

Employing Symmetry Constraints for Improved Frequency Estimation by Eigenanalysis Methods*

Gordon K. Smyth

Department of Mathematics, University of Queensland
Australia 4072

Abstract

The problem of extracting sinusoid signals from noisy observations made at equally spaced times is considered. Eigenanalysis methods, such as Pisarenko's method and the extended Prony method, find the eigenvector with minimum eigenvalue of a suitably chosen matrix, and then obtain the complex sinusoids as the roots of the polynomial which has the components of the eigenvector as coefficients. For the sinusoids to be undamped, it is necessary that the roots lie on the unit circle, and hence that the eigenvector be conjugate symmetric. It is shown how this symmetry constraint can be incorporated into eigenanalysis estimation methods in a routine way. The practical importance of the constraint is investigated by Cramér-Rao variance bound calculations and by simulation. Three data examples are included. The following conclusions are made: (i) The symmetry constraint is straightforward to implement, and reduces the amount of computation required. (ii) The relative reduction in variance of the constrained over the unconstrained frequency estimators is arbitrarily large for frequencies close together or near a multiple of π . However there are also frequency values for which the symmetry constraint gives no reduction in variance at low noise-to-signal ratios. (iii) The relative reduction in variance converges to zero for large sample sizes. (iv) The symmetry constraint increases the breakdown noise-to-signal ratios above which the various methods fail to give useful results.

Keywords: Prony's method; Pisarenko's harmonic decomposition; sinusoid signals; complex exponentials; least squares; discrete spectra.

1 Introduction

Time series consist of observations recorded regularly over time which show ongoing patterns or periodicities that need to be decoded. Time series are common in most

* *Technometrics*, August 2000, **42**, 277–289.

areas of science and engineering, especially in areas such as geophysics, speech recognition and electronic signal processing. A traditional method of identifying periodicities in time series data is to model the series y_t as a sum of sinusoidal signals,

$$y_t = \mu_t + \epsilon_t, \tag{1}$$

$t = 1, \dots, n$, where

$$\mu_t = \sum_{k=1}^p \alpha_k \sin(\omega_k t + \phi_k) \tag{2}$$

and where ϵ_t represents an error process with mean zero and constant variance. The sum of sinusoids model is said to have a *discrete spectra*, as there are a finite number of discrete frequencies ω_k . The sinusoidal signals are sometimes called *hidden periodicities* or *harmonics*.

Discrete spectra are usually generated by mechanisms that operate with high precision and regularity. Thus, light spectra produced by the motion of electrons between prescribed energy levels in atoms are discrete. The frequency decomposition of a tone produced by the human voice or a musical instrument is discrete because only finitely many “modes” of vibration are possible in human vocal cords and the vibrating mechanism of the instrument. Problems involving the identification of discrete frequencies arise in many areas such as laser signals, speech recognition, industrial process control, communications, radio location of objects, seismic signal processing and computer-assisted medical diagnosis.

For example, in speech recognition and speech synthesis, voice tones can be modeled using a small number of resonant frequencies (formants) (Reddy, 1967; Schafer and Rabiner 1969; Benade, 1990). Accurate identification of the frequencies and their corresponding amplitudes allows vowels and other tones to be automatically distinguished and reproduced.

Less regular physical processes often produce time series with a smooth spectral response, which behave as if there are many small contributions at infinitely many frequencies (Priestley, 1981). Even in these situations there is often a discrete spectral component, perhaps because of some characteristic of the measuring equipment. In this case the interest in estimating the discrete component may be to remove it from the series so that the remaining continuous spectra can be studied in isolation, a process which is known as “complex demodulation”. In manufacturing process control, identification of periodic behavior in a production process is often of interest as it may signal a process which is responding to environment conditions and is therefore in an out of control state.

This paper is concerned with the classic problem of extracting sinusoid signals from observations made at equally spaced times, where the ϵ_t are uncorrelated errors with mean zero and constant variance σ^2 . A nonlinear regression approach is taken in which the α_k , ω_k and ϕ_k are viewed as fixed parameters to be estimated by least squares or other appropriate criteria. The ratio of the amplitude to the error standard deviation α_k/σ represents the signal-to-noise ratio for the k th signal. In general, the larger the signal-to-noise ratio, the easier it is to estimate the corresponding frequency for a given number of observations.

It is often convenient to represent (2) as a sum of complex exponentials,

$$\mu_t = \sum_{k=1}^{2p} \alpha_k \exp(i\omega_k t) \quad (3)$$

in which the frequencies ω_k are real but the α_k may be complex (Marple, 1987; Kay, 1988). This complex signal reduces to a real signal of the form (2) if the complex exponentials and corresponding coefficients occur in conjugate pairs. The fully complex version of the frequency estimation problem has also been studied. In the complex version the errors have imaginary parts which are uncorrelated with, and of equal variance as, the real parts. The complex problem with n observations is essentially equivalent to the corresponding real problem with $2n$ observations, and can be obtained from it using an ideal Hilbert transform (James et al, 1994, Appendix A). Complex data might also arise directly where a real signal has been subject to complex modulation before being observed with error.

Both the real and complex frequency estimation problems have attracted enormous attention in the statistical and signal processing literature. If the errors ϵ_t are approximately normal, then maximum likelihood estimation of the frequencies results in a nonlinear least squares problem. This particular problem though is not amenable to standard nonlinear least squares methods because of the existence of many local minima of the sum of squares (Rice and Rosenblatt, 1988; Smyth and Hawkins, 1997). Standard methods such as Gauss-Newton are likely to converge to a local minima rather than the global minimum unless exceptionally good starting values are supplied. There are also difficulties with numerical stability and with convergence when there are frequencies close together. For these reasons, frequency estimation is usually undertaken by autoregressive methods or by eigenanalysis methods (Mackisack et al, 1994). In this paper we show, by combining three eigenanalysis algorithms of increasing complexity, that convergence to the global likelihood maximum can be obtained in virtually all cases without having to supply starting values, except when the noise to signal ratio is extremely high. This essentially solves the starting value problem for maximum likelihood estimation of frequencies which was described by Rice and Rosenblatt (1988).

Most eigenanalysis methods of estimation are based on a parametrization which was proposed in a, now classical, paper by Gaspard Riche, Baron de Prony (Prony, 1795). Gaspard Riche observed that a linear combination of exponential functions satisfies a difference equation of the form

$$c_1 \mu_t + \dots + c_{2p+1} \mu_{t+2p} = 0 \quad (4)$$

for all t , and hence derived a system of linear equations for interpolating a sum of p exponentials through $2p$ points. Modern eigenanalysis methods find $\mathbf{c} = (c_1, \dots, c_{2p+1})^T$ as the eigenvector with minimum eigenvalue of a suitably chosen matrix B , and then find the frequencies as the angular positions of the roots of the polynomial

$$C(z) = c_1 + c_2 z + \dots + c_{2p+1} z^{2p}$$

Best known is Pisarenko's harmonic decomposition method (Kay and Marple, 1981) for which B is an estimate of the "covariance matrix" of the y_t (viewing the y_t as a

stationary process with mean zero). Other eigenanalysis methods have been proposed by Osborne (1970, 1975) in which B is a function of \mathbf{c} and the eigenproblem is solved iteratively. Other methods can be expressed as eigenproblems, although they were originally proposed in terms of quadratic minimizations. These include the extended Prony method (Kay and Marple, 1981), the IQML method of Bresler and Macovski (1986), and methods proposed by Kumaresan and Shaw (1985), Kumaresan, Scharf and Shaw (1986) and Evans and Fischl (1973). Related procedures have been proposed by Matausek, Stankovic and Radivic (1983) and Kay (1984).

It is a feature of the Prony parametrization that it is not restricted to periodic signals. For suitable choices of \mathbf{c} , the signal may include damped sinusoids, real or complex exponentials, and, if there are repeated roots, damped polynomials. See Brockwell and Davis (1987, Section 3.6) or Osborne and Smyth (1995) for some discussion. For general applications, this generally is an advantage of the Prony parametrization. However for time series applications we may wish to consider only periodic signals. A necessary and sufficient condition for the signal to be of the form (2) is that the roots of $C(z)$ are distinct and lie on the unit circle. For the roots to lie on the unit circle is necessary that the vector \mathbf{c} be conjugate symmetric. A number of authors, including Marple (1980), Kay and Marple (1981), Chan, Lavoie and Plant (1981), Nehorai (1985) and Bresler and Macovski (1986), have proposed that this symmetry constraint on \mathbf{c} be incorporated into the estimation method.

Recently Kannan and Kundu (1994) and Kundu and Kannan (1997) have studied the symmetry constraint for a number of frequency estimation algorithms including the eigenanalysis algorithm of Osborne (1975) and Kundu (1993). They show that the constraint-modified eigenanalysis algorithm performs better than the other algorithms in experiments with two simulated models, one with two complex radian frequencies and another with one real frequency. Cramér-Rao variance bounds are given for the simulated models but not for the constrained estimators. The purpose of the current paper is two-fold. We show that the symmetry constraint can be implemented in a unified way for the class of eigenanalysis methods by projecting the matrix B onto a suitable space, thus generalizing the algorithm of Kundu and Kannan. We also study how the efficiency gained from the symmetry constraint varies as the data generating process changes. We show that the effect of the constraint, i.e., the difference in performance between the constrained and unconstrained algorithms, depends critically on the underlying frequencies, on their separations, on the sample size and on the error standard deviation. We show that the constraint is most important when it is most needed, i.e., in the most difficult cases. However there are other parameter values for which the constraint has essentially no effect at all. Variance bound calculations are used to study effect of the constraint as a function of the frequency values and their separations. Simulations are used to assess how closely the actual performance of the algorithms follow the theoretical variance bounds in small samples. The simulations allow further conclusions in terms of the error standard deviation and breakdown noise to signal ratio.

The main conclusions of the paper are summarized below.

- The symmetry constraint is straightforward to implement for eigenanalysis methods, and reduces rather than increases the amount of computation required.

- The relative reduction in variance of the constrained over the unconstrained frequency estimators is arbitrarily large for frequencies close together or, in the real case, near a multiple of π . However, for each n and p , there are also regularly spaced frequency values for which the symmetry constraint gives no reduction in variance at all at low or moderate noise to signal ratios.
- The relative reduction in variance converges to zero as $n \rightarrow \infty$ for any set of frequency values.
- The symmetry constraint increases the breakdown noise level above which each method ceases to give useful results. The constrained algorithms can tolerate a noise to signal ratio nearly twice as high as can the unconstrained algorithms.

The plan of the remainder of the paper is as follows. Section 2 describes the various eigenanalysis methods. Section 3 shows how the symmetry constraint can be incorporated into the methods by projecting the eigenvectors onto a restricted subspace. Section 4 assesses the resultant reduction in the variances of the frequency estimators, using variance bound calculations and simulation. Three data examples are given in Section 5. The paper finishes with a brief summary.

2 Eigenanalysis methods

For any sinusoidal signal (2) there is a unique vector $\mathbf{c} = (c_1, \dots, c_{2p+1})^T$ satisfying the difference equation (4). Let $\mathbf{y}_t = (y_t, y_{t+1}, \dots, y_{t+2p})^T$ where the y_t are as given in (1). The basic idea behind eigenanalysis methods is that $\mathbf{c}^T \mathbf{y}_t$ has expectation zero for all t , and therefore $\|\mathbf{a}^T \mathbf{y}_t\|/\|\mathbf{a}\|$ is in some sense minimized by $\mathbf{a} = \mathbf{c}$. Eigenanalysis methods actually compute a suitable matrix B for which $\mathbf{a}^T E(B)\mathbf{a}/\|\mathbf{a}\|$ is minimized by $\mathbf{a} = \mathbf{c}$, and find the eigenvector of B with smallest eigenvalue. The matrix B is essentially a matrix of sums of squares and cross-products of y_t with y_t at various lags, but possibly normalized in a way which depends on \mathbf{c} .

Consider the following general eigenanalysis iteration. Let \mathbf{c}^k be a working estimate of \mathbf{c} . The updated estimate is defined as the solution \mathbf{c}^{k+1} of

$$(B(\mathbf{c}^k) - \lambda^{k+1}D)\mathbf{c}^{k+1} = 0 \quad (5)$$

where B is a suitably defined matrix function of \mathbf{c} , D is a constant symmetric matrix and λ^{k+1} is the closest to zero of such solutions. The eigenproblem is repeated until convergence. The ω_i are then obtained as the angular positions of the roots of the polynomial $C(z)$, and the α_k and ϕ_k are obtained from a linear least squares calculation. The eigenproblem (5) is equivalent to the constrained quadratic minimization problem

$$\min_{\mathbf{c}^{(k+1)*} D \mathbf{c}^{k+1} = 1} \mathbf{c}^{(k+1)*} B(\mathbf{c}^k) \mathbf{c}^{k+1}$$

with λ^{k+1} the minimized value. Here $\mathbf{c}^{(k+1)*} D \mathbf{c}^{k+1} = 1$ is the scaling constraint. Common choices for D are $D = I$ corresponding to the constraint $\|\mathbf{c}\| = 1$, or $D = \text{diag}(0, \dots, 0, 1)$ corresponding to the constraint $c_{2p+1} = 1$. The implications of the choice of D are discussed below.

The simplest eigenanalysis method is Pisarenko's. Pisarenko's method uses $D = I$, and specifies that B be a sum of squares and cross-products matrix such as

$$B = \frac{1}{n} Y^* Y$$

where

$$Y = \begin{pmatrix} y_1 & \cdots & y_{2p+1} \\ \vdots & & \vdots \\ y_{n-2p} & \cdots & y_n \end{pmatrix}$$

Different choices of B for Pisarenko's method are discussed by Kay and Marple (1981). The above choice for B is called the "covariance matrix" of y_t in the signal processing literature, although the y_t are not corrected for their mean, and the resulting eigenanalysis method is called the Pisarenko covariance method. The use of the term "covariance" is a reference to the mixture model for the y_t which arises when the phases ϕ_k are distributed uniformly on $[0, 2\pi)$. Under this mixture model the y_t form a stationary process with mean zero and the matrix B estimates the covariance matrix of \mathbf{y}_t . An alternative choice for B is the Toeplitz matrix in which elements on the j th off-diagonal are $(1/n) \sum_{t=1}^{n-j} y_t y_{t+j}$. The use of a Toeplitz matrix has the advantage that the eigenvector is automatically conjugate symmetric, but in this case $E(B)$ does not have \mathbf{c} as an eigenvector and the frequency estimators derived from B are biased. Another Toeplitz matrix has elements given by $\{1/(n-j)\} \sum_{t=1}^{n-j} y_t y_{t+j}$. This matrix is an unbiased estimator of the covariance of \mathbf{y}_t under the above mixture model and as a consequence the frequency estimators derived from it are less biased than those from the previous matrix, but it is not guaranteed to be positive definite. In the next section of this paper we show how to achieve the effect of the Toeplitz matrices without introducing bias or non-positive definiteness.

Although not usually expressed as an eigenproblem, the extended Prony method (Kay and Marple, 1981) is as for Pisarenko's method but with scaling constraint $c_{2p+1} = 1$ instead of $\|\mathbf{c}\| = 1$. This choice of scaling has the computational advantage that the eigenproblem (5) can be reformulated as a linear least squares problem. Kahn et al (1992) have examined the asymptotic efficiencies of different scaling constraints, and show that the Pisarenko scaling is preferable to the extended Prony scaling unless the signal to noise ratio is very small.

The Pisarenko covariance method minimizes the sum of squares of $Y\mathbf{c}$ with respect to \mathbf{c} . The algorithm proposed by Osborne (1970) differs from Pisarenko's method in that it minimizes a weighted sum of squares. The vector $Y\mathbf{c}$ is normalized by its covariance matrix $M(\mathbf{c}) = X^* X$, where

$$X^* = \begin{pmatrix} c_1 & \cdots & c_{2p+1} & & 0 \\ & \ddots & & \ddots & \\ 0 & & c_1 & \cdots & c_{2p+1} \end{pmatrix}$$

At each iteration Osborne's algorithm minimizes $\mathbf{c}^{(k+1)*} Y^* M(\mathbf{c}^k) Y \mathbf{c}^{k+1}$ with respect to \mathbf{c}^{k+1} subject to $\|\mathbf{c}^{k+1}\| = 1$, resulting in an eigenproblem of the form (5) with

$$B(\mathbf{c}) = \frac{1}{n} Y^* M(\mathbf{c})^{-1} Y$$

and $D = I$. Bresler and Macovski (1986) independently proposed the same algorithm, as a sequence of constrained quadratic minimizations rather than as a sequence of eigenproblems, and implemented the Prony constraint $\text{real}(c_{2p+1}) = 1$ rather than the Pisarenko constraint $\|\mathbf{c}\| = 1$. Kahn et al (1992) discuss the effectiveness of several scaling constraints, including the two given here. With either scaling constraint, this algorithm produces estimators with a higher order of accuracy than those of the Pisarenko or Prony methods.

An eigenanalysis method which returns the actual least squares estimators of the α_k , ω_k and ϕ_k was proposed by Osborne (1976) and further studied by Osborne and Smyth (1991, 1995) and Kundu (1993). This algorithm uses

$$B = \frac{1}{n}(Y^* M^{-1} Y - V^* V)$$

where $V^* V$ is a term which arises from the differentiation of M^{-1} with respect to \mathbf{c} . This algorithm has the property that the eigenvalue λ^k converges to zero as the algorithm converges. Hence it returns estimators which are invariant with respect to the scaling constraint defined by D (Osborne and Smyth, 1995).

Kumaresan and Shaw (1985) and Kumaresan et al (1986) proposed a two-phase iterative quadratic minimization algorithm. Their first phase can be shown to be equivalent to the Osborne/Bresler/Macovski algorithm, while their second stage is analogous to Osborne's least squares algorithm described above.

The asymptotic properties of least squares estimators for real sinusoids have been considered by, for example, Whittle (1951, 1953), Walker (1971), Hannan (1973) and Hasan (1982). The least squares estimators of the frequencies have variances which are asymptotically $O(n^{-3})$ while those of the α_k and ϕ_k are $O(n^{-1})$. Rice and Rosenblatt (1988) show, however, that the sum of squares has many local minima at a separation of $O(n^{-1})$ in the frequencies, and that the estimates of the amplitudes and phases are consistent only if the frequencies are resolved to the global rather than to a local minimum. Mackisack et al (1991) show that the same qualitative behavior, $O(n^{-3})$ convergence of its estimator but many local minima, is found in terms of the eigenvector \mathbf{c} as well as in terms of the frequencies.

The non-least squares algorithms are generally less efficient. Pisarenko's method, for example, returns frequency estimates with $O(n^{-1})$ variances (Sakai, 1984; Stoica and Nehorai, 1988; Kahn et al, 1992). An exception is the Osborne/Bresler/Macovski algorithm, which Kahn et al (1992) show is comparable in efficiency to least squares.

The extended Prony and Pisarenko algorithms are non-iterative and do not require starting values. Osborne's least squares algorithm, as with all least squares based frequency methods, requires excellent starting values. In this study, the Osborne/Bresler/Macovski algorithm proved to be far less sensitive to starting values than the least squares algorithm, and converged in almost all cases when started from the Pisarenko estimates. The least squares algorithm usually converged to a local minimum if started from the Pisarenko estimates, but almost always found the global minimum when started from the Osborne/Bresler/Macovski estimates, unless the noise level was particularly high.

A three stage algorithm can therefore be recommended for computing the least squares estimators. Pisarenko's method is used to start the Osborne/Bresler/Macovski

algorithm, which is turn is used to start the least squares algorithm. This hybrid algorithm proved to be a very reliable method of computing the least squares estimators without the need to provide starting values.

3 Constrained eigenanalysis methods

Suppose z is a root of the polynomial $C(z)$. If z lies on the unit circle, then $\bar{z} = z^{-1}$, and $0 = \bar{C}(z) = \bar{c}_1 + \bar{c}_2\bar{z} + \dots + \bar{c}_{2p+1}\bar{z}^{2p} = \bar{c}_1 + \bar{c}_2z^{-1} + \dots + \bar{c}_{2p+1}z^{-2p}$, so $\bar{c}_{2p+1} + \bar{c}_{2p}z^2 + \dots + \bar{c}_1z^{2p} = 0$. Comparing this with $C(z) = 0$ shows that $c_j = \bar{c}_{2p+2-j}$ for $j = 1, \dots, p$, i.e., \mathbf{c} must be conjugate symmetric. This condition on \mathbf{c} is necessary but not sufficient for the roots of $C(z)$ to lie on the unit circle. Conjugate symmetry guarantees only that if a root z occurs, then so does it's reciprocal z^{-1} , not that $|z| = 1$. A fully sufficient condition would involve inequality constraints on the c_j which could not easily be incorporated into an eigen-calculation. Conjugate symmetry however is locally sufficient for \mathbf{c} in a neighborhood of the true values.

The symmetry constraint can be incorporated into the eigenproblem using the reduction method suggested by Osborne and Smyth (1995, Section 5). Let Q_1 and Q_2 be the $(2p + 1) \times p + 1$ and $(2p + 1) \times p$ matrices

$$Q_1 = \begin{pmatrix} I_p/\sqrt{2} & 0 \\ 0 & 1 \\ J_p/\sqrt{2} & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} I_p/\sqrt{2} \\ 0 \\ -J_p/\sqrt{2} \end{pmatrix}$$

where I_p is the $p \times p$ identity matrix and J_p is the $p \times p$ anti-diagonal matrix

$$J_p = \begin{pmatrix} 0 & & & 1 \\ & & & \cdot \\ & & \cdot & \\ & 1 & & 0 \end{pmatrix}$$

Then the symmetry constraint can be represented as $\mathbf{c} = Q_1\boldsymbol{\gamma}_1 + iQ_2\boldsymbol{\gamma}_2$ where $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are unrestricted real vectors of dimension $p + 1$ and p respectively.

Suppose for the moment that the observations y_t are real. In that case, B and \mathbf{c} also will be real, so $\boldsymbol{\gamma}_2 = 0$ and the eigenproblem (5) reduces to

$$(Q_1^T B(\mathbf{c}^k) Q_1 - \lambda^{k+1} Q_1^T D Q_1) \boldsymbol{\gamma}_1^{k+1} = 0 \tag{6}$$

A constrained eigenanalysis method consists of iteratively solving (6) for $\boldsymbol{\gamma}_1$, then recovering \mathbf{c} from $\hat{\mathbf{c}} = Q_1\hat{\boldsymbol{\gamma}}_1$. Note that Q_1 has been chosen to satisfy $Q_1^T Q_1 = I$, so (6) is an ordinary eigenproblem if $D = I$.

The transformation from \mathbf{c} to $\boldsymbol{\gamma}$ may be interpreted by noting that $Q_1 Q_1^T$ is the orthogonal projection onto the linear space of symmetric \mathbf{c} . The effect of the transformation is to replace B and D in (5) with $Q_1 Q_1^T B Q_1 Q_1^T$ and $Q_1 Q_1^T D Q_1 Q_1^T$, both of which are Toeplitz matrices. A related but different method of incorporating the symmetry constraint into the extended Prony method is given by Kay and Marple (1981).

Now consider the case in which the y_t are complex. In the complex domain, the symmetry condition on \mathbf{c} is not expressible as a linear constraint. Therefore, in order to use the above linear reduction methods, it is necessary to expand the eigenproblem (5) into a real eigenproblem of twice the dimension. The problem which the eigenproblem solves is to minimize $\mathbf{c}^*B(\mathbf{c})\mathbf{c}$ subject to a scaling constraint. Let \mathbf{c}_1 and \mathbf{c}_2 be the real and imaginary parts of \mathbf{c} and B_1 and B_2 the real and imaginary parts of $B(\mathbf{c})$. Then

$$\mathbf{c}^*B(\mathbf{c})\mathbf{c} = \begin{pmatrix} \mathbf{c}_1^T & \mathbf{c}_2^T \end{pmatrix} \begin{pmatrix} B_1 & -B_2 \\ B_2 & B_1 \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} \quad (7)$$

Suppose the scaling constraint is

$$\begin{pmatrix} \mathbf{c}_1^T & \mathbf{c}_2^T \end{pmatrix} \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = 1$$

where D_{11} and D_{22} are symmetric matrices and $D_{21} = D_{12}^T$. Introducing the Lagrange multiplier λ , the quadratic form (7) is minimized subject to the scaling constraint when $(\mathbf{c}_1^T \ \mathbf{c}_2^T)^T$ satisfies the generalized eigenproblem

$$\left[\begin{pmatrix} B_1 & -B_2 \\ B_2 & B_1 \end{pmatrix} - \lambda \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \right] \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = 0 \quad (8)$$

Introducing now the symmetry constraint $\mathbf{c}_1 = Q_1\boldsymbol{\gamma}_1$ and $\mathbf{c}_2 = Q_2\boldsymbol{\gamma}_2$, the eigenproblem is transformed to

$$\left[\begin{pmatrix} Q_1^T B_1 Q_1 & -Q_1^T B_2 Q_2 \\ Q_2^T B_2 Q_1 & Q_2^T B_1 Q_2 \end{pmatrix} - \lambda \begin{pmatrix} Q_1^T D_{11} Q_1 & Q_1^T D_{12} Q_2 \\ Q_2^T D_{21} Q_2 & Q_2^T D_{22} Q_2 \end{pmatrix} \right] \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{pmatrix} = 0 \quad (9)$$

Since Q_1 and Q_2 have been chosen orthogonal, (9) is an ordinary eigenproblem if $D_{11} = D_{22} = I$ and $D_{12} = D_{21} = 0$. A constrained eigenanalysis method consists of solving (9) iteratively until convergence.

The frequency estimates are obtained by rooting the polynomial $C(z)$. Estimates of the coefficients and phases are obtained by conditional least squares. Let ρ_k be the roots of $C(z)$ and let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^*$. The signal can be written $\boldsymbol{\mu} = A\boldsymbol{\beta}$ where A is the $n \times 2p$ matrix with (j, k) th element ρ_k^j . The least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (A^*A)^{-1}A^*\mathbf{y}$$

and the fitted signal is $\hat{\boldsymbol{\mu}} = P\mathbf{y}$ where $P = A(A^*A)^{-1}A^*$ is the orthogonal projection onto the column space $\mathcal{C}(A)$. In the complex case, the vector $\hat{\boldsymbol{\beta}}$ holds the estimates of the complex coefficients α_k . In the real case, P is real and the elements of $\hat{\boldsymbol{\beta}}$ occur in conjugate pairs. Assuming that the roots ρ_k lie on the unit circle, write $\rho_k = \exp(i\omega_k t)$. Then $\hat{\boldsymbol{\mu}}_t$ contains terms of the form

$$\hat{\beta}_k \exp(i\omega_k t) + \bar{\hat{\beta}}_k \exp(-i\omega_k t) = \alpha_k \cos(\omega_k t + \phi_k)$$

where $\alpha_k \cos(\phi_k) = 2 \text{Real}(\hat{\beta}_k)$ and $\alpha_k \sin(\phi_k) = -2 \text{Imag}(\hat{\beta}_k)$.

4 Variance Reduction

4.1 Crámer-Rao Lower Bounds

The variances of the frequency estimators can be generally expected to decrease when the signal is constrained to be periodic, because the number of freely varying parameters is reduced. Consider a sum of exponentials

$$\mu_t = \sum_{k=1}^{2p} \alpha_k \exp(\omega_k t)$$

where the α_k and ω_k are complex numbers. Let $\boldsymbol{\psi}$ be the vector of the real parts of the ω_k , and let $\boldsymbol{\theta}$ represent the other parameters. The unconstrained least squares estimators $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\theta}}$ are asymptotically multinormal with a covariance matrix which we will write as

$$\text{cov}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) = \begin{pmatrix} \Sigma_{\theta\theta} & \Sigma_{\theta\psi} \\ \Sigma_{\psi\theta} & \Sigma_{\psi\psi} \end{pmatrix}.$$

Here $\Sigma_{\theta\theta}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$, $\Sigma_{\psi\psi}$ is the covariance matrix of $\hat{\boldsymbol{\psi}}$, and $\Sigma_{\theta\psi} = \Sigma_{\psi\theta}^*$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$ with $\hat{\boldsymbol{\psi}}$. For $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\theta}}$ in a suitable neighborhood of the true values, the symmetry constraint on $\hat{\mathbf{c}}$ is equivalent to the constraint $\hat{\boldsymbol{\psi}} = 0$. The covariance matrix of the other parameters conditional on the constraint is

$$\text{cov}(\hat{\boldsymbol{\theta}} | \hat{\boldsymbol{\psi}} = 0) = \Sigma_{\theta\theta} - \Sigma_{\theta\psi} \Sigma_{\psi\psi}^{-1} \Sigma_{\psi\theta}.$$

The variance reduction due to the symmetry constraint is therefore

$$\text{cov}(\hat{\boldsymbol{\theta}}) - \text{cov}(\hat{\boldsymbol{\theta}} | \hat{\boldsymbol{\psi}} = 0) = \Sigma_{\theta\psi} \Sigma_{\psi\psi}^{-1} \Sigma_{\psi\theta}$$

which is non-negative definite. Each individual component $\hat{\theta}_j$ of $\hat{\boldsymbol{\theta}}$ will enjoy a reduced variance, unless it is entirely uncorrelated with $\hat{\boldsymbol{\psi}}$.

An asymptotically accurate expression for $\text{cov}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ is given by the Cramér-Rao lower bound for the variances of unbiased estimators. Let $\dot{\boldsymbol{\mu}}_{\theta}$ and $\dot{\boldsymbol{\mu}}_{\psi}$ be the gradient matrices of $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^*$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ respectively. The Cramér-Rao bound for the variance of the unconstrained estimators is the inverse Fisher information matrix,

$$V = \sigma^2 \begin{pmatrix} \dot{\boldsymbol{\mu}}_{\theta}^* \dot{\boldsymbol{\mu}}_{\theta} & \dot{\boldsymbol{\mu}}_{\theta}^* \dot{\boldsymbol{\mu}}_{\psi} \\ \dot{\boldsymbol{\mu}}_{\psi}^* \dot{\boldsymbol{\mu}}_{\theta} & \dot{\boldsymbol{\mu}}_{\psi}^* \dot{\boldsymbol{\mu}}_{\psi} \end{pmatrix}^{-1}$$

where σ^2 is the variance of the observational errors. In particular the bound for the variance of the unconstrained estimator of $\boldsymbol{\theta}$ is the submatrix

$$V_{\theta\theta} = \sigma^2 \left\{ \dot{\boldsymbol{\mu}}_{\theta}^* \dot{\boldsymbol{\mu}}_{\theta} - \dot{\boldsymbol{\mu}}_{\theta}^* \dot{\boldsymbol{\mu}}_{\psi} (\dot{\boldsymbol{\mu}}_{\psi}^* \dot{\boldsymbol{\mu}}_{\psi})^{-1} \dot{\boldsymbol{\mu}}_{\psi}^* \dot{\boldsymbol{\mu}}_{\theta} \right\}^{-1}$$

On the other hand, the Cramér-Rao bound for the constrained estimators is simply

$$V_c = \sigma^2 (\dot{\boldsymbol{\mu}}_{\theta}^* \dot{\boldsymbol{\mu}}_{\theta})^{-1}$$

so the increase in the Fisher information for $\boldsymbol{\theta}$ due to the symmetry constraint is

$$V_c^{-1} - V_{\theta\theta}^{-1} = \frac{1}{\sigma^2} \dot{\boldsymbol{\mu}}_{\theta}^* \dot{\boldsymbol{\mu}}_{\psi} (\dot{\boldsymbol{\mu}}_{\psi}^* \dot{\boldsymbol{\mu}}_{\psi})^{-1} \dot{\boldsymbol{\mu}}_{\psi}^* \dot{\boldsymbol{\mu}}_{\theta}$$

The size of this term depends on the level of the collinearity between $\dot{\boldsymbol{\mu}}_{\theta}$ and $\dot{\boldsymbol{\mu}}_{\psi}$. The higher the collinearity the greater will be the reduction in variance of the constrained over the unconstrained estimator of $\boldsymbol{\theta}$.

The essential features regarding collinearity and variance reduction can be seen in the case of one real sinusoid. If $p = 1$ then we can write

$$\mu_t = \alpha e^{(\omega+i\psi)t} + \bar{\alpha} e^{(\omega-i\psi)t} = e^{\psi t} \rho \sin(\omega t + \phi)$$

where α is an arbitrary complex number, ω and ψ are real numbers, and $\rho \sin(\phi)$ and $-\rho \cos(\phi)$ are the real and imaginary parts of α . In this case we can put $\boldsymbol{\theta} = (\rho, \phi, \omega)^T$. If the signal actually is sinusoidal, then $\psi = 0$ and we wish to impose the constraint $\hat{\psi} = 0$. The relative reduction in the variance of the frequency estimate $\hat{\omega}$ due to the constraint is simply the squared correlation between $\hat{\omega}$ and $\hat{\psi}$. The correlation depends on the collinearity of $\partial\mu_t/\partial\omega = \rho t \cos(\omega t + \phi)$ and $\partial\mu_t/\partial\psi = \rho t \sin(\omega t + \phi)$. (All quantities are evaluated at $\psi = 0$.) Now $\partial\mu_t/\partial\omega$ and $\partial\mu_t/\partial\psi$ become uncorrelated as $n \rightarrow \infty$ for nonzero ω , so the variance reduction must become small as n becomes large. On the other hand, $\partial\mu_t/\partial\omega$ and $\partial\mu_t/\partial\psi$ become collinear as ω approaches zero, so the relative reduction in variance approaches 100%. Putting this together, we can expect the symmetry constraint to be most important for estimating a small frequency from a small data set. The symmetry constraint helps in the most difficult case.

As a function of ω , the correlation between $\partial\mu_t/\partial\omega$ and $\partial\mu_t/\partial\psi$ also passes through zero at values close to the fundamental frequencies $j\pi/(2n)$, $j = 1, \dots, n-1$. We can therefore expect a series of discrete frequency values for which the symmetry constraint has little effect.

The remaining gradients are $\partial\mu_t/\partial\rho = \sin(\omega t + \phi)$ and $\partial\mu_t/\partial\phi = \rho \cos(\omega t + \phi)$. The first will be generally be partially collinear with $\partial\mu_t/\partial\psi$ even for large n , while the second will not. Therefore the constraint will generally reduce the variance of the amplitude $\hat{\rho}$ more than that of the phase $\hat{\phi}$.

The same sort of results apply when there are several frequencies. The pair $\hat{\omega}, \hat{\psi}$ for each frequency is asymptotically independent of all other pairs in addition to $\hat{\omega}$ and $\hat{\psi}$ being independent of each other. Therefore the relative variance reduction will become small as n becomes large. On the other hand, the relative variance reduction will be large when there are nearly equal frequencies or when there are frequencies near zero. Note that the relative reduction in variance does not depend on ρ or σ^2 , and depends only marginally on ϕ .

4.2 Numerical Calculations

In this section and the next, variance calculations are done for sample sizes $n = 20$ and $n = 50$. Even for $n = 20$, maximum likelihood estimation is able to return fairly precise estimates for reasonable signal-to-noise levels. Samples of this size are of

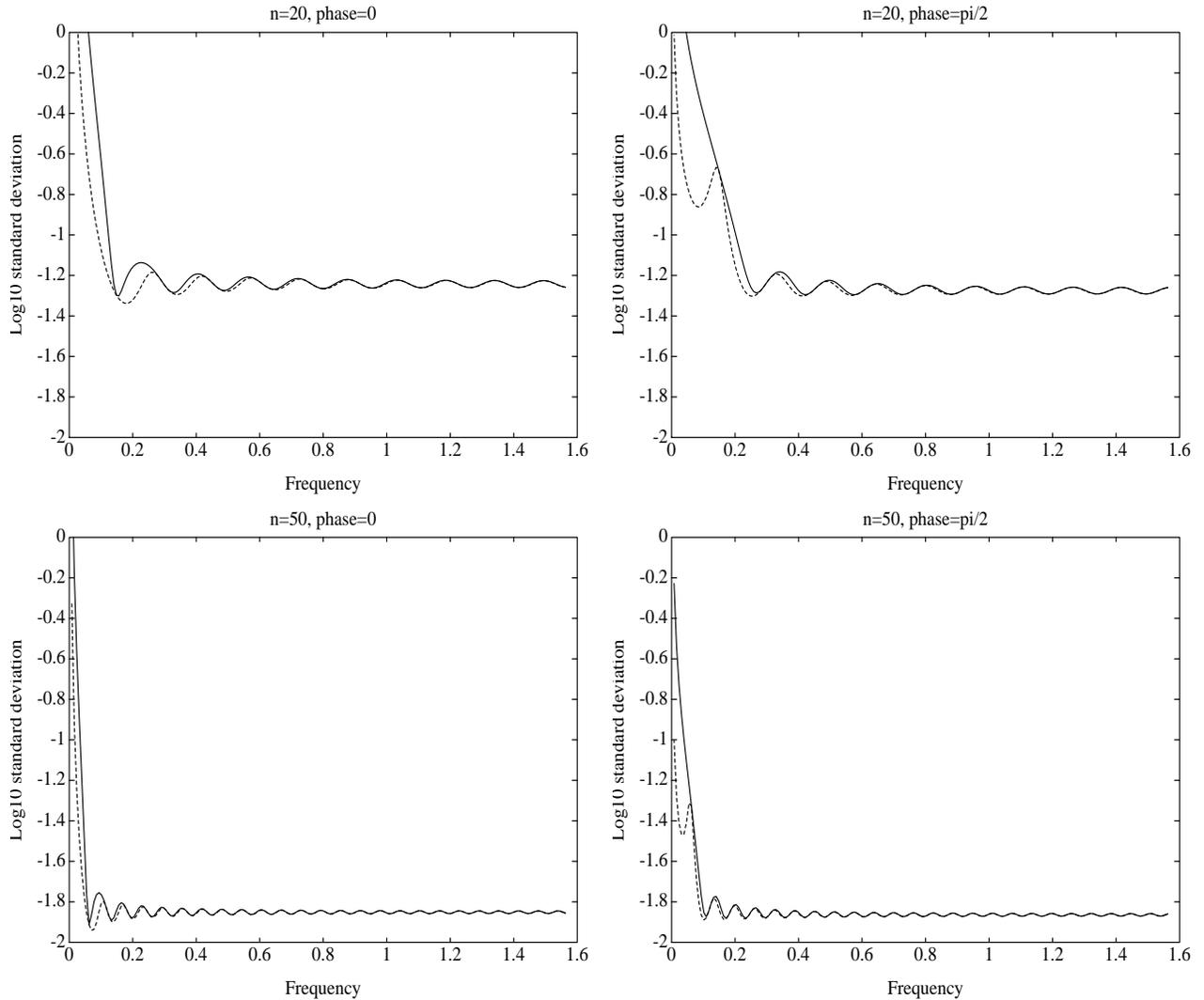


Figure 1: Cramér-Rao lower bounds for the frequency estimator for a single sinusoid signal. Sample sizes of 20 and 50 and phases of zero and $\pi/4$ are used. The dotted line gives bounds for constrained estimators, the solid line for unconstrained estimators. The assumed noise-to-signal-ratio is unity.

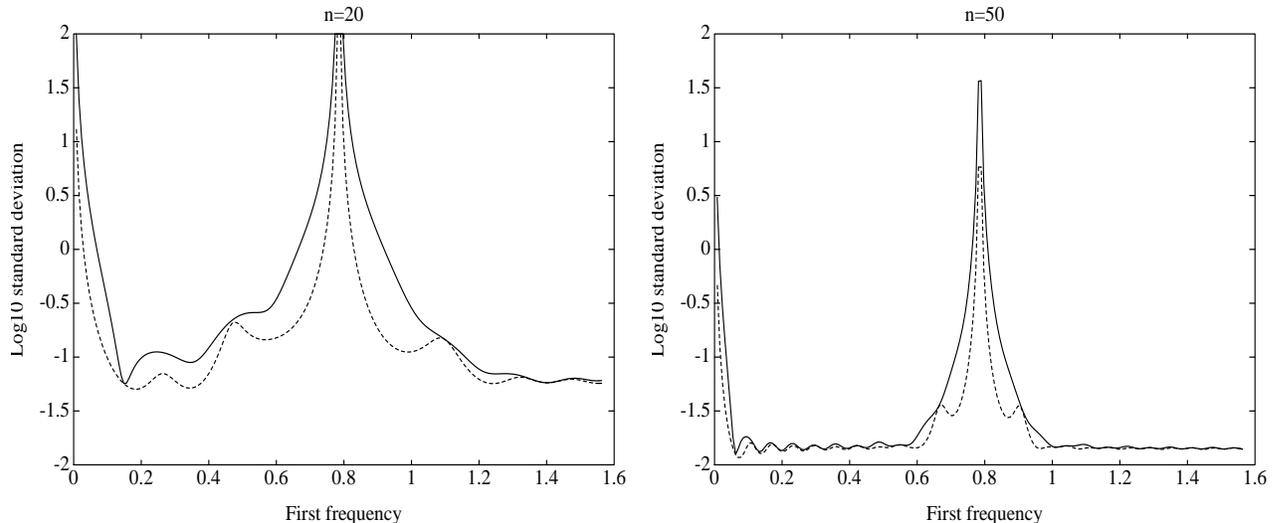


Figure 2: Cramér-Rao lower bounds for the frequency estimator when there is a second frequency at $\pi/4$. Both phases are set to zero and the noise-to-signal-ratio is unity. Sample sizes of 20 and 50 are used. The dotted line gives bounds for constrained estimators, the solid line for unconstrained estimators.

interest in many applications, especially when it is necessary to detect the existence of or a *change* in an input frequency before many new observations are received.

The constrained and unconstrained Cramér-Rao lower bounds for one sinusoid with $n = 20$ and $n = 50$ are displayed in Figure 1. The vertical axis is on a logarithmic-scale, so the difference in height between the curves measures relative variance reduction. The assumed noise-to-signal ratio is unity. It can be seen that the relative gain in precision of the constrained over the unconstrained estimator is arbitrarily large for ω near 0 (and hence also near multiples of π), but decreases as ω approaches $\pi/2$. On the other hand there are intermediate values, at spacings of about π/n , for which the gain is exactly zero. Moving the phase ϕ from zero to $\pi/2$ shifts the zero-gain points slightly, but does not alter the qualitative behavior. Comparing $n = 50$ with $n = 20$, the variances are smaller, the relative gain is lower, and the zero-gain points are more closely spaced, but the qualitative picture is the same.

When there are two or more sinusoids, the relative variance reduction depends on the separations between the frequencies as well as on their proximity to zero. Figure 2 displays constrained and unconstrained Cramér-Rao lower bounds for $n = 20$ and $n = 50$ when there is a second frequency at $\pi/4$. The horizontal axis gives the first frequency, which varies from 0 to $\pi/2$ while the second frequency is held fixed. The frequencies are equally weighted, both phases are zero, and the noise-to-signal ratio is unity. The second frequency introduces an additional point of singularity near which the relative gain in precision of the constrained over the unconstrained estimator becomes arbitrarily large. There are still intermediate points of zero-gain, but these are less apparent than in the single sinusoid case.

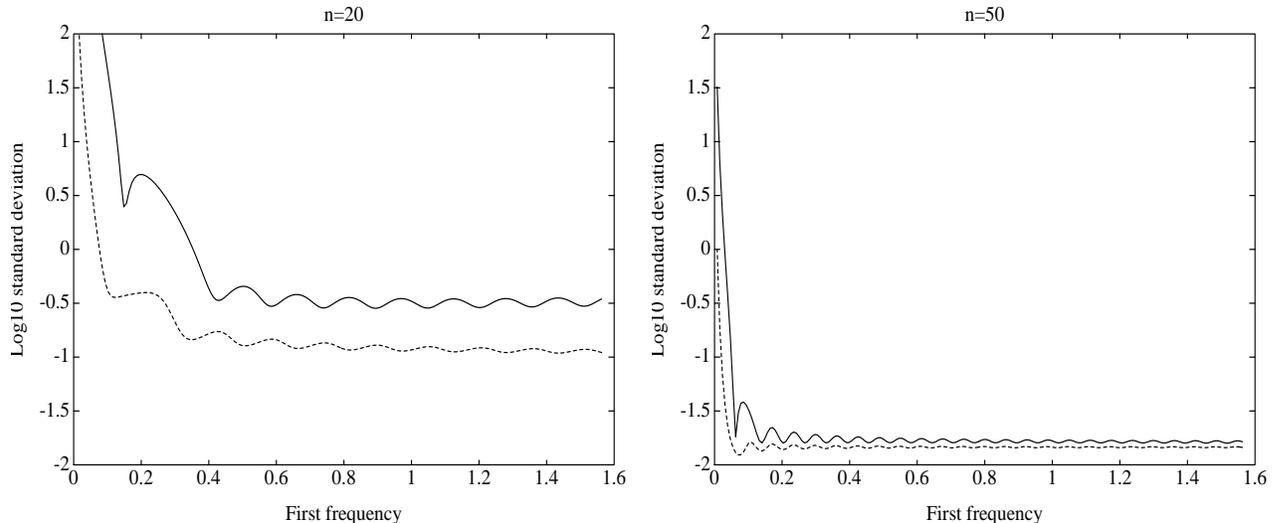


Figure 3: Cramér-Rao lower bounds for the lower frequency estimator when there are two frequencies at a spacing of 0.2. Both phases are set to zero and the noise-to-signal-ratio is unity. Sample sizes of 20 and 50 are used. The dotted line gives bounds for constrained estimators, the solid line for unconstrained estimators.

Figure 3 gives constrained and unconstrained Cramér-Rao lower bounds for two sinusoids when the frequencies are held at a fixed separation of 0.2. The horizontal axis gives the lower of the two frequencies. The relative gain in precision of the constrained over the unconstrained estimator remains substantial throughout the frequency range as well as becoming arbitrarily large as for frequencies near zero or multiples of π .

4.3 Simulations

The above results use Cramér-Rao lower bounds, which are asymptotic expressions. Simulations were carried out to determine how accurate are the asymptotic results for moderate sample sizes. In the case of a single sinusoid, two frequency values were chosen for simulation: $\omega = 0.2$ for which the Cramér-Rao variance bound of the unconstrained estimator is 2.1 times that of the constrained, and $\omega = 0.3$ for which the two variance bounds are equal to three decimal places. 1000 data sets were generated in each case. (More accurate results though were sought for $\omega = 0.2$ with the two highest and two lowest noise to signal ratios. These results are based on 40,000 simulated data sets.) For each simulated data set, a frequency estimate was obtained from Osborne’s least squares algorithm. Starting values for the algorithm were obtained from Pisarenko’s method and the Osborne/Bresler/Macovski algorithm, as described in Section 2. To check that the least squares algorithm was indeed finding the least squares estimator rather than a local minima, the algorithm was also started at the known true frequency. This lead in almost all cases to the same final frequency estimate.

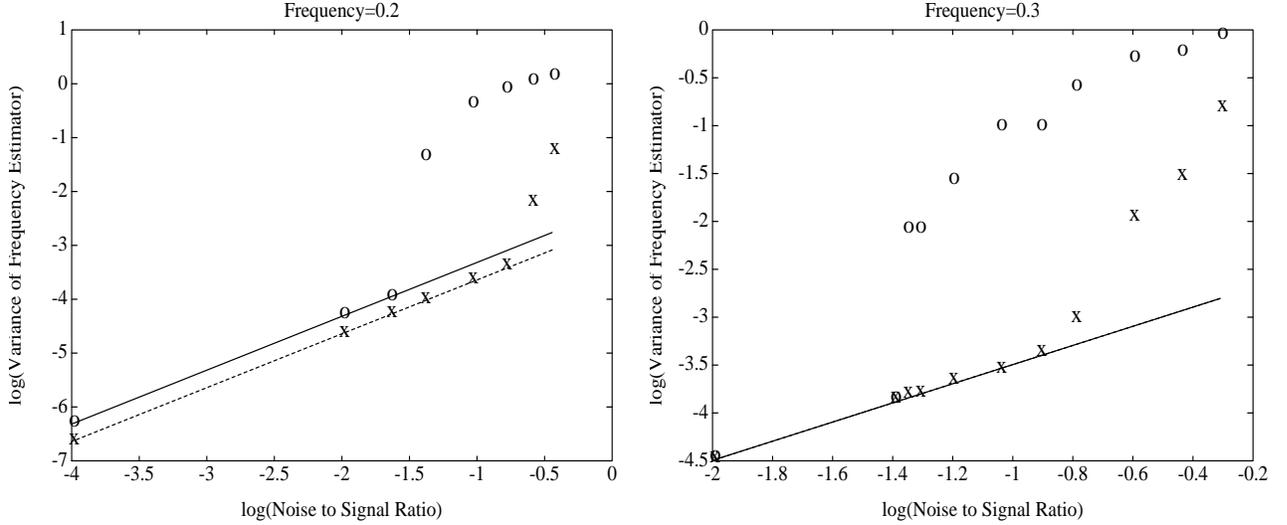


Figure 4: Variances of constrained (crosses) and unconstrained (circles) frequency estimates from the simulations. The true signal is a single sinusoid with frequency 0.2 and 0.3 respectively. The phase is zero and the sample size is 20, The horizontal axis is $\log_{10}(\sigma^2/\rho^2)$. The dotted line gives the Cramér-Rao lower bound for constrained estimators, the solid line for unconstrained estimators.

Figure 4 plots the simulation results for the least squares algorithm. The vertical axis is \log_{10} of the variance of the frequency estimates and the horizontal axis is $\log_{10}(\sigma^2/\rho^2)$, which is a representation of the noise to signal ratio. The Cramér-Rao lower bounds are represented by parallel lines. It can be seen that the variances follow the Cramér-Rao lower bounds very closely for low noise to signal ratios. The breakdown point for σ/ρ above which the actual variance rises above the Cramér-Rao lower bound is about 0.16 and 0.20 for the unconstrained estimator at the two frequencies and about 0.42 and 0.37 for the constrained estimator. Above the breakdown points, the estimation method often produced real exponentials (i.e., frequencies at zero or π) and associated sums of squares much above the theoretical variance bound.

Detailed simulation results are given in Table 1 ($\omega = 0.2$) and in Table 2 ($\omega = 0.3$). The Osborne/Bresler/Macovski algorithm (Osborne/BM) is seen to be at least equivalent in performance to the least squares algorithm (Least Squares). Starting the least squares algorithm from the true values (LS from true) improves on the automatically generated started values only when the symmetry constraint is not used and the noise to signal ratio is at the higher values. As expected, the Pisarenko algorithm is less efficient than either of the Osborne algorithms. The unconstrained Pisarenko algorithm is as good as its constrained version for low noise levels, but quickly breaks down as the noise-to-signal level increases above $\sigma/\rho = 0.1$. The performance of the constrained Pisarenko algorithm degrades very much more slowly as the noise level increases.

The simulations show that the Cramér-Rao lower bounds are an accurate indica-

Table 1: Simulation results one real sinusoid. The true frequency is 0.2, the phase is zero and the sample size is 20. Part (a) of the table gives the means and part (b) the standard deviations of the frequency estimates for various estimation methods. 1000 data sets were generated for each noise level¹. In part (b), Cramér-Rao lower bounds are also given.

(a) Means.

	Noise to Signal Ratio (σ/ρ)							
	0.01	0.1	0.15	0.2	0.3	0.4	0.5	0.6
Constrained								
Pisarenko	0.2000	0.1960	0.1887	0.1853	0.1876	0.1941	0.2065	0.2362
Osborne/BM	0.2000	0.2001	0.2002	0.2002	0.2007	0.2009	0.2031	0.2157
Least Squares	0.2000	0.2000	0.2000	0.1998	0.1997	0.1992	0.2017	0.2229
LS from true	0.2000	0.2000	0.2000	0.1998	0.1997	0.1992	0.2044	0.2302
Unconstrained								
Pisarenko	0.1999	0.1886	0.1741	0.2443	0.6450	1.0181	1.2615	1.4291
Osborne/BM	0.2000	0.1999	0.1998	0.2145	0.3757	0.5539	0.7730	0.9516
Least Squares	0.2000	0.1997	0.1995	0.2139	0.3480	0.4945	0.6408	0.7831
LS from true	0.2000	0.1997	0.1995	0.1991	0.2026	0.2403	0.3226	0.4618

(b) Standard deviations.

	Noise to signal ratio (σ/ρ)							
	0.01	0.1	0.15	0.2	0.3	0.4	0.5	0.6
Constrained								
Pisarenko	0.00318	0.0379	0.0665	0.0904	0.1245	0.1517	0.1918	0.2860
Osborne/BM	0.00048	0.0048	0.0073	0.0099	0.0149	0.0200	0.0574	0.1801
Least Squares	0.00048	0.0048	0.0073	0.0099	0.0149	0.0201	0.0788	0.2364
LS from true	0.00048	0.0048	0.0073	0.0099	0.0149	0.0201	0.1085	0.2658
CRLB	0.00048	0.0048	0.0072	0.0096	0.0143	0.0191	0.0239	0.0287
Unconstrained								
Pisarenko	0.00280	0.0444	0.1550	0.5108	1.1047	1.3311	1.4005	1.4096
Osborne/BM	0.00070	0.0070	0.0104	0.2081	0.6996	0.9612	1.1675	1.2836
Least Squares	0.00070	0.0071	0.0105	0.2082	0.6468	0.8822	1.0430	1.1571
LS from true	0.00070	0.0071	0.0105	0.0146	0.1352	0.3500	0.5874	0.8259
CRLB	0.00069	0.0069	0.0104	0.0139	0.0208	0.0278	0.0347	0.0416

¹ Half of the results, those for $\sigma/\rho = 0.01, 0.1, 0.5$ and 0.6 , are based on 40,000 simulations.

Table 2: Simulation results for one real sinusoid. The true frequency is 0.3, the phase is zero and the sample size is 20. Part (a) of the table gives the means and part (b) the standard deviations of the frequency estimates for various estimation methods. 1000 data sets were generated for each noise level. In part (b), Cramér-Rao lower bounds are also given.

(a) Means.

	Noise to signal ratio (σ/ρ)								
	0.1	0.2	0.25	0.3	0.35	0.4	0.5	0.6	0.7
Constrained									
Pisarenko	0.2992	0.2952	0.2893	0.2808	0.2812	0.2724	0.2641	0.2726	0.3164
Osborne/BM	0.2998	0.2996	0.2986	0.2991	0.2967	0.2979	0.2935	0.2925	0.3157
Least Squares	0.2999	0.2999	0.2992	0.2998	0.2976	0.2992	0.2994	0.3059	0.3610
LS from true	0.2999	0.2999	0.2992	0.2998	0.2976	0.2990	0.3003	0.3094	0.3482
Unconstrained									
Pisarenko	0.2962	0.2830	0.3441	0.4793	0.5598	0.7826	1.1266	1.2320	1.4171
Osborne/BM	0.3000	0.3003	0.3141	0.3548	0.3589	0.4692	0.6491	0.7507	0.9467
Least Squares	0.2999	0.2999	0.3103	0.3385	0.3385	0.4100	0.5303	0.5872	0.7524
LS from true	0.2999	0.2999	0.2992	0.2998	0.2977	0.3023	0.3117	0.3521	0.4196

(b) Standard deviations.

	Noise to signal ratio (σ/ρ)								
	0.1	0.2	0.25	0.3	0.35	0.4	0.5	0.6	0.7
Constrained									
Pisarenko	0.0161	0.0434	0.0654	0.0976	0.1090	0.1339	0.1742	0.2053	0.3408
Osborne/BM	0.0058	0.0117	0.0150	0.0173	0.0211	0.0261	0.0346	0.0439	0.2483
Least Squares	0.0058	0.0117	0.0146	0.0168	0.0206	0.0310	0.1042	0.1712	0.3914
LS from true	0.0058	0.0117	0.0146	0.0168	0.0206	0.0260	0.1055	0.1821	0.3531
CRLB	0.0057	0.0113	0.0141	0.0170	0.0198	0.0226	0.0283	0.0339	0.0396
Unconstrained									
Pisarenko	0.0160	0.1953	0.5103	0.7993	0.9119	1.1321	1.3214	1.3481	1.3684
Osborne/BM	0.0058	0.0116	0.2011	0.3884	0.4082	0.6699	0.9343	1.0367	1.1841
Least Squares	0.0058	0.0117	0.1618	0.3091	0.3089	0.5033	0.7083	0.7640	0.9323
LS from true	0.0058	0.0117	0.0146	0.0167	0.0204	0.0932	0.2072	0.3862	0.5361
CRLB	0.0057	0.0113	0.0141	0.0170	0.0198	0.0226	0.0283	0.0339	0.0396

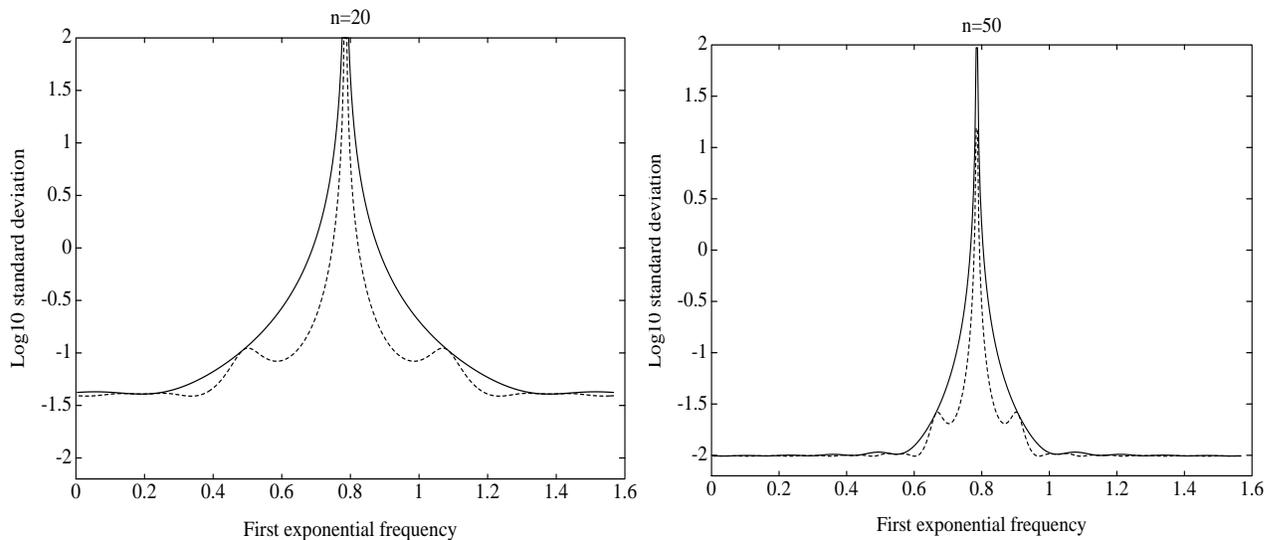


Figure 5: Cramér-Rao lower bounds for the frequency of a complex sinusoid when there is a second frequency at $\pi/4$. Both phases are set to zero and the noise-to-signal-ratio is unity. Sample sizes of 20 and 50 are used. The dotted line gives bounds for constrained estimators, the solid line for unconstrained estimators.

tion of the performance of the constrained and unconstrained least squares frequency estimators for low noise-to-signal ratios, but understate the advantage of the constrained estimators at higher noise levels. It appears that the constrained estimators can tolerate noise-to-signal ratios very roughly twice as high as can the unconstrained estimators. Simulation results for signals made up of two sinusoids were similar to those with one sinusoid, and are not otherwise reported here.

4.4 The Fully Complex Case

In the case of complex signals in complex noise, frequencies near zero or π do not cause a problem, but the situation is otherwise similar to the real case. If there is only one complex exponential, the Cramér-Rao lower bound is constant irrespective of the frequency or symmetry constraint. With two or more frequencies the Cramér-Rao lower bound, and the reduction achieved by the symmetry constraint, depend on how close the frequencies are. Figure 5 plots the constrained and unconstrained variance bounds with two equally weighted frequencies. The second frequency is fixed at $\pi/4$, and the signal-to-noise-ratio is one. This figure shows the same behavior as in the real case, except that there is no singularity at zero.

4.5 Variance of the Eigenvector

It is also of interest to study the variance of the eigenvector estimator $\hat{\mathbf{c}}$. Mathematically, the variance reduction due to the symmetry constraint can be studied more

Table 3: Estimated frequencies for the Inner Diameter data. NA indicates that the roots of the Prony polynomial were not a conjugate pair so that a frequency could not be determined.

	Pisarenko	Osborne/BM	Least Squares	CRLB
Constrained	0.2711	0.3369	0.3379	(0.0225)
Unconstrained	NA	NA	2.9713	(0.0227)

directly for \mathbf{c} than for the frequencies, because the symmetry constraint is linear in $\hat{\mathbf{c}}$. The variance of $\hat{\mathbf{c}}$ is of interest in its own right because the estimated signal μ_t can be calculated directly from \mathbf{c} without rooting the polynomial to find the frequencies (Osborne and Smyth, 1991, Section 3) and because \mathbf{c} defines the transfer function for y_t viewed as an ARMA process.

The Fisher information matrix for \mathbf{c} , adjusted for the α_k and the ϕ_k , is

$$\mathcal{I}_{c|\alpha,\phi} = \frac{1}{\sigma^2} \frac{\partial \boldsymbol{\mu}^*}{\partial \mathbf{c}} (I - P) \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{c}} = \frac{1}{\sigma^2} \frac{\partial \boldsymbol{\mu}^*}{\partial \mathbf{c}} P_X \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{c}}$$

where $P_X = X(X^*X)^{-1}X^*$ is the orthogonal projection onto the column space of X . Now $X^*\boldsymbol{\mu} = 0$, so

$$X^* \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{c}} = -\frac{\partial X^*}{\partial \mathbf{c}} \boldsymbol{\mu} = \mathcal{M}$$

where

$$\mathcal{M} = \begin{pmatrix} \mu_1 & \cdots & \mu_{2p+1} \\ \vdots & & \vdots \\ \mu_{n-2p} & \cdots & \mu_n \end{pmatrix}$$

Hence

$$\mathcal{I}_{c|\alpha,\phi} = \sigma^{-2} \mathcal{M}^* (X^*X)^{-1} \mathcal{M}$$

The Cramér-Rao lower bound for \mathbf{c} is the Moore-Penrose inverse $V_0 = \mathcal{I}_{c|\alpha,\phi}^+$. The Cramér-Rao lower bound for $\boldsymbol{\gamma}_1$ becomes $(Q_1^T \mathcal{I}_{c|\alpha,\phi} Q_1)^+$, so the bound for the symmetrically constrained $\mathbf{c} = Q\boldsymbol{\gamma}_1$ is $V_1 = Q_1(Q_1^T \mathcal{I}_{c|\alpha,\phi} Q_1)^+ Q_1^T$. The matrix V_0 has $2p$ non-zero eigenvalues, and one zero eigenvalue corresponding to the scaling constraint. The effect of the projection reduction here is to put half the positive eigenvalues of V_0 to zero while leaving the others unchanged.

5 Examples

5.1 Inner Diameter of a Landing Gear Triunion

The data give 29 consecutive measurements of the average inner diameter of a landing gear triunion (Spurrier and Thombs, 1990). It is of interest to detect periodic behavior here as it indicates an out of control state which may not be detected by Shewhart charts. The constrained least squares algorithm finds a frequency of 0.337

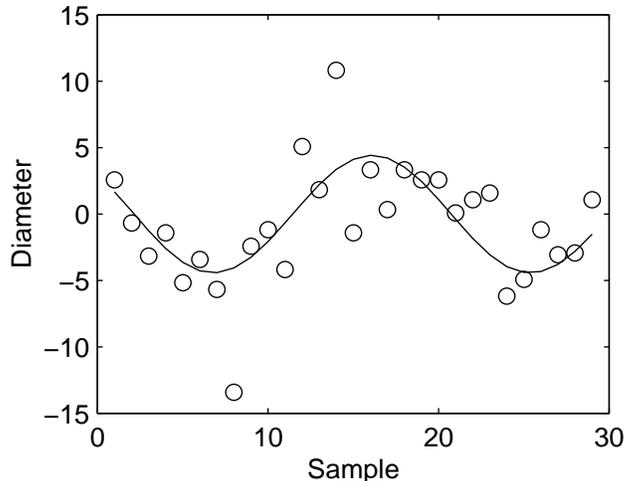


Figure 6: Inner diameter data with one sinusoid.

corresponding to period $2\pi/0.337 = 18.6$ (Figure 6). Although the frequency is highly significant ($P = 0.0008$), the unconstrained algorithms fail on this data, returning the wrong frequency or failing to find frequencies at all (Table 3). Table 3 also gives the Cramér-Rao lower bounds for the standard deviations of the frequency estimators.

5.2 Forces on a Cylinder

Newton (1988) studied waves emanating from a cylinder suspended in a tank of water. The data give the relative vertical displacement at equally spaced times. The waves were believed to be subject to high frequency vibration, as an artifact of the experiment equipment, as well as lower frequency vibration which reflects forces acting on the cylinder. It is of interest to identify and to filter out the high frequency vibration.

Use of the constrained least squares algorithm finds six significant frequencies (Figure 7). The high frequency oscillation is found to be at 2.34 with a possible harmonic at 1.23. The slower oscillations are represented by a range of frequencies from 0.24 to 0.81, which may represent a continuous peak in the spectra around these values (Table 4). The Cramér-Rao lower bounds for the standard deviations, computed using the constrained least squares estimates, suggest that the two high frequencies are equally well estimated with or without the symmetry constraint while the four slower frequencies are much less well determined when the constraint is not used. In the practice, estimation using the constraint proves to be more stable. If the number of frequencies is reduced from six to five, then the frequencies returned by the constrained algorithms are, to sampling accuracy, a subset of the six given here, while those returned by the unconstrained algorithms are different. All of the unconstrained algorithms returned frequencies with damping factors which were ignored when extracting the frequencies.

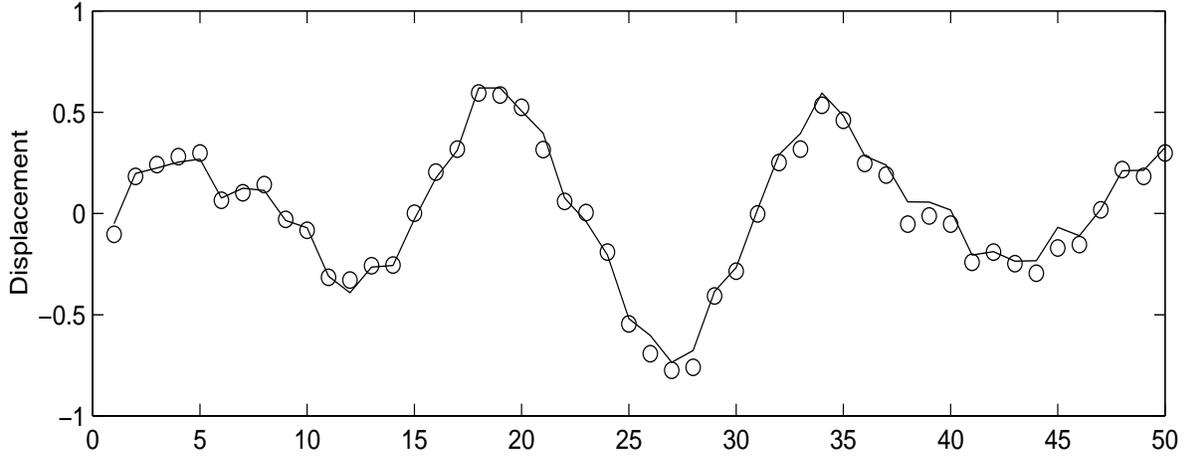


Figure 7: Vertical displacements of waves emanating from a cylinder suspended in water. The estimated signal consists of six sinusoidal frequencies.

Table 4: Estimated frequencies of waves emanating from a cylinder suspended in water.

	Frequencies					
	1	2	3	4	5	6
Constrained						
Pisarenko	0.3305	0.4332	0.9611	1.4347	2.0939	2.3696
Osborne/BM	0.2148	0.3906	0.5184	0.7881	1.2063	2.3237
Least Squares	0.2381	0.4045	0.5475	0.8142	1.2341	2.3440
CRLB	(0.0086)	(0.0051)	(0.0144)	(0.0149)	(0.0192)	(0.0133)
Unconstrained						
Pisarenko	0.3268	0.4373	0.9576	1.4438	2.0924	2.3705
Osborne/BM	0.1737	0.3764	0.5015	0.8264	1.2166	2.3330
Least Squares	0.1681	0.3722	0.4982	0.8128	1.2299	2.3429
CRLB	(0.0335)	(0.0227)	(0.0532)	(0.0248)	(0.0195)	(0.0136)

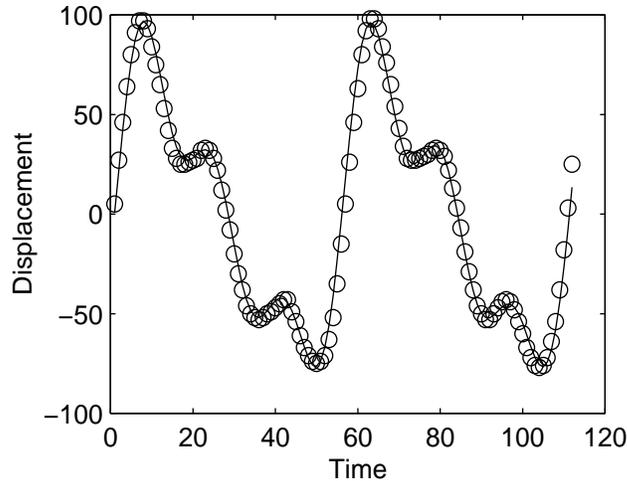


Figure 8: Voice data from singing the vowel 'ooh'. The estimated signal has three frequencies.

5.3 Voice Data from Singing the Vowel 'ooh'

The data give the normalized magnitudes of voice data when the vowel 'ooh' was sung at a pitch of 290 Hz. A sampler/synthesizer was used to capture and to store the data (Oliver, 1997). The frequencies and amplitudes found in the signal are used to identify the phonetic vowel and are of interest in voice synthesis, therapy and training. In this case the data are very regular and are well modeled by a frequency and two harmonics (Figure 8). The symmetry constraint turns out to be of only marginal importance here. The constrained and unconstrained algorithms give similar results, even though the unconstrained algorithms returned damping factors which were ignored when extracting the frequencies (Table 5). Even for this very regular data, Pisarenko's method is less efficient than the iterative algorithms and fails to correctly identify the third frequency.

6 Summary

It has been shown that the symmetry constraint on the eigenvector \mathbf{c} is a natural modification to eigenanalysis based algorithms for estimating frequencies. In most cases this is equivalent to constraining the signal to be periodic.

The constraint has been shown to substantially improve the precision of the frequency estimates when the frequencies are difficult to resolve, that is when the noise to signal ratio is high, when the frequencies are nearly equal, or when the frequencies are close to zero. In other cases the gain from imposing the constraint depends on the actual values of the frequencies and the sample size. The gain is generally much less evident for larger sample sizes.

Table 5: Estimated frequencies for the voice data.

	Frequencies		
	1	2	3
Constrained			
Pisarenko	0.1361	0.3827	1.8136
Osborne/BM	0.1134	0.2299	0.3402
Least Squares	0.1134	0.2297	0.3408
CRLB	(0.0003231)	(0.0009395)	(0.0008023)
Unconstrained			
Pisarenko	0.1358	0.3832	1.8147
Osborne/BM	0.1134	0.2298	0.3401
Least Squares	0.1134	0.2299	0.3400
CRLB	(0.0003234)	(0.0009400)	(0.0008033)

Perhaps the most significant result is that the constraint extends the usefulness of all the eigenanalysis methods, most notably Pisarenko’s method, into higher noise situations. It increases the breakdown noise to signal ratio above which the methods fail to give useful results. This result is particularly important if Pisarenko’s method is used to find starting values with which to initialize the other frequency estimation algorithms. In that case the constraint roughly doubles the noise to signal ratio at which the methods can be used.

A three stage algorithm, in which Pisarenko’s method is used to start the Osborne/Bresler/Macovski algorithm, which in turn is used to start the least squares algorithm, proved to be a very reliable method of computing the least squares estimators without the need to provide starting values.

Frequency estimation software which implements the symmetry constraint can be obtained from <http://www.statsci.org/other/prony.html>. The software is available for the S-Plus and Matlab programming environments.

References

- Benade, A. (1990). *Fundamentals of Musical Acoustics, 2nd ed.*, Dover Publications, New York.
- Bresler, Y. and Macovski, A. (1986). Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34, 1081–1089.
- Brockwell, P. J., and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.
- Chan, Y. T., Lavoie, J. M. M., and Plant, J. B. (1981). A parameter estimation approach to estimation of frequencies of sinusoids. *IEEE Trans. Acoust., Speech*

- and Sig. Proc.*, ASSP-**29**, 214–219.
- Evans, A. G., and Fischl, R. (1973). Optimal least squares time-domain synthesis of recursive digital filters. *IEEE Trans. Audio Electroacoust.*, AU-**21**, 61–65.
- Hannan, E. J. (1973). The estimation of frequency. *J. Appl. Prob.*, **10**, 510–519.
- Hasan, T. (1982). Nonlinear time series regression for a class of amplitude modulated cosinusoids. *J. Time Series Anal.*, **3**, 109–122.
- James, B., Anderson, B. D. O., and Williamson, R. C. (1994). Conditional mean and maximum likelihood approaches to multiharmonic frequency estimation. *IEEE Transactions on Signal Processing*, **42**, 1366–1375.
- Kahn, M., Mackisack, M.S., Osborne, M.R. and Smyth, G.K. (1992). On the consistency of Prony’s method and related algorithms. *J. Comput. Graph. Statist.*, **1**, 329–349.
- Kannan, N., and Kundu, D. (1994). On modified EVLP and ML methods for estimating superimposed exponential signals. *Signal Processing*, **39**, 223–233.
- Kay, S. M. (1984). Accurate frequency estimation at low signal-to-noise ratio. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-**31**, 540–547.
- Kay, S.M. (1988). *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs.
- Kay, S.M. and Marple, S.L. (1981). Spectrum analysis—a modern perspective. *Proc. IEEE*, **11**, 1380–1419.
- Kumaresan, R., and Shaw, A. K. (1985). High resolution bearing estimation without eigendecomposition. *Proc. IEEE ICASSP 85*, Tampa, FL, 576–579.
- Kumaresan, R., Scharf, L. L., and Shaw, A. K. (1986). An algorithm for pole-zero modeling and spectral analysis. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-**34**, 637–640.
- Kundu, D. (1993). Estimating the parameters of undamped exponential signals. *Technometrics*, **35**, 215–218.
- Kundu, D., and Kannan, N. (1997). Constrained maximum likelihood estimator for superimposed exponential signals. *Comm. Statist. Simulation Comput.*, **26**, 733–764.
- Mackisack, M. S., Osborne, M. R., and Smyth, G. K. (1994). A modified Prony algorithm for estimating sinusoidal frequencies. *J. Statist. Comput. Simul.*, **49**, 111–124.
- Marple, S.L. (1980). A new autoregressive spectrum analysis algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-**28**, 441–454.
- Marple, S.L. (1987). *Digital Spectral Analysis*, Prentice-Hall, Englewood Cliffs.

- Matausek, M. R., Stankovic, S. S., and Radivic, D. V. (1983). Iterative inverse filtering approach to the estimation of frequencies of noisy sinusoids. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-**31**, 1456–1463.
- Nehorai, A. (1985). A minimal parameter adaptive notch filter with constrained poles and zeros. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-**33**, 983–996.
- Newton, H. J. (1988). *TIMESLAB: A Time Series Analysis Laboratory*. Wadsworth, Pacific Grove, California.
- Oliver, W. D. (1997). The Singing Tree: A Novel Interactive Musical Interface. Master of Science Thesis, Massachusetts Institute of Technology.
- Osborne, M.R. (1970). A class of nonlinear regression problems. In: *Data Representation*, R.S. Anderssen and M.R. Osborne, eds., University of Queensland Press, pp. 94–101.
- Osborne, M.R. (1975). Some special nonlinear least squares problems. *SIAM J. Numer. Anal.*, **12**, 571–592.
- Osborne, M. R. and Smyth, G. K. (1991). A modified Prony algorithm for fitting functions defined by difference equations, *SIAM J. Sci. Stat. Comput.*, **12**, 362–382.
- Osborne, M. R., and Smyth, G. K. (1995). A modified Prony algorithm for exponential function fitting. *SIAM J. Sci. Statist. Comput.* **16**, 119–138.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.
- Prony, R. (1795). Essai expérimental et analytique. *J. de L'École Polytechnique*, **2**, 24–76.
- Reddy, D. R. (1967). Computer Recognition of Connected Speech. *Journal of the Acoustical Society of America*, **42**, 329–347.
- Rice, J. A., and Rosenblatt, M. (1988). On frequency estimation. *Biometrika*, **75**, 477–484.
- Sakai, H. (1984). Statistical analysis of Pisarenko's method for sinusoidal frequency estimation. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-**32**, 95–101.
- Schafer, R., and Rabiner, L. (1969). System for Automatic Formant Analysis of Voiced Speech. *Journal of the Acoustical Society of America*, **47**, 634–648.
- Smyth, G. K., and Hawkins, D. M. (1997). Robust Frequency Estimation Using Elemental Sets. *Computing Science and Statistics*, **28**, 659–662.
- Spurrier, J. D., and Thombs, L. A. (1990). Control charts for detecting cyclical behavior. *Technometrics* **32**, 163–170.
- Stoica, P. and Nehorai, A. (1988). Study of the statistical performance of the Pisarenko harmonic decomposition method. *Communications, Radar and Signal Processing*, **153**, 161–168.

- Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika*, **58**, 21–36.
- Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Almqvist and Wiksells Boktryckeri, Uppsala.
- Whittle, P. (1953). The simultaneous estimation of a time series' harmonic components and covariance structure. *Trab. Estad.*, **3**, 43–57.