EDICS category SP 3.6.1

A Subspace Approach to Estimation of Autoregressive Parameters From Noisy Measurements [1]

Carlos E. Davila

Electrical Engineering Department, Southern Methodist University
Dallas, Texas 75275-0338

E-mail: cd@seas.smu.edu
Phone: (214) 768-3197
Fax: (214) 768-3573

Abstract

This paper describes a method for estimating the parameters of an autoregressive (AR) process from a finite number of noisy measurements. The method uses a modified set of Yule-Walker equations that lead to a quadratic eigenvalue problem which when solved, gives estimates of the AR parameters and the measurement noise variance.

Version 2, submitted to *IEEE Transactions on Signal Processing*
January 4, 1998

[1]This work was supported in part by a grant from the National Science Foundation (BCS-9308028) and a grant from the Whitaker Foundation.

I. Background

There are a number of applications where signals are modeled as autoregressive (AR) random processes. These include linear predictive coding (LPC) of speech [1], spectral estimation [2], biomedical signal processing [3], and time series forecasting [4]. When the signal to be modeled is observed in noise, the AR parameter estimates are biased and can produce misleading results. In this section, the problem of estimating AR parameters from noisy observations is formulated and some of the solutions that have been proposed are briefly described. In Sections II and III, a new method of estimating AR parameters is presented. This method is seen to be closely related to signal/noise subspace approaches used to estimate sinusoidal signal parameters. In Section IV the new method is compared with several well-known techniques and is seen to outperform them. A summary of the method is found in Section V. The following notation will be used throught: matrices are denoted by upper case letters (e.g., R_y, B), vectors are denoted by lower case letters (e.g., a, v); scalar elements of a sequence are denoted by lower case indexed letters (e.g., $r_y(n)$), all other scalars are denoted by lower case greek letters (e.g., λ).

The p^{th}-order AR Random Process is defined by

$$x(n) = -a(1)x(n-1) - a(2)x(n-2) - \cdots - a(p)x(n-p) + v(n) \tag{1}$$

where $v(n)$ is white noise having variance σ_v^2 and $a(k), k = 1, \ldots, p$ are the AR parameters. Using the fact that the autocorrelation function, $r_x(k)$ of the AR process also satisfies the autoregressive property

$$r_x(k) = -\sum_{i=1}^{p} a(i) r_x(k-i), \quad k \geq 1 \tag{2}$$

leads to the well-known *Yule-Walker* equations for the AR parameters

$$\begin{bmatrix} r_x(0) & \cdots & r_x(-(p-1)) \\ \vdots & \ddots & \vdots \\ r_x(p-1) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ \vdots \\ r_x(p) \end{bmatrix} \tag{3}$$

Let

$$y(n) = x(n) + z(n) \tag{4}$$

be a noisy measurement of the AR process where $z(n)$ is white noise having variance σ_z^2. The Yule-Walker equations become:

$$\begin{bmatrix} r_y(0) & \cdots & r_y(-(p-1)) \\ \vdots & \ddots & \vdots \\ r_y(p-1) & \cdots & r_y(0) \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_y(1) \\ \vdots \\ r_y(p) \end{bmatrix} \tag{5}$$

where $r_y(k)$ is the autocorrelation function of $y(n)$. It is assumed that σ_z^2 is unknown and is not easily measured. Hence, the AR parameter estimates derived from (5) will be biased since,

$$r_y(k) = r_x(k) + \delta(k)\sigma_z^2 \tag{6}$$

where $\delta(k)$ is the Dirac delta function. It has been shown that the biased AR parameters produce a "flatter" AR spectrum since the poles of the AR process are biased toward the center of the unit circle [5]. In practice, $r_y(k)$ must be estimated using the available measurements by incorporating one of several autocorrelation function estimators.

A number of approaches to AR parameter estimation for the case of noisy measurements have been developed. One approach is to treat the additive noise as a moving average component and estimate the ARMA parameters [2]. It has been observed that these ARMA parameter estimates may have an unacceptably high variance [6]. The noise compensation technique of Kay assumes σ_z^2 is known or can be estimated. This method then modifies the reflection coefficients for the effects of the noise [5]. Another approach requiring *a priori* knowledge of the additive noise variance is given in [7]. Measurement of the additive noise variance during no-signal intervals however can be problematic, additional processing is required to detect signal loss, and if the noise is nonstationary, incorrect variance estimates may result. Several authors have used the fact that for large values of k, the relationship in (2) is lacking $r_x(0)$, leading to the so-called Extended Yule-Walker equations [8, 9].

$$\begin{bmatrix} r_y(p) & \cdots & r_y(1) \\ \vdots & \ddots & \vdots \\ r_y(2p-1) & \cdots & r_y(p) \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_y(p+1) \\ \vdots \\ r_y(2p) \end{bmatrix} \tag{7}$$

A problem with this method is that less data is available to compute higher autocorrelation lag

estimates. To compensate for errors in the estimated autocorrelation lags, the Overdetermined Extended Yule-Walker equations have been proposed [10, 11],

$$\begin{bmatrix} r_y(p) & \cdots & r_y(1) \\ \vdots & \ddots & \vdots \\ r_y(p+q-1) & \cdots & r_y(q) \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_y(p+1) \\ \vdots \\ r_y(p+q) \end{bmatrix} \quad q > p \tag{8}$$

This method also avoids $r_y(0)$ and compensates for errors in $r_y(k)$ by using more equations than unknowns. The overdetermined system is then solved via a constrained least squares algorithm. The Overdetermined Extended Yule-Walker equations also suffer from the need to compute a large number of high-order autocorrelation lags. Higher-order cummulants have also been used [12, 13]. The cummulant method assumes that the additive noise $z(n)$ has a non-skewed density function, and that the driving noise $v(n)$ is non-Gaussian. It then uses odd-order cummulants, to generate a set of Yule-Walker-like equations. The motivation underlying this method is that since the odd-order cummulants of the measurement noise are zero, the cummulant Yule-Walker equations are independent of the measurement noise. The non-skewed density assumption on $z(n)$ however may not always hold. Moreover, some time series can be very close to Gaussian, for example the electroencephalogram (EEG) has been shown to close to Gaussian over time intervals of two seconds or less [14]. Several other methods are described in [2, 15].

II. Noise-Compensated Yule-Walker Equations

The Noise-Compensated Yule-Walker (NCYW) equations are defined as

$$\left[R_y - \lambda I_p \right] a = \tilde{g}_1 \tag{9}$$

$$\begin{aligned} g_1 a &= r_y(p+1) \\ &\vdots \\ g_q a &= r_y(p+q) \end{aligned} \tag{10}$$

where $a = \left[\begin{array}{ccc} -a(1) & \cdots & -a(p) \end{array} \right]^T$, I_p is the $p \times p$ identity matrix,

$$g_i = \left[\begin{array}{ccc} r_y(p+i-1) & \cdots & r_y(i) \end{array} \right], \quad i = 1, \ldots, q \tag{11}$$

$\tilde{g}_1$ is an element-reversed version of g_1^T,

$$R_y = \left[\begin{array}{ccc} r_y(0) & \cdots & r_y(-(p-1)) \\ \vdots & \ddots & \vdots \\ r_y(p-1) & \cdots & r_y(0) \end{array} \right] \tag{12}$$

and λ is an estimate of the noise power, σ_z^2. The unknowns in (9)-(10) are a and λ. Since the NCYW equations are nonlinear in the AR parameters and λ, it is useful to know the minimum value of q, the number of auxiliary linear equations, which will guarantee a unique solution for the $p+1$ unknowns. Clearly, $q = p$ is sufficient to guarantee a unique solution since this gives p linear equations in a. In fact, the Extended Yule-Walker equations use these same p equations, ignoring the first p nonlinear equations. However these nonlinear equations are based on relatively low autocorrelation lag estimates which tend to be more accurate than the autocorrelation lag estimates used in the Extended Yule-Walker equations. Hence if a method can be found to solve the combined $p+q$ nonlinear and linear equations, one might anticipate improved performance relative to the Extended Yule-Walker equations. The following section describes such a method.

III. Subspace Method

The Noise-Compensated Yule-Walker equations can be written as

$$\left(\overline{R}_y - \lambda B \right) v = 0_{p+q} \tag{13}$$

where

$$\overline{R}_y = \begin{bmatrix} r_y(1) & r_y(0) & r_y(-1) & \cdots & r_y(-(p-1)) \\ r_y(2) & r_y(1) & r_y(0) & \cdots & r_y(-(p-2)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_y(p) & r_y(p-1) & r_y(p-2) & \cdots & r_y(0) \\ r_y(p+1) & r_y(p) & r_y(p-1) & \cdots & r_y(1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_y(p+q) & r_y(p+q-1) & r_y(p+q-2) & \cdots & r_y(q) \end{bmatrix} \tag{14}$$

$$B = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad v = \begin{bmatrix} 1 \\ a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} \tag{15}$$

0_{p+q} is a column vector having $p+q$ zeros, and $q \geq p$; so that a unique solution is assured. The dimensions of $\overline{R}_y, B$, and v are $p+q \times p+1, p+q \times p+1$, and $p+1 \times 1$, respectively. Multiplying both sides of (13) by $\left(\overline{R}_y - \lambda B\right)^T$ leads to the quadratic eigenvalue problem:

$$\left(A_0 + \lambda A_1 + \lambda^2 A_2\right) v = 0_{p+1} \tag{16}$$

where

$$A_0 = \overline{R}_y^T \overline{R}_y, \;\; A_1 = -\left(R_y^T B_1 + B_1^T R_y\right), \;\; A_2 = B_1^T B_1 \tag{17}$$

Each of the matrices in (17) has dimension $p+1 \times p+1$. There are several ways to solve the quadratic eigenvalue problem [16]. One method is to define

$$P = \begin{bmatrix} A_0 & 0 \\ 0 & I \end{bmatrix}, \;\; Q = \begin{bmatrix} -A_1 & -A_2 \\ I & 0 \end{bmatrix} \tag{18}$$

It is readily verified that solving (16) is equivalent to solving the $2p+2$-dimensional linear eigenvalue problem

$$(P - \lambda Q)\,\overline{v} = 0 \tag{19}$$

Let $\overline{v}_k, \quad k = 1, \ldots, 2p+2$ be eigenvectors solving (19), these along with their corresponding eigenvalues appear in complex conjugate pairs [16]. Only one of the $\overline{v}_k$ will correspond to the correct solution to (9)-(10). Recalling that the eigenvalues solving the quadratic eigenvalue problem (16) are estimates of the additive noise power σ_z^2, it is clear that for the noiseless case, only two complex conjugate eigenvalues will be zero since there is only one real solution to equations (9)-(10), all the other eigenvalues will be complex. It follows that for the noisy case, all of the eigenvalues associated with the noiseless case will be increased by an amount equal to the noise power. Hence, for the noisy case, the correct solution is the eigenvector corresponding to the only real eigenvalue solving the quadratic eigenvalue problem (16). In practice, errors in estimating the autocorrelation matrix due to finite data lengths tend to introduce small imaginary components in the eigenvalues solving (16). Hence, a more suitable criterion is to select the eigenvalue with the smallest imaginary part. Unfortunately, for high noise levels, this can lead to the selection of an eigenvalue with a small imaginary part but with a large real part which does not correspond to σ_z^2. A selection criterion which appears to be very robust is to select the eigenvalue having the smallest modulus. This criterion has the effect of eliminating eigenvalues with large real components which may have small imagingary components. The selection criterion could be expanded, for example, one could look at the error associated with the last q linear equations in (10). However, choosing the eigenvalue having minimum modulus seems to work well, and no other criteria were used in the simulations below. Note the similarity between the proposed method and noise subspace methods used to estimate the frequencies of sinusoids observed in white noise [2]. The method described here is analogous to the Pisarenko harmonic retrieval method, wherein a basis for the one-dimensional noise subspace is found as the eigenvector associated with the minimum eigenvalue of the data autocorrelation matrix [2, 17]. The dimension of the noise subspace for the proposed method can be increased beyond one by increasing the dimension of $\overline{R}_y$. For higher-order noise subspaces, the AR model order and parameters can be estimated from common zeros of the noise subspace eigenvectors. Higher-order noise subspaces will not be considered here. The proposed Subspace

Method can be summarized as follows:

1. Form estimates of the autocorrelation matrix $\overline{R}_y$ defined in (14).

2. Form the matrices A_0, A_1, and A_2, defined in (17).

3. Form the matrices P and Q, defined in (18).

4. Solve the generalized eigenvalue problem

$$P\overline{v} = \lambda Q\overline{v} \tag{20}$$

5. The AR parameter vectors are the elements of the generalized eigenvector $\overline{v}_{min}$ associated with the generalized eigenvalue having minimum modulus. The AR parameters are given by

$$\hat{a}(i) = \frac{[\overline{v}_{min}]_i}{[\overline{v}_{min}]_1}, \quad i = 2, \ldots, p+1 \tag{21}$$

   where $[\overline{v}_{min}]_i$ denotes the i^{th} entry of $\overline{v}_{min}$.

6. The estimate of the noise power σ_z^2, is the modulus of the generalized eigenvalue in (20) having minimum modulus.

IV. Experimental Results

Experiment 1: A 10,000-point sequence was synthesized by applying uniformly distributed white noise to the all-pole system having poles at $0.75e^{\pm j0.2\pi}$, $0.8e^{\pm j0.4\pi}$, and $0.85e^{\pm j0.7\pi}$ giving $p = 6$. In order to allow initial transients to die out, the first 9,500 samples of the sequence were discarded and the last 500 samples were selected as an exemplar of a stationary AR process. After normalizing this segment to have zero mean and unit variance, additive, zero mean white Gaussian noise having a variance of 0.316 was added so as to produce an SNR of approximately 5 dB. The $N = 500$-sample noisy AR data, $y(n)$, was then submitted for AR parameter estimation via the following methods:

- Subspace Method (SS)

- Yule-Walker Equations (YW)

- Overdetermined Extended Yule-Walker Equations (OD)

For all methods tested, the autocorrelation function was estimated using the autocorrelation method [2]

$$\hat{r}_y(k) = \frac{1}{N} \sum_{n=1}^{N-k} y(n)y(n+k) \tag{22}$$

For the SS method, $q = p = 6$ linear auxiliary equations equations (10) in addition to the 6 nonlinear equations (9). For the OD method, a total of 113 equations was found to give the best results. The equations in the OD method were exponentially downweighted with a weighting parameter of 0.98 as described in [2]. In all, 20 independent trials were run. Fig. 1 shows the location of the estimated model poles with respect to the unit circle and the corresponding AR spectra for each of these experiments. It can be seen that the proposed SS method provides a more compact clustering of the estimated poles about the true poles than the other methods tested. The YW method produces poles which are severely biased toward the origin as expected.

Experiment 2: The experiment was then repeated with the actual poles positioned closer to the unit circle at $0.85e^{\pm j0.2\pi}, 0.9e^{\pm j0.4\pi}$, and $0.95e^{\pm j0.7\pi}$ Again, the SS method out-performs the other methods as seen in Fig. 2.

Experiment 3: Next 10 runs using the AR model poles in *Experiment 1* were performed. The measurement noise variance σ_z^2 was varied over the 10 runs from 0.05 to 0.5 in steps of 0.05. For each run, 50 independent trials were performed from which the mean and standard deviation of the estimated measurement noise variance, $\hat{\sigma}_z^2$ was computed. One trial consisted of AR parameter estimation with the SS method from N=500 data points as described in *Experiment 1.* The same experiment was then repeated with the AR model poles used in *Experiment 2.* These results are shown in Fig 3. The estimated noise variance is seen to have a mean that is very close to the actual measurement noise variance.

V. Summary

A method for estimating the parameters of an autoregressive process from a finite number of noisy measurements was described. Estimates of the AR parameters and measurement noise variance result by solving a quadratic eigenvalue problem. The new method was demonstrated to outperform several popular methods.

References

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Prentice Hall, 1978.

[2] S. Kay, *Modern Spectral Estimation*. Prentice Hall, 1988.

[3] A. Cohen, *Biomedical Signal Processing, Vol. 1*. CRC Press, 1986.

[4] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[5] S. Kay, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. ASSP-28, pp. 292–303, June 1980.

[6] S. Kay, "The effects of noise on the autoregressive spectral estimator," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. ASSP-27, pp. 478–485, Oct. 1979.

[7] W. R. Wu and P. C. Chen, "Adaptive AR modeling in white gaussian noise," *IEEE Trans. on Signal Proc.*, vol. 45, pp. 1184–1192, May 1997.

[8] Y. T. Chan and R. Langford, "Spectral estimation via the high-order Yule-Walker equations," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. ASSP-30, pp. 689–698, 1980.

[9] D. Gingras, "Estimation of the autoregressive parameter from observations of a noise-corrupted autoregressive time series," in *Proc. ICASSP-82*, (Paris), 1982.

[10] H. Ogura and H. Yoshida, "Spectral analysis and subtraction of noise in radar signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 62–71, 1981.

[11] J. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, Sept. 1982.

[12] D. Brillinger, "An introduction to polyspectra," *Ann. Math. Statist.*, vol. 36, pp. 1351–1374, 1965.

[13] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust. , Speech, Signal Processing*, vol. ASSP-37, pp. 360–377, March 1989.

[14] J. McEwen and G. Anderson, "Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity," *IEEE Trans. Biomed. Eng.*, vol. BME-22, pp. 361–369, Sept. 1975.

[15] D. P. McGinn and D. H. Johnson, "Estimation of all-pole model parameters from noise-corrupted sequences," *IEEE Trans. Acoust. , Speech, Signal Processing*, vol. ASSP-37, pp. 433–435, March 1989.

[16] F. Chatelin, *Eigenvalues of Matrices*. Wiley, 1993.

[17] V. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, pp. 347–366, 1973.

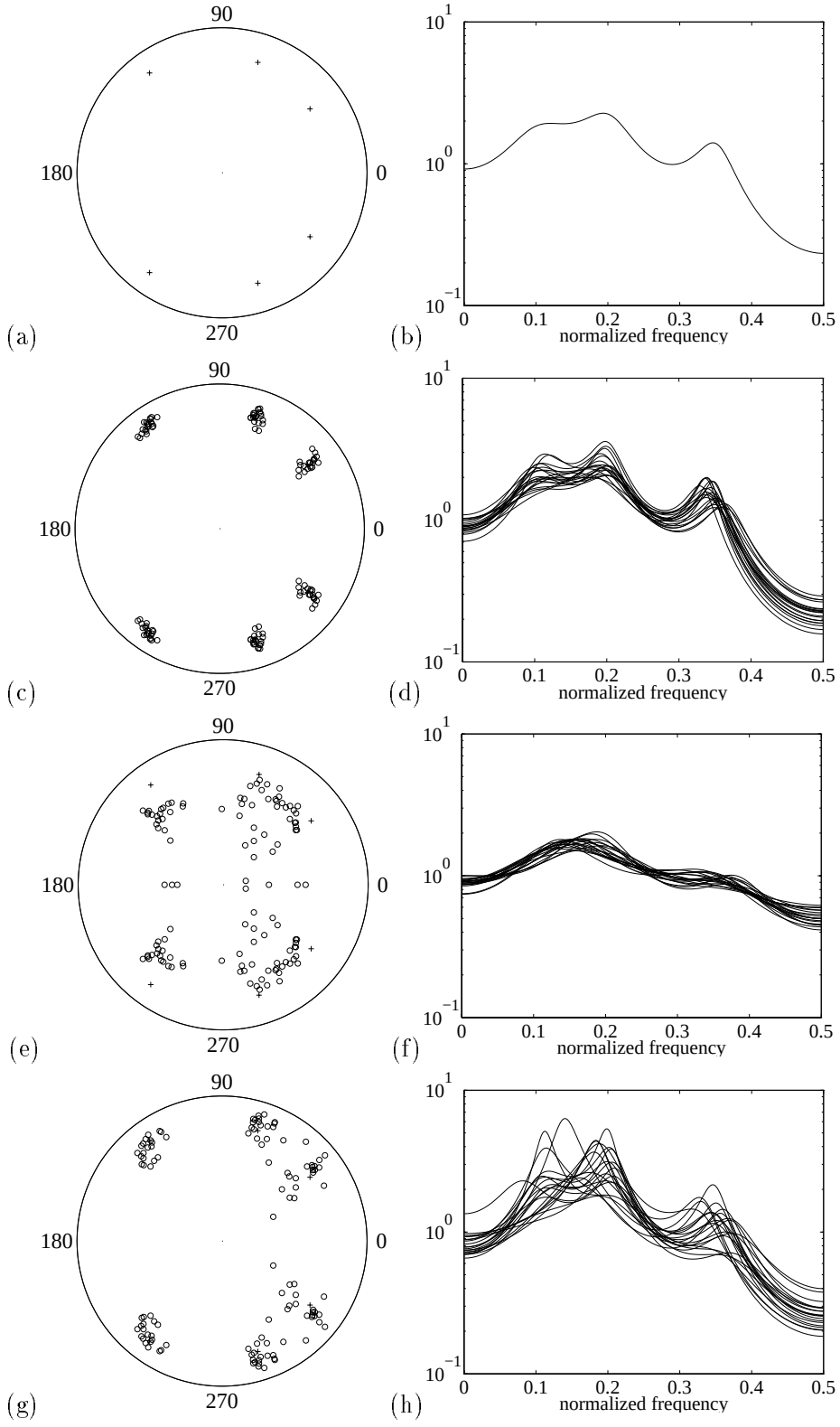


Figure 1: *Experiment 1* results: (a) actual poles, (b) actual AR spectrum, (c) SS poles, (d) SS spectra, (e) YW poles, (f) YW spectra, (g) OD poles, (h) OD spectra.

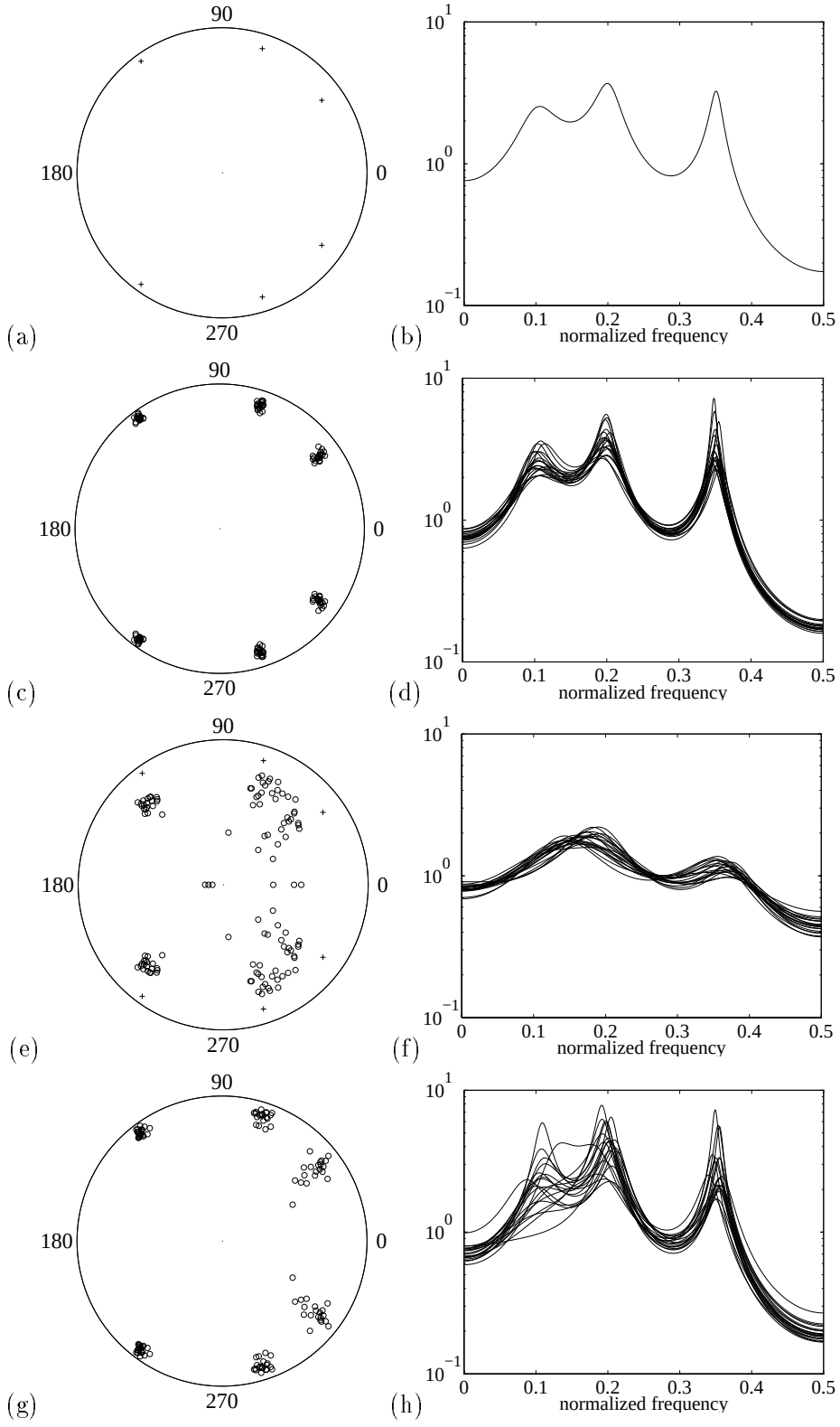
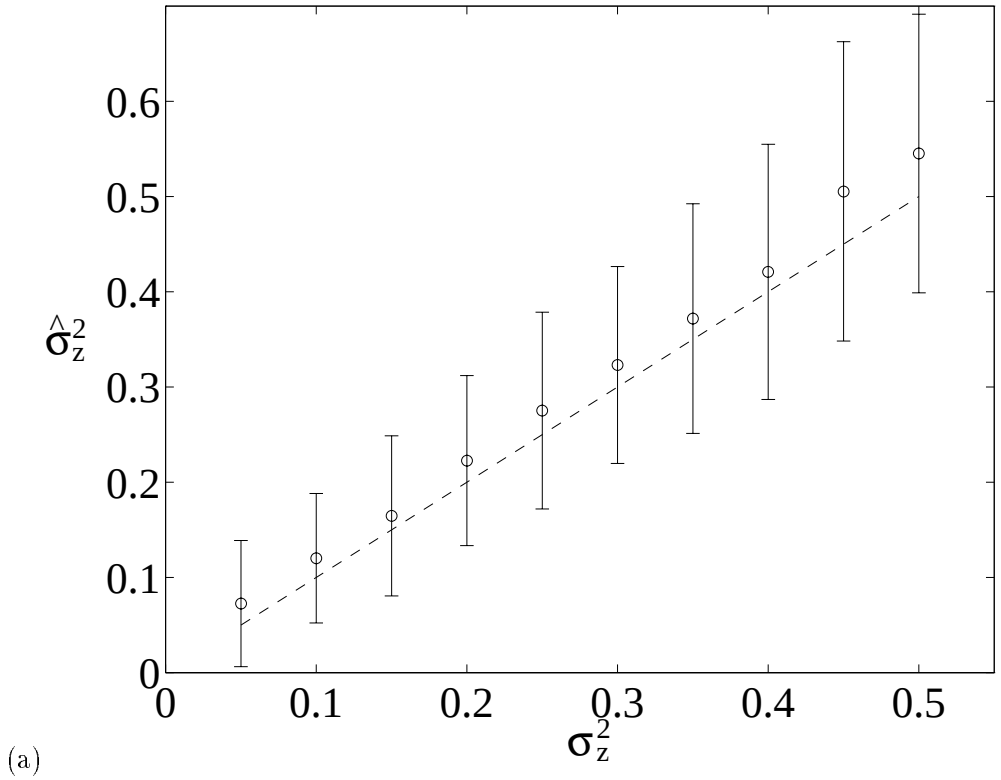
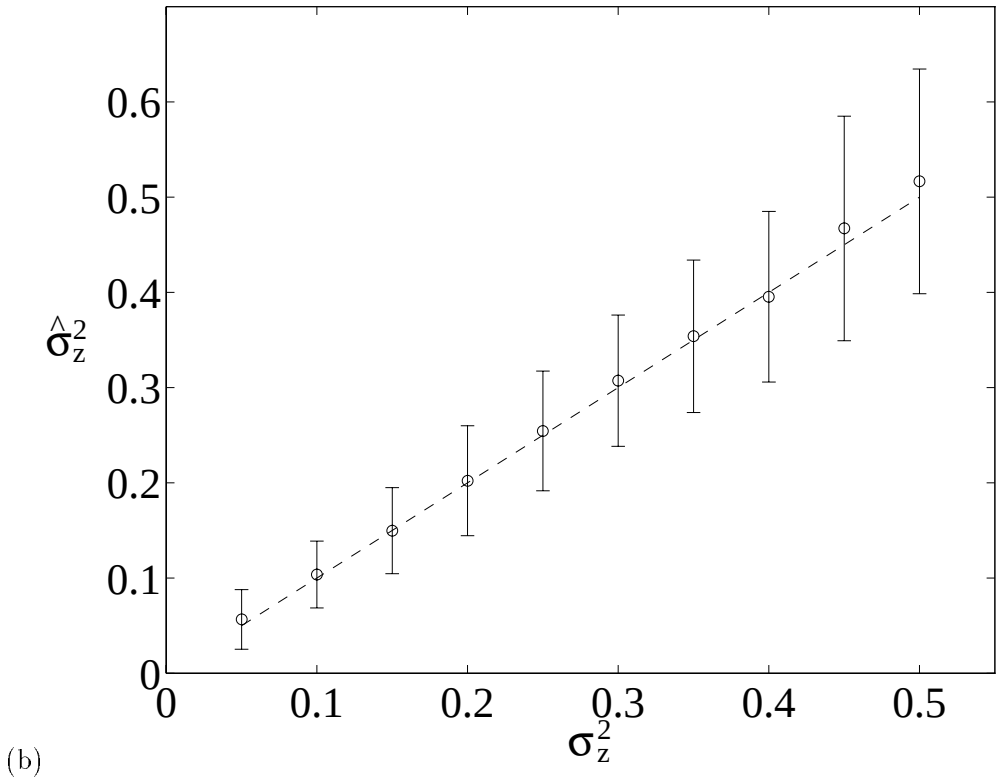


Figure 2: *Experiment 2* results: (a) actual poles, (b) actual AR spectrum, (c) SS poles, (d) SS spectra, (e) YW poles, (f) YW spectra, (g) OD poles, (h) OD spectra.



(a)



(b)

Figure 3: Mean and standard deviation of measurement noise variance estimates for (a) *Experiment 1* poles and (b) *Experiment 2* poles. The dashed line corresponds to $\hat{\sigma}_z^2 = \sigma_z^2$.