. 68, pp.1354-1355, Oct. 1980.

37. S.Y. Kung,K.S. Arun, and D.V. Bhaskar Rao,"State-space and singular value decomposition based approximation methods for the harmonic retrieval problem", Opt. Soc. of Amer., Vol. 73, No. 12, pp.1799-1811, Dec. 1983.

38. S.Y. Kung,"A new identification and model reduction algorithm via singular value decomposition", IEEE ICASSP, pp.705-714, 1979.

39. S.Y. Kung and Y.H. Hu,"Improved Pisarenko's sinusoidal spectrum estimate via SVD subspace approximation methods", IEEE ICASSP, pp.1312-1314, 1982.

40. R.T. Lacoss,"Data adaptive spectral analysis methods", Geophysics., Vol. 36, No. 4, pp.661-675, Aug. 1971.

41. S.L. Marple, Jr.,"Digital Spectral Analysis with Applications", Prentice-Hall Inc., New Jersey, 1987.

42. A.V. Oppenheim and R.W. Schafer,"Digital Signal Processing", Prentice-Hall Inc., New Jersey, 1975.

43. M.D. Ortiguerra and J.M. Tribolet,"On the evaluation of spectral analysis techniques", Elseiver Science Publishers B.V. (North Holland), Digital Signal Processing, pp.117-120, 1984.

44. A. Papoulis,"Maximum entropy estimation: a review", IEEE Trans. ASSP, Vol. 29, No. 6, pp.1176-1186, Dec. 1981.

45. V.F. Pisarenko,"The retrieval of harmonics from a covariance function",Geophysical J. Royal Astro. Soc., Vol. 33, pp.347-366, 1973.

46. B. Porat and B. Friedlander,"A modification of Kumaresan-Tufts method for estimating rational impulse responses", IEEE Trans. ASSP, Vol. 34, No. 5, pp.1336-1342, Oct. 1986.

47. ,"Computation of the exact information matrix of Gaussian time series with stationary random components", IEEE Trans. ASSP, Vol. 34, No. 1, pp.118-130, Feb. 1986.

48. ,"On the accuracy of the Kumaresan-Tufts method for estimating complex damped exponentials", IEEE Trans. ASSP, Vol. 35, No. 2, pp.231-235, Feb. 1987.

49. J.E. Purviance,"Determining if noise corrupted matrix is singular ",Proc. IEEE, pp.430-431, 1983

50. S.S. Rao and D.C. Gnanaprakasm,"A criterion for identifying dominant singular values in the SVD based method of harmonic retrieval" ,IEEE ICASSP, pp.2460-2463, 1988.

51. H. Sakai,"Statistical analysis of Pisarenko's method for sinusoidal frequency estimation", IEEE Trans. ASSP, Vol. 32, No. 1, pp.95-101, Feb. 1984.

52. R.O. Schmidt,"Multiple emitter location and signal parameter estimation", IEEE Trans. Antenna and Prop., Vol. AP-34, No. 3, pp.276-290, Mar. 1986.

53. L.M. Silverman and M. Bettayeb,"Optimal approximation of linear systems",Proc. JACC San Franscisco, CA, 1980.

54. B.J. Sullivan and B. Liu,"On the use of singular value decomposition and decimtion in discrete time band limited signal extrapolation", IEEE Trans. ASSP, Vol. 32, No. 61, pp.1201-1212, Dec. 1984.

55. D.W. Tufts and R. Kumaresan,"Estimation of multiple sinusoids: making linear prediction perform like maximum likelihood", Proc. IEEE, Vol. 70, No. 9, pp.975-989, Sep. 1982.

56. D.W. Tufts and C.D. Melissinos,"Simple , effective computation of principal eigenvectors and their eigenvalues and application to high- resolution estimation of frequencies", IEEE Trans. ASSP, Vol. 34, No. 5, pp.1046-1053, Oct. 1986.

57. T.J. Ulrych and T.N. Bishop,"Maximum entropy spectral analysis and autoregressive decomposition",Rev. Geophysics space Phys., Vol. 13, pp.183-200, Feb. 1975.

58. R.J. Vaccaro,"Modeling of perturbed covariance sequences",Proc. IEEE , Vol. 74, No. 4, pp.617-619, Apr. 1986.

59. A. Van den Bos,"Alternative interpretation of maximum entropy spectral estimate",IEEE Trans. Inform. Theory, Vol. IT-74, pp.493-494, July. 1971.

60. G. Walker,"On periodicity in series of related terms",Proc. Roy. Soc. London, series A, Vol. 131, pp.518-532,1931

61. P.D. Welch,"The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,"IEEE Trans. Audio-Electro-Acoust., Vol. AU-15, pp.70-73, June 1967.

62. G.U. Yule,"On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers", Philosophical Trans. Royal Soc. London, series A, Vol. 226, pp.267-298, July 1927.

63. G. Tacconi and D. Reidel,"Aspects of Signal Processing-Part 1",D. Reidel Publishing Co., Dordrecht-Holland/Boston , 1976.

64. T. Kailath,"Modern Signal Processing",Hemisphere Publishing Corp., New York , 1985.

65. S. Haykin,Ed.,"Nonlinear Methods of Spectral Analysis",2nd edition, New York , Springer-Verlag, 1983.

66. G.H. Golub and C.F. Van Loan,"Matrix Computations",Baltimore, MD: Jhons Hopkins University Press , 1983.

67. C.H. Chen,"Digital Waveform Processing and Recognition",CRC Press, Inc., Boca Raton, Florida, 1982.

68. R. Ray , A. Paulraj , and T. Kailath,"ESPRIT-A subspace rotation approach to estimation of parameters of cisoids in noise",IEEE Trans. ASSP, Vol. 34, No. 5, pp.1340-1342, Oct. 1986.

69. R. Kumaresan.Ph.D Dissertation , University of Rhode Island, Sept. 1982.

61. P.D. Welch,"The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,"IEEE Trans. Audio-Electro-Acoust., Vol. AU-15, pp.70-73, June 1967.

62. G.U. Yule,"On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers", Philosophical Trans. Royal Soc. London, series A, Vol. 226, pp.267-298, July 1927.

63. G. Tacconi and D. Reidel,"Aspects of Signal Processing-Part 1",D. Reidel Publishing Co., Dordrecht-Holland/Boston , 1976.

64. T. Kailath,"Modern Signal Processing",Hemisphere Publishing Corp., New York , 1985.

65. S. Haykin,Ed.,"Nonlinear Methods of Spectral Analysis",2nd edition, New York , Springer-Verlag, 1983.

66. G.H. Golub and C.F. Van Loan,"Matrix Computations",Baltimore, MD: Jhons Hopkins University Press , 1983.

67. C.H. Chen,"Digital Waveform Processing and Recognition",CRC Press, Inc., Boca Raton, Florida, |982.

68. R. Ray , A. Paulraj , and T. Kailath,"ESPRIT-A subspace rotation approach to estimation of parameters of cisoids in noise",IEEE Trans. ASSP, Vol. 34, No. 5, pp.1340-1342, Oct. 1986.

69. R. Kumaresan.Ph.D Dissertation , University of Rhode Island, Sept. 1982.