Study of the statistical performance of the Pisarenko harmonic decomposition method

Petre Stoica
Arye Nehorai

Indexing terms: Signal processing, Noise and interference

Abstract: A self-contained statistical analysis of the Pisarenko method for estimating sinusoidal frequencies from signal measurements corrupted by white noise is presented. An explicit formula is provided for the asymptotic covariance matrix of the joint estimation errors of both the minimum eigenvector and the minimum eigenfilter coefficients of the data covariance matrix. Our theoretical results extend and reinforce previous results obtained by Sakai [1]. A numerical study of the performance of the Pisarenko method, and a comparison with the performance achieved by the Yule–Walker method are also reported.

1 Introduction

Determination of angular frequencies of sinusoidal signals from noisy measurements is a topic of considerable recent interest [2]. When the measurement noise is white, consistent estimates of the sinusoidal frequencies can be obtained by the so-called Pisarenko harmonic decomposition method [1, 3]. This method consists of determining the minimum eigenvalue of the data covariance matrix and its associated eigenfilter. Then the sinusoidal frequencies are determined as the angular positions of the eigenfilter zeros.

Recently, Sakai [1] presented a formula for the asymptotic covariance matrix of the estimation errors of the minimum eigenfilter coefficients. The analysis of Reference 1 relies on some referenced results, which makes it somewhat difficult to follow. In this paper, we present a self-contained analysis of the statistical performance of the Pisarenko method. Our analysis is more general than that of Reference 1 since we consider the joint estimation errors of both the minimum eigenvector and the minimum eigenvalue. The extended result that we obtain exhibits an interesting feature. While the covariance matrix of the estimated eigenvector depends only on the second order moments of the data, the joint covariance matrix of the estimated eigenvector and eigenvalue also depends on the fourth order moment of the measurement noise.

2 Preliminaries

Let \( x(t) \) denote a signal consisting of \( m \) sinusoids

\[
    x(t) = \sum_{k=1}^{m} a_k \sin(\omega_k t + \varphi_k), \quad \omega_k \in (0, \pi)
\]

where \( t \) is integer valued \( t = 1, 2, \ldots \) (for simplicity of notation we therefore assume a normalised time scale for which the sampling interval equals unit). Let \( y(t) \) denote the noisy measurement of \( x(t) \)

\[
    y(t) = x(t) + \eta(t)
\]

The noise \( \eta(t) \) is assumed to be white and of zero mean; its variance is denoted by \( \lambda \). It is also assumed that \( x(t) \) and \( \eta(s) \) are uncorrelated for all \( t \) and \( s \).

It is well known that \( x(t) \) obeys the following homogeneous autoregressive (AR) equation of order \( n \)

\[
    A(q^{-1})x(t) = 0
\]

where \( q^{-1} \) denotes the unit delay operator, and

\[
    A(q^{-1}) = \prod_{k=1}^{m} (1 - 2 \cos \omega_k q^{-1} + q^{-2})
\]

\[
    = 1 + a_1 q^{-1} + \cdots + a_m q^{-m}
\]

To see this, observe that the 'annihilating filter' \( A(q^{-1}) \) has zeros in its frequency characteristic at the sinusoidal frequencies: \( A(e^{i\omega}) = \prod_{k=1}^{m} [1 - e^{i(\omega - \omega_k)}][1 - e^{i(\omega + \omega_k)}] = 0 \) for \( \omega = \omega_k \). Inserting eqn. 1 into eqn. 2 one obtains the following special autoregressive moving-average (ARMA) equation for \( y(t) \)

\[
    A(q^{-1})y(t) = A(q^{-1})x(t)
\]

The poles and the zeros of eqn. 3 lie on the unit circle at \( e^{i\omega_k} \) for \( k = 1, \ldots, m \). Thus, the problem of estimating the sinusoidal frequencies \( \{\omega_k\} \) from the noisy measurements...
(\{\gamma(t)\}) can be, and often is, reduced to that of estimating the AR parameters in eqn. 3. Once the \{a_i\} coefficients are determined, the frequencies \{\omega_k\} may be obtained either (a) by looking at the peaks of \(|1/|A(e^{i\omega})|^2|\); or (b) as the angular positions of the zeros of \(A(z)\). The estimation errors of \{\omega_k\} will clearly depend on the estimation errors of \{a_i\}. Asymptotically valid formulas relating the covariance matrix of \{\omega_k\} to that of \{\hat{a}_i\} when \{\hat{a}_i\} are determined from \{\hat{a}_i\} by procedures (a) and (b) above, are presented in Reference 1 (for (a)) and Reference 20 (for (b)). The availability of these formulas makes it possible to restrict attention to the problem of analysing the estimation errors of \{a_i\} and that is what we will do in this paper.

Pisarenko method for consistently estimating \(a_i\) is based on the following simple observations. If we premultiply eqn. 3 by \(|y(t), \ldots, y(t-n)|^T\) and take expectation then we get

\[
\begin{bmatrix}
\Omega - \lambda I
\end{bmatrix}
\begin{bmatrix}
1
\end{bmatrix}
= 0
\tag{4}
\]

where

\[\theta = [a_1, \ldots, a_n]^T\]

is the parameter vector,

and \(\Omega\) is the \((n+1)\)th order theoretical covariance matrix of the data

\[\Omega_{ij} = E[y(t-i)\gamma(t-j)]
= Ex(t-\delta)x(t-j) + \lambda \delta_{i,j} \quad 1 \leq i, j \leq n + 1
\tag{5}
\]

(here \(\delta_{i,j}\) is the Dirac delta; \(\delta_{i,j} = 1\) for \(i = j\) and 0 otherwise). Since \(\lambda_{max}(\Omega) \geq \lambda\) [see, for example, eqn. 5] and \(\Omega - \lambda I\) must be singular [see eqn. 4], it readily follows that \(\lambda\) (the noise power) is equal to the minimum eigenvalue of \(\Omega\), and \([10^2]^T\) (the AR parameter vector) is the corresponding eigenvector properly normalised. In practice, consistent estimates of \(\lambda\) and \(\theta\) can therefore be obtained by determining the minimum eigenvalue and eigenvector of a consistent sample estimate of \(\Omega\). In the following, we will consider a specific form for the sample estimator of \(\Omega\). However, since our analysis is asymptotic, the results that we obtain apply to all currently used sample covariance estimators (which are asymptotically equivalent).

Introduce the following notation

\[R = \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-n) \end{bmatrix} = \begin{bmatrix} x(t-1), \ldots, x(t-n) \end{bmatrix} \]

\[r = \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y(t) \end{bmatrix} \]

\[s = \frac{1}{N} \sum_{t=1}^{N} y^2(t) \]

where \(N\) denotes the number of data points. According to the analysis above, consistent estimates \(\hat{\lambda}\) and \(\hat{\theta}\) of \(\lambda\) and \(\theta\) can be obtained from the following equation

\[
\begin{bmatrix}
s - \hat{\lambda} \\
r
\end{bmatrix}
\begin{bmatrix}
r^T \\
r - \hat{\lambda} I
\end{bmatrix}
= 0
\tag{7}
\]

which can be written as

\[
\begin{cases}
s - \hat{\lambda} + r^T \hat{\theta} = 0 \\
r + (R - \hat{\lambda} I) \hat{\theta} = 0
\end{cases}
\tag{8}
\]

Recall that \(\hat{\lambda}\) is the smallest solution of eqn. 8. The asymptotic behaviour of the estimation errors \([\hat{\lambda} - \lambda](\hat{\theta} - \theta)^T\) properly normalised, is established in the next section.

3 Asymptotic analysis of the joint estimation errors

Let us introduce the following additional notation

\[
\Gamma = E \begin{bmatrix} x(t-1) \\ \vdots \\ x(t-n) \end{bmatrix} = \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-n) \end{bmatrix} - \lambda I
\]

\[
\rho = E \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-n) \end{bmatrix} = E \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-n) \end{bmatrix}
\tag{9}
\]

It follows from eqn. 8 that for sufficiently large \(N\), one can write

\[
0 = s - \hat{\lambda} + r^T \hat{\theta} + r^T (\theta - \theta)
= s - \hat{\lambda} + r^T \hat{\theta} + r^T (\theta - \theta)
\tag{10a}
\]

and

\[
0 = r + (R - \hat{\lambda} I) (\hat{\theta} - \theta) + (R - \hat{\lambda} I) \theta
= r + \Gamma (\hat{\theta} - \theta) + [(R - \hat{\lambda} I) - \Gamma] (\theta - \theta)
\tag{10b}
\]

The terms neglected in eqn. 10 go to zero faster than \([\hat{\lambda} - \lambda](\theta - \theta)^T\) as \(N\) tends to infinity (this is so since \(\hat{\theta}\), \(r\) and \(R - \hat{\lambda} I\) are consistent estimates of \(\theta\), \(\rho\) and \(\Gamma\), respectively). It follows from eqn. 10 that for large enough \(N\) (neglecting the higher-order terms), we have

\[
\begin{bmatrix}
(\hat{\lambda} - \lambda) - r^T (\theta - \theta) = s + r^T (\hat{\theta} - \theta)
(\hat{\lambda} - \hat{\lambda}) \theta - \Gamma (\hat{\theta} - \theta) = (r + R \theta) - \hat{\lambda} \theta
\end{bmatrix}
\]

which can be written more compactly as

\[
\begin{bmatrix}
1 \\
\theta
\end{bmatrix}
\begin{bmatrix}
\rho^T \\
\theta - \hat{\lambda}
\end{bmatrix}
= \begin{bmatrix}
s + r^T (\hat{\theta} - \lambda) \\
(r + R \theta) - \lambda \theta
\end{bmatrix}
\tag{11}
\]

The matrix

\[Q = \begin{bmatrix}
1 & \rho^T \\
\theta & \Gamma
\end{bmatrix}
\tag{12}
\]

which appears in eqn. 11, is nonsingular. This can be shown as follows. Using a standard result for the determinant of partitioned matrices, we obtain

\[\det Q = \det (\Gamma - \theta \rho^T),\]

which is nonzero since the matrix \(\Gamma - \theta \rho^T\) is nonsingular [1]. Note, for completeness, that the nonsingularity of \(\Gamma - \theta \rho^T\) can readily be shown. Since the matrix \(\Gamma\) is positive definite and since \(\rho = -\Gamma \theta\) [see eqns. 2a or 4] it follows that

\[\det Q = \det (\Gamma + \theta \rho^T \Gamma) = \det (I + \theta \rho^T) \cdot \det \Gamma > 0\]

which concludes the proof that \(Q\) is nonsingular.

We can now state the main result of this paper.
Theorem: Consider the estimates \( \hat{\lambda} \) and \( \hat{\theta} \) of \( \lambda \) and \( \theta \) obtained by the Pisarenko method (eqn. 8). When \( N \) tends to infinity, the normalised estimation error vector
\[
v = \sqrt{(N)} \begin{bmatrix} \hat{\lambda} - \lambda \\ -\hat{\theta} - \theta \end{bmatrix}
\]
converges in law to a normal distribution of zero mean and covariance \( P \)
\[
v \xrightarrow{law} \mathcal{N}(0, P) \quad (13)
\]
where the covariance matrix \( P \) is given by
\[
P = Q^{-1} P_0 Q^{-T} \quad (14a)
\]
The \( i, j \) element of \( P_0 \) in eqn. 14a has the following expression
\[
[P_0]_{ij} = \lambda^2 \sum_{k=0}^{n} \left[ a_i a_{i+j-k} + a_j a_{i+j-k} + \mu - 3 \lambda^2 a_i a_j \right] 0 \leq i, j \leq n \quad (14b)
\]
where \( \mu \) denotes the fourth order moment of \( e(t) \), \( \mu = E(e(t))^4 \), and where the following convention was used: \( a_0 = 1, a_k = 0 \) for \( k < 0 \) or \( k > n \).

Proof: It follows from eqns. 3 and 6 that
\[
s + r^2 \theta = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-n) \end{bmatrix} \cdot A(q^{-1}) e(t) = \frac{1}{N} \sum_{i=1}^{N} \phi(t) \cdot A(q^{-1}) e(t) \quad (15)
\]
where
\[
\phi(t) = [y(t), y(t-1), \ldots, y(t-n)]^T
\]
Inserting eqn. 15 into eqn. 11 we get
\[
v = Q^{-1} \frac{1}{\sqrt{(N)}} \sum_{i=1}^{N} \left( \phi(t) \cdot A(q^{-1}) e(t) - \frac{1}{\theta} \right) \quad (16)
\]
Clearly, from eqns. 3 and 4, the quantity between the curly brackets has zero mean. Thus \( Ev = 0 \) for all \( N \). It follows from eqn. 16, from some standard results on convergence in law of stochastic variables (see, e.g., Reference 10), and from the central limit theorem in Reference 9 (also see Reference 5) that, when \( N \) tends to infinity, \( v \) converges in law to a normal distribution with zero mean and covariance matrix given by
\[
P = Q^{-1} P_0 Q^{-T} \quad (17)
\]
where
\[
P_0 \triangleq \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{N} E\left[ \phi(t) \cdot A(q^{-1}) e(t) - \frac{1}{\theta} \right]^T
\]
\[
\times \left( \phi(s) \cdot A(q^{-1}) e(s) - \frac{1}{\theta} \right)^T
\]
Note that \( P_0 \) is the asymptotic covariance matrix of the vector that multiplies \( Q^{-1} \) in eqn. 16.

Introduce the following definitions
\[
\phi(t) = \begin{bmatrix} x(t) \\ x(t-1) \vdots x(t-n) \end{bmatrix} = \text{the signal part of } \phi(t)
\]
\[
\psi(t) = \begin{bmatrix} e(t) \\ e(t-1) \vdots e(t-n) \end{bmatrix} = \text{the noise part of } \phi(t)
\]
Since by assumption \( \phi(t) \) and \( \psi(s) \) are uncorrelated for all \( t \) and \( s \), it follows that
\[
P_0 = P_1 + P_2 \quad (18)
\]
where
\[
P_1 \triangleq \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{N} E(\phi(t) \cdot A(q^{-1}) e(t)) \times \phi^T(s) \cdot A(q^{-1}) e(s)
\]
\[
\times \left( \phi^T(s) \cdot A(q^{-1}) e(s) - \frac{1}{\theta} \right) \quad (19)
\]
\[
P_2 \triangleq \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{N} E(\psi(t) \cdot A(q^{-1}) e(t) - \frac{1}{\theta}) \times \left( \psi^T(s) \cdot A(q^{-1}) e(s) - \frac{1}{\theta} \right) \quad (20)
\]
The limit matrix \( P_1 \) is equal to zero. To see this, first note that
\[
P_1 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{N} E(\phi(t) \cdot A(q^{-1}) e(t) \cdot A(q^{-1}) e(t))
\]
\[
\times \left( \phi^T(s) \cdot A(q^{-1}) e(s) - \frac{1}{\theta} \right) \quad (21)
\]
A simple calculation gives
\[
E(A(q^{-1}) e(t) \cdot A(q^{-1}) e(t)) = \sum_{i=0}^{n} \sum_{j=0}^{n} a_i a_j E(e(t-i) e(t-j)) = \lambda \sum_{i=0}^{n} a_i a_i \quad (22)
\]
which clearly is zero for \( |t| > n \) (recall the convention that \( a_0 = 1 \) and \( a_k = 0 \) for \( k < 0 \) or \( k > n \)). Inserting eqn. 22 into eqn. 21, we get
\[
P_1 = \lim_{N \to \infty} \frac{N}{N} \sum_{i=1}^{N} \sum_{s=1}^{N} E(\phi(t) \cdot A(q^{-1}) e(t)) \times \left( \phi^T(s) \cdot A(q^{-1}) e(s) - \frac{1}{\theta} \right) \quad (23)
\]
Next, consider evaluation of \( P_2 \). It can be readily verified that (see, for example, References 5 and 6)
\[
E(e(t-i) e(t-j)) = \lambda^2 \delta_{i,j} + \delta_{i-L, j} + \delta_{i-L-k, j} + \delta_{i-L-k, j} \quad (24)
\]
Let \([P_{ij}]_{ij}\) denote the \(i,j\) element of \(P\), for \(i,j = 0, 1, \ldots, n\). Identity expr. 24 above and some straightforward calculations give

\[
[P_{ij}] = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \left[ \sum_{k=0}^{N} \sum_{l=0}^{N} a_{ik} a_{jk} \right] \\
= \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \left[ \sum_{k=0}^{N} \sum_{l=0}^{N} a_{ik} a_{jk} \right] + \lambda^2 a_{ii} a_{jj} \\
+ \lambda^2 a_{ii} a_{jj} + (\mu - 3\lambda^2) a_{ij} \delta_{ij} - \lambda^2 a_{ij}
\]

To complete the proof note that

\[
\sum_{t=-n}^{n} a_{++}(a_{++}) = \sum_{t=-n}^{n} a_{ii} a_{jj}
\]

Expr. 14 for \(P\) now follows from eqns. 17, 18, and 25, and the proof is finished.

4 Asymptotic covariance of the estimated AR parameters

The asymptotic covariance matrix of the estimated AR parameter vector \(\theta\) can be obtained from \(P\) by straightforward manipulations. We state the relevant result in the following.

**Lemma:** The asymptotic covariance matrix of the normalised estimation errors \(\sqrt{N(\theta - \theta)}\) is given by

\[
P_0 = \begin{pmatrix}
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\vdots & \vdots & \ddots & \vdots \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2
\end{pmatrix}
\]

or

\[
P_0 = \begin{pmatrix}
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\vdots & \vdots & \ddots & \vdots \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2
\end{pmatrix}
\]

Proof: By definition we have

\[
P_0 = \begin{pmatrix}
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\vdots & \vdots & \ddots & \vdots \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2
\end{pmatrix}
\]

where \(P_0 = \begin{pmatrix}
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\vdots & \vdots & \ddots & \vdots \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2
\end{pmatrix}\) is given by eqn. 14. Application of a standard formula for partitioned inverses to \(Q\) (see, for example, Reference 21) gives

\[
Q^{-1} = \begin{pmatrix}
1 & \rho \\
\rho & \Gamma\end{pmatrix}
\]

Using eqn. 30 in eqn. 29, we get

\[
P_0 = \begin{pmatrix}
1 & \rho \\
\rho & \Gamma\end{pmatrix}^{-1}
\]

where

\[
P_0 = \begin{pmatrix}
1 & \rho \\
\rho & \Gamma\end{pmatrix}^{-1}
\]

Let \(u_i\) denote the \(i\)th unit vector in \(R^n\) (i.e., the \(i\)th column of the \(n \times n\) identity matrix \(I\)). Inserting the formula eqn. 27 of \(P_0\) into eqn. 32 leads to the following expression for

\[
P_0 = \lambda^2 \begin{pmatrix}
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2 \\
\vdots & \vdots & \ddots & \vdots \\
\lambda^2 & \lambda^2 & \cdots & \lambda^2
\end{pmatrix}
\]
where

$$[\hat{P}_v]_{ik} = [-a_i a_k]^T P_r [-a_k]$$

$$= \frac{1}{2} \left[ E[A(q^{-1})][a_i e(t) + e(t) - i] \cdot A(q^{-1}) \right] \times \left[ -a_k e(t) + e(t - k) + E[A(q^{-1})][a_i e(t) + e(t) - i] \right]$$

$$\times \left[ -a_k e(t) + e(t - k) - E[A(q^{-1})][a_i e(t) + e(t) - i] \right]$$

$$= \frac{1}{2} \left[ a_i a_k \cdot E[A(q^{-1})] \right]^2 \cdot a_i a_k \cdot E[A(q^{-1})]$$

$$\times \left[ A(q^{-1}) e(t - k) - a_i E[A(q^{-1})] e(t) \right]$$

$$\times \left[ A(q^{-1}) e(t) - a_k E[A(q^{-1})] e(t-k) \right]$$

$$+ a_i a_k \cdot E[A(q^{-1})] e(t) - a_k E[A(q^{-1})] e(t-k)$$

$$- a_i E[A(q^{-1})] e(t-k) - a_k E[A(q^{-1})] e(t)$$

$$+ E[A(q^{-1})] e(t-k) - E[A(q^{-1})] e(t)$$

$$= \frac{1}{2} \left[ a_i a_k \cdot E[A(q^{-1})] \right]^2 (33)$$

The covariances occurring in eqn. 33 can be evaluated using the formulas eqns. 26a and b. The result is precisely eqn. 28b, which concludes the proof.

Eqn. 28 above is identical to the expression for $P_v$ derived by Sakai [1]. Since Sakai used a different approach to obtain $P_v$, the Lemma above provides a useful cross-checking of his and our methods for studying the asymptotic estimation errors. Note that our approach appears to be more general: we were able to consider the joint error vector $v = \sqrt{N} (\hat{\lambda} - \lambda - (\hat{\theta} - \theta)^T)$ for which we established the asymptotic distribution and not only a formula for the covariance matrix. Note also from eqns. 14 and 28 that while the covariance matrix $P_v$ depends on the fourth-order moment of the noise $e(t)$, the covariance matrix $P_v$ of $\sqrt{N}(\hat{\theta} - \theta)$ depends only on the second-order properties of the data. Thus, while $P_v$ can be consistently estimated from the estimates of $\mu$ and $F$ obtained within the Pisarenko method, the covariance matrix $P_v$ cannot (since an estimate of $\mu$ is not available, unless we have information on the noise properties (such as its skewness)). However, the difficulty in estimating $P_v$ is not a major problem. Indeed, in applications we are less concerned with the noise power $\sigma^2$ than we are with the parameters $\theta$ (or equivalent in the sinusoidal frequencies for which we may want to determine concentration ellipsoids, to perform hypothesis testing, etc. For such purposes a consistent estimate of $P_v$ will suffice, estimation of the whole covariance matrix $P_v$ begin unnecessary.

Finally, we present the formula of Reference 20, which relates the covariance matrix $P_v$ and the covariance matrix $P_v$ of the vector $\sqrt{N} [\hat{\omega}_1 - \omega_1, \ldots, \hat{\omega}_m - \omega_m]^T$, with $[\hat{\omega}_i]$ being determined as the angular positions of the $m$ pairs of complex-conjugate zeros of $A(z)$. We have

$$P_v = FP_v G^T F^T$$

(34)

where

$$G = \begin{bmatrix}
\cos \omega_1 \cos 2\omega_1 & \cdots & \cos \omega_1 \omega_m \\
\vdots & \ddots & \vdots \\
\sin \omega_1 \sin 2\omega_1 & \cdots & \sin \omega_1 \omega_m \\
\sin \omega_m \cos 2\omega_1 & \cdots & \sin \omega_m \omega_m
\end{bmatrix} (2m \times 2m)$$

The formula above for $P_v$ will be useful in the numerical performance study reported in the next section.

5 Numerical performance study of the Pisarenko method and comparison with the Yule-Walker method

The statistical analysis of estimation errors of the Pisarenko method, presented in this paper and in Reference 1, is useful for studies of performance and comparison with the performance achieved by other competing methods.

A close competitor to the Pisarenko harmonic decomposition method is the Yule-Walker (YW) method. The YW method is based on the simple observation that the AR parameter vector $\theta$ satisifies the following system of linear equations (called YW equations).

$$\begin{bmatrix}
r_1 & \cdots & r_1 \\
\vdots & \ddots & \vdots \\
r_1 & \cdots & r_1 \\
\end{bmatrix} \theta = - \begin{bmatrix}
r_{n+1} \\
\vdots \\
r_{n+1} \\
\end{bmatrix}$$

(35)

Eqn. 35 is obtained by premultiplying eqn. 2 by $[y(t - n), y(t - 2n), \ldots, y(t - 2n)]^T$ and taking expectation. The YW estimator of $\theta$ is the solution of eqn. 35 where $\{r_i\}$ are replaced by some consistent sample estimates. The YW estimator is more simple to compute than the Pisarenko estimator. Fast algorithms exist that need about $2n^2$ arithmetic operations (adds and multiplies) to solve a YW system of the type of eqn. 35 (see for example, Reference 21). The YW method may, however, give poor estimates in some cases. This was, in fact, the reason for introducing some variants of the YW method, such as the overdetermined and high-order YW methods, which have improved accuracy properties [7, 17-20]. The statistical performance of the YW method for sinusoidal frequency estimation was analysed in References 8 and 19 (also see References 5 and 6). The asymptotic covariance matrix of the corresponding estimation errors $\sqrt{N}(\hat{\theta} - \theta)$ is given by

$$P_{YW} = \lambda C^{-1} S C^{-1}$$

(36)

where

$$C = \begin{bmatrix}
r_1 & \cdots & r_1 \\
\vdots & \ddots & \vdots \\
r_1 & \cdots & r_1 \\
\end{bmatrix}$$

and

$$S = E[A(q^{-1})] e(t - 1) \cdots e(t - n)$$

Note that the elements of $S$ can be evaluated using the formula of eqn. 26a, and that the covariances $r_i$, for $i \neq 1$
0, occurring in \( C \) are given by

\[
\gamma_i = \sum_{k=1}^{m} \frac{a_i^2}{2} \cos \left( \phi_k \right) \quad (\tau \neq 0)
\]  

(37)

General order relationships between the covariance matrix \( P \) of the Pisarenko estimator and \( P_{yw} \) above do not seem to exist. In most cases the difference matrix \( P_{yw} - P \) is indefinite. This means that each of these two methods may be better than the other depending on the case under consideration. In general the two methods are expected to give comparable accuracies. Concerning their statistical efficiency, both methods are expected to be quite inefficient. That is to say, their error variances are much larger than the Cramér–Rao bound (CRB). These claims are supported by the numerical evaluation of \( P \) and \( P_{yw} \) presented below.

Note that the asymptotic CRB on the variance of the estimation error \( \phi_k - \omega_k \) is given by Reference 15 (see also Reference 22)

\[
\text{CRB} (\phi_k) = \frac{12}{N^3 \cdot \text{SNR}_k}
\]  

(38)

where \( \text{SNR}_k = \frac{\alpha_k^2}{2L} \). Thus the CRB for the standard deviation of the normalised estimation error \( \sqrt{(N)(\phi_k - \omega_k)} \) is

\[
\sqrt{(N \cdot \text{CRB} (\phi_k))} = \frac{1}{N} \sqrt{\left( \frac{12}{\text{SNR}_k} \right)}
\]

The ultimate performance corresponding to the CRB eqn. 38 is achieved by the maximum likelihood estimator, for reasonably large values of \( N \) (see Reference 15).

5.1 Example: Verification of the theoretical formula for the variance of the Pisarenko estimate and comparison with the CRB

This example compares the accuracy of the Pisarenko method with the CRB. It also verifies the theoretical formula of eqn. 28 and 34 for the asymptotic variance of the estimation errors \( (\phi_k - \omega_k) \) corresponding to the Pisarenko method. With this twofold purpose in mind, we ran a number of Monte Carlo experiments from which the sample statistics of the Pisarenko frequency estimates were obtained. Each of these Monte Carlo experiments consisted of 100 independent simulation runs. The input data were

\[
y(t) = \sqrt{2} \cos \omega t + \epsilon(t)
\]  

(39)

where \( \epsilon(t) \) is a zero-mean white Gaussian noise whose variance was selected to obtain various signal-to-noise ratios (SNR). Since both \( P \) and the CRB (eqn. 38) do not depend on the phase of the sine wave, we simply choose \( \phi_k = 0 \) in eqn. 39.

Fig. 1 shows the experimental standard deviation of the normalised estimation error \( \sqrt{(N)(\phi_k - \omega_k)} \) corresponding to the Pisarenko method, determined from the Monte Carlo experiments for two values of \( \omega_k \) (\( \omega_k = 0.271 \) and \( \omega_k = 0.4n \)), two values of SNR (SNR = 0 dB and SNR = 10 dB).

---

**Fig. 1** Comparison between the experimental standard deviation (obtained from 100 independent runs) of the Pisarenko method for the data (eqn. 39), the theoretical (asymptotic) standard deviation and the (asymptotic) CRB

- a) \( \omega_k = 0.2\pi \) and SNR = 0 dB
- b) \( \omega_k = 0.2\pi \) and SNR = 10 dB
- c) \( \omega_k = 0.4\pi \) and SNR = 0 dB
- d) \( \omega_k = 0.4\pi \) and SNR = 10 dB

---
and SNR = 10 dB), and varying \( N \). Also shown in this figure are the theoretical (asymptotic) standard deviations and the CRB. Note that there is a good accordance between the experimental and the theoretically expected standard deviations. Note also that the Pisarenko method is statistically inefficient and that the departure from the CRB increases with increasing \( N \). Finally, note from Fig. 1 that the variance of the Pisarenko estimate increases faster than the CRB when SNR decreases. (See the Appendix and the next Section for more details on and an analysis of this aspect.)

5.2 Example: comparison of the accuracies of the Pisarenko and the Yule–Walker methods

We consider the case of two sine waves in additive white noise, and we compare the accuracies of the Pisarenko and the Yule–Walker estimates of the two frequencies using the ratio of their associated standard deviations. The standard deviations for the Pisarenko method can be obtained from eqns. 28 and 34, and for the Yule–Walker method from eqns. 36 and 34.

Fig. 2 shows the surface

\[
\gamma_1 = \frac{\text{St. Dev. } (\hat{\omega}_1)_{PW}}{\text{St. Dev. } (\hat{\omega}_1)_{YW}}
\]

as a function of \( \omega_1 \) and \( \omega_2 \), from three different angles. Note that the two methods have comparable accuracies for most pairs of frequencies \( \{\omega_1, \omega_2\} \). For \( \omega_1 + \omega_2 \approx \pi \), the YW method is slightly better than the Pisarenko method. For \( \omega_1 \) (or \( \omega_2 \)) close to 0 or \( \pi \), the Pisarenko method is much more accurate than the YW procedure. There are also a few other (small) areas of the square \((0, \pi) \times (0, \pi)\) for which the Pisarenko method is significantly better than YW (see Fig. 2).

It is interesting to note that the ratio of eqn. 40 does not depend on SNR. This is a consequence of the fact shown in the Appendix, that the variance of the estimation error associated with both methods is proportional to \( 1/(SNR)^2 \). Another consequence of this fact is that the error variance of these two methods increases faster than the CRB when SNR decreases. Note from eqn. 38 that the CRB is proportional to \( 1/(SNR) \).

Finally, note that the surface \( \gamma_2(\omega_1, \omega_2) \) is simply obtained as \( \gamma_2(\omega_1, \omega_2) = \gamma_1(\omega_2, \omega_1) \). For \( SNR_1 = SNR_2 \), the equality above clearly holds due to the symmetry which the problem of estimating \( \omega_1 \) and \( \omega_2 \) has in such a case. As neither \( \gamma_1 \) nor \( \gamma_2 \) depends on SNR, the equality must hold for all SNR values.

6 Conclusion

In this paper we have presented an asymptotic statistical analysis of the Pisarenko method and a comparison with the Yule–Walker method for estimating sinusoidal frequencies from noisy data. For the case of two sine waves in noise, we performed an extensive comparison between the performance of Pisarenko and YW methods delineating areas in the frequency domain where one of these two methods out-performs the other, or where they give similar accuracies.

Our results show that both the Pisarenko and the Yule–Walker methods are quite inefficient in the statistical sense. Their error variances are much larger than the Cramer–Rao bound for large data samples. It was also shown that the variances of these two methods increase faster than the CRB when the signal-to-noise ratio decreases.

Since both the Pisarenko and the Yule–Walker methods are quite inefficient statistically, they are attractive only when computational simplicity is a must. For cases where estimation accuracy is more important, the off-line maximum likelihood method of Reference 15 which is efficient, or the on-line notch filter method of Reference 16 which is nearly efficient, should be used.

7 References

8 Appendix: analysis of the dependence of $P_\omega$ on the signal-to-noise ratios

First, note that the covariance (eqn. 37) can be written as

$$r_{i-j} = \sum_{k=1}^{m} \frac{a_k^2}{2} (\cos io_k \cos jio_k + \sin io_k \sin jio_k)$$

$$= [\cos jio_1 \cdots \cos jio_m \sin jio_1 \cdots \sin jio_m] \times M$$

where

$$M = \text{diag} \left\{ \frac{a_1^2}{2}, \ldots, \frac{a_m^2}{2} \right\}$$

It follows from eqn. 41 that

$$\Gamma = [r_{i-j}]_{i,j=1}^{m} = G^M G$$

$$C = [r_{i-j}]_{i,j=1}^{m} = H M G$$

where $G$ is the matrix defined in eqn. 34, and $H$ is a matrix whose elements depend on $\{\omega_i\}$ but not on $\{\omega_i\}$ or $\lambda$; the expression of $H$ is not important for the present discussion.

Next, using eqn. 42 in eqns. 28 and 36 we get

$$(P_d)_{BS} = \Gamma^{-1} (I + \theta H^{-1})^{-1} \bar{P}_d (I + \theta H^{-1})^{-1} \Gamma^{-1}$$

$$= \lambda^2 G^{-1} M^{-1} W M^{-1} G^{-1}$$

and

$$(P_d)_{YW} = \lambda^2 G^{-1} M^{-1} H^{-1} S H^{-1} T M^{-1} G^{-1}$$

$$= \lambda^2 G^{-1} M^{-1} \bar{W} M^{-1} G^{-1}$$

where $W$ and $\bar{W}$ are $(n \times n)$ matrices whose elements do not depend on $\{\omega_i\}$ or $\lambda$; they depend on $\{\omega_i\}$ only. Inserting eqn. 43 into eqn. 34, we readily reach the conclusion that the $i,j$-element of $P_\omega$ has, for both the Pisarenko and the YW methods, the following general expression

$$[P_\omega]_{ij} = f(\omega_1, \ldots, \omega_m) (\text{SNR} \cdot \text{SNR})$$

where $\text{SNR} \triangleq \frac{\omega_i}{2\lambda}$. The functions $f(\cdot, \ldots, \cdot)$ corresponding to the two methods will in general be different. However, they do not have simple expressions and hence the dependence of $P_\omega$ on $\{\omega_i\}$ is rather involved for both methods. However, the dependence of $P_\omega$ on $\{\text{SNR}\}$ is simple. Moreover, both $(P_d)_{BS}$ and $(P_d)_{YW}$ depend on $\{\text{SNR}\}$ in the same way. Note in particular that for both methods

$$[P_\omega]_{ij} = \text{Var}(\hat{\omega}_i) \sim 1/\text{SNR}^2$$

The dependence of $(P_d)_{BS}$ on SNR has also been studied in Reference 1 for the case of two sinusoids ($m = 2$).