ACCURACY PROPERTIES OF HIGH ORDER YULE WALKER EQUATION ESTIMATORS OF SINUSOIDAL FREQUENCIES

Randolph L. Moses*, Jian Li*, and Petre Stoica**

*Department of Electrical Engineering, The Ohio State University, 2015 Neil Avenue, Columbus, Ohio 43210 USA

**Department of Automatic Control, Polytechnic Institute of Bucharest, Splaiul Independentei 313, Bucharest, R77206 ROMANIA

ABSTRACT. This paper considers the problem of estimating the frequencies of multiple sinusoids using the High Order Yule-Walker equations. Asymptotic expressions for the variances of frequency estimates obtained from two HOYW-based algorithms are presented. Two-step estimators which minimize these variances are also developed. The theoretical variances and Monte-Carlo simulation results are compared.

Keywords. Identification; parameter estimation, accuracy analysis, computational methods

I. Introduction

The problem considered in this paper is that of estimating the frequencies of sinusoids from a finite set of noisy measurements. A popular method involves estimating autocorrelation coefficients from the data, and solving a set of Yule-Walker equations [1,2]. If the data consists of m sinusoids in additive white noise, the order of the Yule-Walker equations must be greater than or equal to $2m$. It has been empirically found [1,2,3,4] and theoretically confirmed [5,6,7] that if the order is chosen to be larger than $2m$, (leading to the so-called High Order Yule-Walker estimates) the estimation accuracy improves.

In this paper we present asymptotic variance expressions for the frequencies of sinusoids estimated from High Order Yule-Walker equations [6,7]. We consider two forms of the HOYW estimator, a minimum norm solution (using singular value decomposition techniques) and a computationally more efficient sparse solution [5]. We present a two-step algorithm that gives asymptotically minimum variance frequency estimates. We compare the theoretical variances for several specific algorithms, and with results obtained from Monte-Carlo simulations. We also discuss the tradeoffs among these methods in terms of both accuracy and computations.

II. The High Order Yule Walker Equation Estimates

Assume we are given N measurements of a noisy sinusoidal signal:

$$y(t) = \sum_{k=1}^{m} \gamma_k \sin(\omega_k t + \phi_k) + e(t) = x(t) + e(t), \qquad (1)$$

where $e(t)$ is white noise with zero mean and variance λ^2. The goal is to estimate the frequencies $\{\omega_k\}_{k=1}^m$ of the sinusoids. One popular procedure for estimating these frequencies is to use the so-called High Order Yule-Walker equations. To develop this approach, let

$$A(z) = a_0 + \cdots + a_n z^n, \quad n \triangleq 2m$$

be a polynomial with roots at $e^{\pm j\omega_k}$, $k = 1, 2, \ldots, m$. Then since $A(q^{-1})x(t) = 0$ (where q^{-1} is the unit delay operator: $q^{-1}x(t) = x(t-1)$), it follows that

$$\begin{aligned} A(q^{-1})y(t) &= A(q^{-1})e(t) \\ &\Rightarrow A(q^{-1})r_k = 0, \quad k > n \end{aligned}$$

where $r_k = E[y(t)y(t-k)]$ is the autocorrelation sequence of $y(t)$. Moreover, if we define

$$C(z) = A(z)B(z) = c_0 + c_1 z + \cdots + c_L z^L \qquad (2)$$

for any polynomial $B(z) = b_0 + b_1 + \cdots + b_{L-n} z^{L-n}$ we have

$$C(q^{-1})r_k = 0, \quad k > L \qquad (3)$$

Equation (3) can be used to solve for the coefficients of $C(z)$ from autocorrelations via the High Order Yule Walker (HOYW) equations:

$$Q \begin{bmatrix} r_L & \cdots & r_1 \\ \vdots & & \vdots \\ r_{L+M-1} & \cdots & r_M \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_L \end{bmatrix} = -Q \begin{bmatrix} r_{L+1} \\ \vdots \\ r_{L+M} \end{bmatrix}, \qquad (4)$$

or

The work of J. Li was supported in part by a fellowship from the People's Republic of China

$$\begin{bmatrix} \gamma_1^2 \sin(L+1)\omega_1 & \cdots & \gamma_m^2 \sin(L+1)\omega_m \\ \vdots & & \vdots \\ \gamma_1^2 \sin(L+M)\omega_1 & \cdots & \gamma_m^2 \sin(L+M)\omega_m \end{bmatrix}$$

$$S = E\left\{v(t)v^T(t)\right\} \quad (12)$$

$$v(t) = C(q^{-1})[\ e(t-1) \ \cdots \ e(t-M)\ ]^T$$

Then, for finite L and M,

$$\sqrt{N}(\hat{\omega} - \omega) \to \mathcal{N}(0, P_1) \text{ as } N \to \infty$$

where

$$P_1 = \lambda^2 F(H^T W H)^{-1} H^T W S W H (H^T W H)^{-1} F^T \quad (13)$$

$$W = Q^T Q$$

Theorem 2: *Let θ_2 be the solution to (10) and $\hat{\theta}_2$ be the solution to (11). Let $\hat{C}(z)$ denote the polynomial corresponding to $\hat{\theta}_2$, and let $\{\hat{\omega}_k\}_{k=1}^m$ denote the positive angular positions of the m complex-conjugate pairs of roots of $\hat{C}(z)$ which are closest to the unit circle. Assume that no zeroes of $B(z)$ lie on the unit circle. Then, for finite L and M,*

$$\sqrt{N}(\hat{\omega} - \omega) \to \mathcal{N}(0, P_2) \text{ as } N \to \infty$$

where

$$P_2 = \lambda^2 F(H^T W H)^{-1} H^T W S W H (H^T W H)^{-1} F^T \quad (14)$$

and where F, H, S, and W are defined as in Theorem 1, but with θ_2 replacing θ_1 there.

These theorems give the asymptotic variance of the HOYW frequency estimates for any weighting matrix Q. The proofs of these theorems can be found in [6,7,8,10]. Note that the only difference in the variance expressions for the estimators is the particular $C(z)$ polynomial coefficients used.

Optimal Choice of Q

We next discuss the choice of the weighting matrix Q in the estimation. A common choice is $Q = I$. However, other choices of Q can lower the variance of the frequency estimates. It is readily shown that the P matrices in (13) and (14) satisfy:

$$P \geq \lambda^2 F(H^T S^{-1} H)^{-1} F^T \quad (15)$$

where equality holds for

$$W = Q^T Q = \lambda^2 S^{-1} \quad (16)$$

Thus, if $Q = \lambda S^{-1/2}$ is used, the resulting frequency estimates have (asymptotically) minimum variance. Note, however, that S is a function of the (unknown) c_i coefficients; S is given by $S = \lambda^2 \phi^T \phi$ where (c.f. (12)):

$$\phi^T = \begin{bmatrix} 1 & c_1 & \cdots & c_L & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & c_1 & \cdots & c_L \end{bmatrix} \quad M \times (L+M) \quad (17)$$

Thus, the optimal choice of Q can be obtained with the following two-step procedure [8]:

Step 1: Set $Q = I$, and calculate initial estimates of the c_i parameters using the HOYW equations.

Step 2: Use the c_i estimates from Step 1 to form $\hat{\phi}$, and repeat the estimation procedure with $Q = (\hat{\phi}^T \hat{\phi})^{-1/2}$. (Note that $\hat{\phi}^T \hat{\phi}$ is a symmetric Toeplitz matrix, so Q can be computed efficiently using a modification of the Levinson-Durbin algorithm.) Finally, obtain the frequencies estimates $\hat{\omega}_i$ from the new $\hat{C}(z)$ polynomial.

It can be shown [8] that the use of estimated c_i coefficients instead of exact ones to form Q in the second step does not affect the asymptotic variance expressions.

IV. Numerical Examples

In this section we present some numerical examples which illustrate the performance of the above algorithms. We compare the theoretical asymptotic frequency variances with each other and with estimated variances obtained via Monte-Carlo simulation.

The example consists of two closely-spaced sinusoids in white noise. The frequencies and amplitudes of the sinusoids are given by: $\omega_1 = 1.00$, $\omega_2 = 1.05$, $\gamma_1 = \gamma_2 = 1$. The noise variance (λ^2) is chosen to give a signal-to-noise ratio of 10dB (where SNR $= 10\log(\gamma_1^2/2\lambda^2)$). In all examples $N = 150$ data points were used; we have also performed simulations with $N = 1500$, and there the estimated variances match the theoretical variances much more closely than for $N = 150$.

The tables and figures below give normalized variances $N \cdot \text{var}(\hat{\omega}_1)$ (i.e. the (1,1) element of P_1 or P_2). Simulation results were obtained by averaging 50 Monte-Carlo experiments. Only results for $\hat{\omega}_1$ are shown, but the variances for $\hat{\omega}_2$ are similar. All variance values are given in dB. For this example, $N \cdot$CRLB$= -42$dB, where CRLB denotes the asymptotic Cramèr-Rao Lower Bound for the variance of $\hat{\omega}_1$.

Table 1 shows the theoretical variances of $\hat{\omega}_1$ for several values of L and M, and using the direct minimum norm estimate ($\hat{\theta}_1$ with $Q = I$). Note that the variance decreases by nearly 50dB as L and M increase. Table 2 shows the variances for the two-step optimum minimum norm estimate; here, similar decrease in variance is seen for increasing L and M. Note that the optimal variances are significantly smaller than the direct variances for large M and small L; in other ranges the direct and optimum variances are almost equal.

Figure 1 shows theoretical and simulation variances of $\hat{\omega}_1$ for the minimum norm ($\hat{\theta}_1$) estimator with $M = 20$. It can be seen that the optimal variance is significantly smaller than the direct variance only for small L. The simulated variances follow theoretical results only approximately; however, for $N = 1500$ data points, the agreement between simulation and theoretical variances is nearly exact.

Figure 2 shows similar variance curves for the sparse matrix ($\hat{\theta}_2$) estimate with $M = 20$. Note that while the theoretical optimal variances for $\hat{\theta}_1$ and $\hat{\theta}_2$ are similar, the theoretical direct variance is much larger for $\hat{\theta}_2$. Moreover, the direct variance is sensitive to the particular choice of L. Both the direct and optimal estimated variances in Figure 2 exhibit a large sensitivity to L. Comparing Figures 1 and 2, we see that the minimum norm (SVD) approach generally produces lower variances for this example.

The higher variances from $\hat{\theta}_2$ and the large swings in the performance as a function of L are primarily caused

Table 1. Theoretical Variances of $\hat{\omega}_1$ for the Direct, Minimum Norm Estimator.

$L\backslash M$	5	9	13	17	21
5	23.3	5.99	1.37	-1.93	-7.78
9	7.20	.34	-6.02	-12.1	-16.0
13	2.01	-5.08	-11.6	-16.7	-19.6
17	1.63	-9.14	-13.9	-17.3	-21.9
21	-4.48	-13.7	-17.5	-21.2	-25.2

Table 2. Theoretical Variances of $\hat{\omega}_1$ for the Two-Step Optimal, Minimum Norm Estimator.

$L\backslash M$	5	9	13	17	21
5	23.3	5.42	-1.55	-7.20	-13.6
9	7.19	.32	-6.11	-12.6	-17.5
13	2.01	-5.10	-11.8	-16.7	-19.9
17	1.63	-9.15	-13.9	-17.3	-22.0
21	-4.48	-13.8	-17.5	-21.3	-25.3

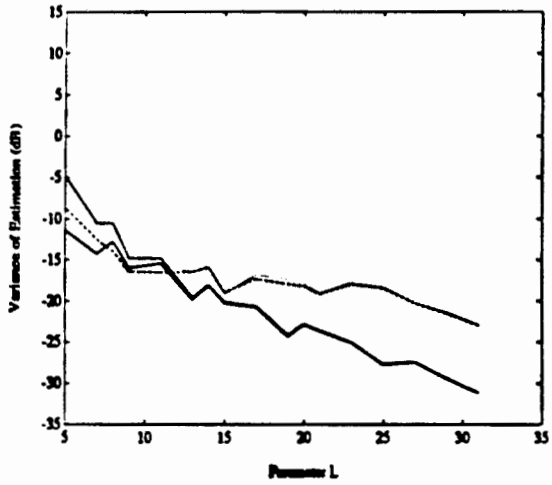

Figure 1. Variances of $\hat{\omega}_1$ using the Minimum-Norm Estimator. Solid lines are theoretical variances (direct and optimal). The dotted line is the estimated direct variance, and the dashed line is the estimated two-step optimal variance.

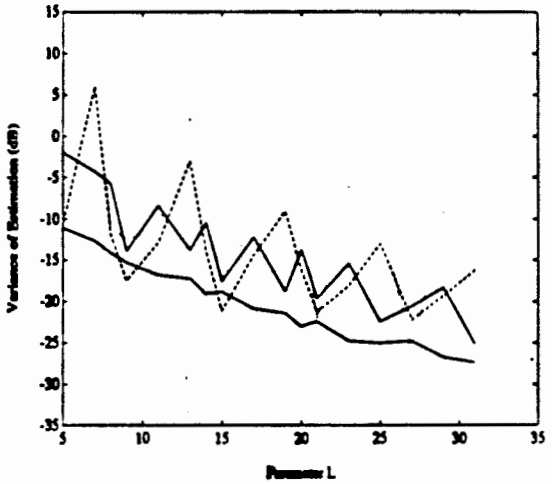

Figure 2. Variances of $\hat{\omega}_1$ using Sparse Matrix Estimator with $J = I$. Lines are as in Figure 1.

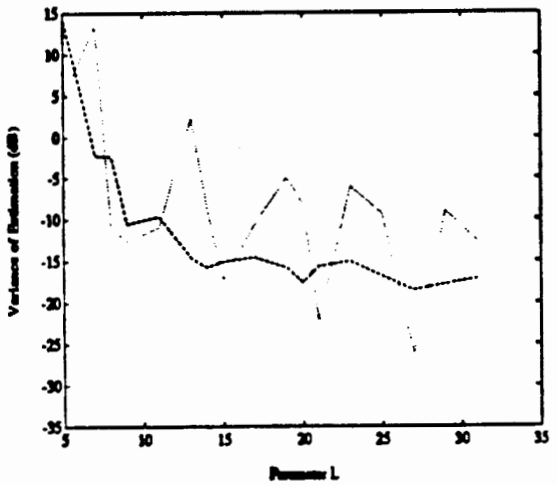

Figure 3. Variances of $\hat{\omega}_1$ for two Sparse Matrix Estimators, Direct Solution. Dotted line is the $\hat{\theta}_2$ estimated variance. Dashed line is the $\hat{\theta}_3$ estimated variance.

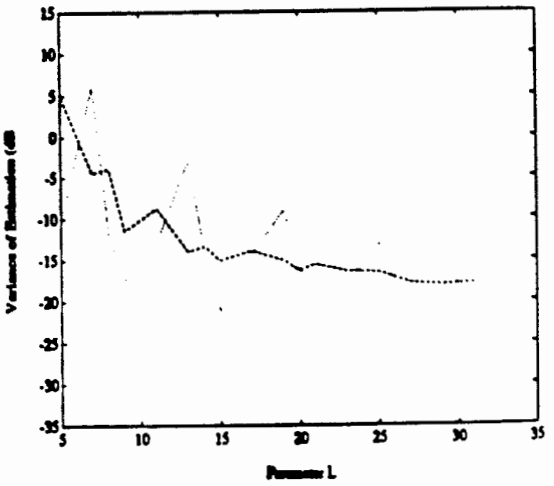

Figure 4. Variances of $\hat{\omega}_1$ for two Sparse Matrix Estimators, Two-Step Procedure.