UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Analytical Bounds on Least Squares and Total Least Squares Methods for the Linear Prediction Problem

Ricardo D. Fierro
Kung Yao

May 1993

CAM Report 93-10

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

ANALYTICAL BOUNDS ON LEAST SQUARES AND TOTAL LEAST SQUARES METHODS FOR THE LINEAR PREDICTION PROBLEM

Ricardo D. Fierro
Dept. of Mathematics
University of California, Los Angeles
Los Angeles, CA 90024-1555

Kung Yao
Electrical Engineering Dept.
University of California, Los Angeles
Los Angeles, CA 90024-1594

Abstract. The least squares (LS) and total least squares (TLS) methods are commonly used to solve the linear prediction (LP) equations in frequency estimation problems. We examine how increasing the prediction order or the number of observations may reduce the sensitivity of the linear prediction coefficients to noise. We provide a new upper bound for the difference between the LS and TLS solutions, which explains their similarities/differences in a high/low SNR environment. Our analysis is intimately linked to the concept of the subspace angle.

1. Introduction. In this paper we consider the problem of resolving closely spaced (in frequencies) sinusoids in a noisy environment using the linear prediction (LP) model and the root form. The least squares (LS) and total least squares (TLS) methods are commonly used to provide a solution to an overdetermined system of linear equations, which then determine the coefficients of a polynomial whose roots must be calculated. It is well known that the estimated frequencies are then determined by the angular positions of the roots closest to the unit cicle.

Asymptotic statistical analysis and computer simulations (e.g., see [14], [10], [8], [11], [13], [9]) show that increasing the prediction order (number of columns) or the number of observations (number of equations) may improve the LS or TLS frequency estimates significantly. The accuracy of the frequency estimates by the root method is inherently related to the accuracy of the computed polynomial coefficients [9], [12], [13].

In this paper we shall show from a numerical linear algebra point of view how increasing the number of observations or the prediction order in the LP model *reduces the sensitivity* of the computed polynomial coefficients to noise in the data. Fierro and Bunch [2] showed how subspace sensitivity translates to solution sensitivity in orthogonal projection methods such as LS and TLS. Using this result as a platform, our analysis will focus on the accuracy of the polynomial coefficients in a noisy environment by examining the sensitivity of certain subspaces. This approach provides a new perspective on the role of the dimensions of the LP problem. Our numerical simulations verify the correlation between decreased subspace sensitivity and improved frequency estimates. This correlation helps explain why increasing the dimensions of the LP model can lead to improved frequency estimates, and especially provides new insight on the role of the prediction order. In addition, we provide a new upper bound for the difference between the LS and TLS solutions, which explains their similarities/differences in a high/low SNR environment. Our analysis is intimately linked to the concept of the subspace angle.

2. LP Models and Notation. For purposes of illustration, we consider the real case of the forward linear prediction (FLP) and forward-backward linear prediction (FBLP) models. The noiseless FLP model is denoted by

$$A_0\, g = b_0$$

where it is assumed A_0 is $(N-L)\times L$, g is $L\times 1$, b_0 is $(N-L)\times 1$, p is the number of complex exponentials, $(N-L) > L > p$, and $\text{rank}(A_0) = p$. Due to the presence of noise, however, one has to estimate the FLP coefficients $\{g_i\}$ from the perturbed problem

$$(1) \qquad A\,g \approx b,$$

where $A = A_0 + \delta A_0$, $b = b_0 + \delta b_0$, and $\text{rank}(A) \geq p$. The singular value decomposition SVD (see [4]) is a powerful analytical and computational tool in numerical linear algebra and engineering applications. Let A_0 and $[A_0\; b_0]$ have the dyadic SVD's $A_0 = \sum_{i=1}^{p} \sigma_{0_i} u_{0_i} v_{0_i}^T$ and $[A_0\; b_0] = \sum_{i=1}^{p} \bar{\sigma}_{0_i} \bar{u}_{0_i} \bar{v}_{0_i}{}^H$. Let A and $[A\; b]$ have the dyadic SVD's $A = \sum_{i=1}^{L} \sigma_i u_i v_i{}^H$ and $[A\; b] = \sum_{i=1}^{L+1} \bar{\sigma}_i \bar{u}_i \bar{v}_i^H$.

The FBLP model was formulated by Nuttall [7] and Ulrych and Clayton [15] as a means to resolve closely spaced sinusoids with a limited data sample size (provided the SNR is sufficiently high). The noiseless FBLP model is denoted

$$(2) \qquad C_0\, h = d_0$$

where it is assumed C_0 is $2(N-L) \times L$, h is $L \times 1$, d_0 is $2(N-L) \times 1$, p is the number of complex exponentials, $2(N-L) > L > p$, and $\text{rank}(C_0) = p$. As in the FLP case, one has to estimate the FBLP coefficients $\{h_i\}$ from a perturbed problem, denoted by

$$(3) \qquad C\,h \approx d,$$

where $C = C_0 + \delta C_0$, $d = d_0 + \delta d_0$, and $\text{rank}(C) \geq p$. Let C_0 and $[C_0\; d_0]$ have the dyadic SVD's $C_0 = \sum_{i=1}^{p} \alpha_{0_i} w_{0_i} y_{0_i}{}^H$ and $[C_0\; d_0] = \sum_{i=1}^{p} \bar{\alpha}_{0_i} \bar{w}_{0_i} \bar{y}_{0_i}^H$. Let C and $[C\; d]$ have the dyadic SVD's $C = \sum_{i=1}^{L} \alpha_i w_i y_i{}^H$ and $[C\; d] = \sum_{i=1}^{L+1} \bar{\alpha}_i \bar{w}_i \bar{y}_i^H$.

The rank of the coefficient matrices in the noise-free FLP and FLPB models is p. The noise usually increases the rank and the ordinary LS and TLS solutions can be unstable due to ill-conditioning. The stabilizing remedy is to restore the rank, i.e., replace the noisy model with a nearby model of the correct rank. For SVD-based methods such as truncated LS and TLS, this is accomplished by zeroing the singular values in the matrix that are below a specified tolerance. (There are other alternatives to the SVD for obtaining a nearby model of reduced rank, such as two-sided orthogonal decompositions (*) or rank revealing QR factorizations (*).)

Provided $\sigma_p > \bar{\sigma}_{p+1}$, the truncated TLS approach to (1) requires that we solve the compatible system

$$(4) \qquad \hat{A}\,g = \hat{b}$$

where $[\hat{A}\; \hat{b}] = \sum_{i=1}^{p} \bar{\sigma}_i \bar{u}_i \bar{v}_i^H$ is the rank p matrix nearest to $[A\; b]$ in the 2-norm. Let g_{TLS} denote the minimum norm TLS solution to (4). Provided $\sigma_p > \sigma_{p+1}$, the LS approach requires that we solve the truncated LS problem

$$(5) \qquad A_p\, g \approx b$$

where $A_p = \sum_{i=1}^{p} \sigma_i u_i v_i{}^H$ is the rank p matrix nearest to A in the 2-norm. Let g_{LS} denote the minimum norm LS solution to (5). If b_p denotes the orthogonal projection of b onto $R(A_p)$, then $A_p g_{LS} = b_p$ is compatible and

$$A_p\, g_{LS} = b_p \iff [A_p\; b_p] \begin{bmatrix} g_{LS} \\ -1 \end{bmatrix} = 0.$$

In this formulation it is shown in [2] that g_{LS} can be determined from an orthonormal basis for the nullspace of $[A_p\; b_p]$, denoted $\text{N}([A_p\; b_p])$.

Similarly, let h_{TLS} denote the minimum norm TLS solution to the FBLP system in (3), i.e., it is the minimum norm solution to

$$(6) \qquad \hat{C}\,h = \hat{d}$$

where $[\hat{C}\ \hat{d}] = \sum_{i=1}^{p} \bar{\alpha}_i \bar{w}_i \bar{y}_i^H$. Finally, let h_{LS} denote the minimum norm truncated LS solution to the FBLP system in (3), i.e., it solves the related LS problem

$$C_p\, h \approx d \tag{7}$$

where $C_p = \sum_{i=1}^{p} \alpha_i w_i {y_i}^H$.

3. Increasing the Number of Observations. It has been empirically observed by many that increasing the number of equations in the LP model may result in more accurate frequency estimates. For fixed L, this is easily achieved by increasing the measured data sample size. Recently, Stoica and Soderstrom (*) derived a first-order expression for unbiased estimators which relate the large-sample covariance matrix of the frequencies, denoted $C_{f,N}$, to the large-sample covariance matrix of the LP coefficients. They showed that for $t \geq N$, then $C_{f,t} \leq C_{f,N}$.

In this section we will show that increasing the number of equations or the data sample size correspondingly results in more accurate estimates of the LP coefficients, i.e., polynomial coefficients. Our analysis does not require the large-sample assumption. We merely need a compatible system of linear equations $\tilde{A}g = \tilde{b}$ with an important property: the coefficient matrix restores the correct rank in an acute manner, i.e., $\text{rank}(\tilde{A}) = p$, $\|A_0^\dagger A_0 - \tilde{A}^\dagger \tilde{A}\| < 1$ and $\|A_0 A_0^\dagger - \tilde{A}\tilde{A}^\dagger\| < 1$ where $\dagger$ denotes the pseudoinverse of a matrix (see (*) for definition of acutenes). In the root form of the LP approach to frequency estimation the accuracy of the polynomial coefficients is paramount because once they are supplied, the frequencies are automatically determined from the angular positions of the roots closest to the unit circle.

In [2] Fierro and Bunch showed that for orthogonal projection methods such as LS or TLS, the solution is completely determined from an approximate nullspace of $[A\ b]$ or an exact nullspace of a compatible rank p matrix approximation of $[A\ b]$, and lower and upper perturbation bounds are derived for orthogonal projection methods. The perturbation bounds are in terms of subspace angles and provide insight into problems that might be otherwise difficult to obtain. As an example, it has been observed by many in various applications that TLS often provides more accurate linear parameter estimates than LS. The analysis is [2] shows that the TLS subspace angle is generally smaller than the LS subspaces angle and justifies the improved performance of TLS over LS in a noisy environment. As we will see below, these perturbation bounds in terms of subspace angles provide theoretical insight into the role of the dimensions of the LP problem.

Let A_0^t and b_0^t denote the FLP coefficient matrix and right hand side vector, repectively, with t rows such that $t \geq (N-L)$. In Matlab notation, $A_0 = A_0^t(1 : N-L, :)$ and $b_0 = b_0^t(1 : N-L)$. Denote by $\sigma_{0_p}^t$ the p^{th} singular value of A_0^t; note that $\sigma_{0_p} \leq \sigma_{0_p}^t$. Further, if $A^t = A_0^t + \delta A_0^t$ and $b^t = b_0^t + \delta b_0^t$ denote the perturbed version of A_0^t and b_0^t, respectively, let g_{LS}^t denote the truncated LS solution to $A_p^t g \approx b^t$, or equivalently, the minimum norm solution to the compatible system $A_p^t g = b_p^t$. The following application of a result by Fierro and Bunch [2, Theorem 5] shows how much the LS estimates g_{LS} and g_{LS}^t can differ from g, the true FLP coefficients.

THEOREM 3.1. *Let g_{LS} denote the truncated LS solution to (1), and let g_{LS}^t denote the minimum norm solution to solution to $A_p^t g = b_p^t$ as described above, with $t \geq N-L$. Let $\eta_{LS} = \sqrt{1+\|g_{LS}\|^2}\sqrt{1+\|g\|^2}$ and $\beta_{LS} = \sqrt{1+\|g_{LS}^t\|^2}\sqrt{1+\|g\|^2}$. If $\|\delta A_0^t\| < \sigma_{0_p}$, then*

$$(i)\ \sin\theta_{FLS} \;\leq\; \|g_{LS} - g\| \;\leq\; \eta_{LS}\sin\theta_{FLS} \;\leq\; \eta_{LS}\,\frac{\|[\delta A_0\ \delta b_0]\|}{\sigma_{0_p} - \|\delta A_0\|}$$

$$(ii)\ \sin\theta_{FLS}^t \;\leq\; \|g_{LS}^t - g\| \;\leq\; \beta_{LS}\sin\theta_{FLS}^t \;\leq\; \beta_{LS}\,\frac{\|[\delta A_0^t\ \delta b_0^t]\|}{\sigma_{0_p}^t - \|\delta A_0^t\|}$$

where θ_{FLS} is the subspace angle between $N([A_p\ b_p])$ and $N([A_0\ b_0])$, and θ_{FLS}^t is the subspace angle between $N([A_p^t\ b_p^t])$ and $N([A_0^t\ b_0^t])$.

Corresponding results hold for the TLS solutions to these equations. This result has several important implications. First, it shows how the solution accuracy is directly related to the subspace angle. The lower bound means *sensitivity of the subspace* to noise translates to *sensitivity of the solutions.* Further, once the subspace is perturbed, i.e., $0 < \sin\theta_M$, for M =LS or TLS, then the LS or TLS solution *cannot* coincide with the true solution g and one cannot extract any more accuracy from the data than the subspace will allow under acute perturbations. Second, provided the problems are not too ill-conditioned in the sense that $\sqrt{1+\|\cdot\|^2}$ is not too large, a small noise level $\|[\delta A_0^t \;\; \delta b_0^t]\|$ will produce a correspondingly small error in the estimates of the true coefficients. As the noise level grows, we expect corresponding larger errors in the LS or TLS estimates of g. Finally, since usually $\|\delta A_0\| \approx \|[\delta A_0^t \;\; \delta b_0^t]\|$ and $\|g_{LS}\| \approx \|g_{LS}^t\|$, the upper bounds reveal the potential role of the singular values in comparing the accuracy of the methods in estimating the polynomial coefficients. We shall now explore this issue.

It is easy to verify (e.g., see [4, p.231]) that $\sigma_{0p} \le \sigma_{0_p}^t$. By the nesting property of the submatrices, we have

$$\sigma_{0p} = \sigma_{0_p}^{N-L} \le \sigma_{0_p}^{N-L+1} \le \cdots \le \sigma_{0_p}^t.$$

Increasing t lengthens the string of inequalities, and for larger values of t we expect $\sin\theta_{FLS}^t$ to be smaller than $\sin\theta_{FLS}$. In other words, we expect the solution g_{LS}^t to be more accurate (or less sensitive to noise) than g_{LS}.

Now, we examine the consequences of $\sigma_{0p} = \sigma_{0_p}^t$, and we shall see that the circumstance under which this equality holds may be rare in practice. We assume $\sigma_{0_p}^t$ is an isolated singular value. Partition A_0^t by

$$A_0^t = \begin{bmatrix} A_0 \\ A_* \end{bmatrix} \begin{matrix} N-L \\ t-(N-L). \end{matrix}$$

Let A_0^t have the dyadic SVD $A_0^t = \sum_{i=1}^p \sigma_{0_i}^t u_{0_i}^t v_{0_i}^t$. Let $u_{0_p}^t$, the p^{th} left singular of A_0^t, have the partition

$$u_{0_p}^t = \begin{bmatrix} u^{t1} \\ u^{t2} \end{bmatrix} \begin{matrix} N-L \\ t-(N-L). \end{matrix}$$

Then from the eigenequations $A_0^t A_0^{t\,H} u_{0_p}^t = {\sigma^t}_{0_p}^2 u_{0_p}^t$ we get

$$A_0 A_0^H u^{t1} + A_0 A_*^H u^{t2} \;\; = \;\; {\sigma^t}_{0_p}^2 u^{t1}. \tag{8}$$

Premultiplying both sides by $u_{0_p}^H$, where u_{0_p} is the p^{th} left singular vector of A_0, yields

$$u_{0_p}^H A_0 A_0^H u^{t1} + u_{0_p}^H A_0 A_*^H u^{t2} = {\sigma^t}_{0_p}^2 u_{0_p}^H u^{t1}, \tag{9}$$

and under the assumption $\sigma_{0p} = \sigma_{0_p}^t$ the equation simplifies to

$$u_{0_p}^H A_0 A_*^H u^{t2} = 0, \quad \text{or} \quad v_{0_p}^H A_*^H u^{t2} = 0. \tag{10}$$

There are three possible cases:

Case 1. $A_*^H u^{t2} = 0$
Case 2. $u^{t2} = 0$
Case 3. $v_{0_p}^H A_*^H u^{t2} = 0$.

It can be shown that each case implies $(\sigma_{0_p}^t,\, u^{t1},\, v_{0_p}^t)$ is a singular triplet of A_0. In other words, $\sigma_{0_p} = \sigma_{0_p}^t$ implies each row of A_* is orthogonal to the right singular vector $v_{0_p}^t$ of A_0.

The previous analysis is also useful if we consider the matrix A_0^t to actually represent the FBLP matrix C_0, since $A_0 = C_0(1 : N-L, :)$. Thus we immediately get the following result.

COROLLARY 3.2. *Let A_0 and C_0 represent the unperturbed FLP and FBLP coefficient matrices in Section 2, and let the noisy LP problems in (1) and (3) have the truncated LS solutions g_{LS} and h_{LS}. Let $\eta_{LS} = \sqrt{1+\|g_{LS}\|^2}\sqrt{1+\|g\|^2}$ and $\beta_{LS} = \sqrt{1+\|h_{LS}\|^2}\sqrt{1+\|h\|^2}$. If $\|\delta C_0\| < \sigma_{0_p}$ then*

(i) $\sin\theta_{FLS} \;\le\; \|g_{LS}-g\| \;\le\; \eta_{LS}\sin\theta_{FLS} \;\le\; \eta_{LS}\frac{\|[\delta A_0\ \delta b_0]\|}{\sigma_{0_p}-\|\delta A_0\|}$

(ii) $\sin\theta_{FBLS} \;\le\; \|h_{LS}-h\| \;\le\; \beta_{LS}\sin\theta_{FBLS}^t \;\le\; \beta_{LS}\frac{\|[\delta C_0\ \delta d_0]\|}{\alpha_{0_p}-\|\delta C_0\|}$

where θ_{FLS} is the subspace angle between $N([A_p\ b_p])$ and $N([A_0\ b_0])$, and θ_{FBLS} is the subspace angle between $N([C_p\ d_p])$ and $N([C_0\ d_0])$.

If the noise level is not too large such that $\|[\delta A_0\ \delta b_0]\| \approx \|[\delta C_0\ \delta d_0]\|$ then we see that the improved performance of the FBLP model can be explained in terms of reduced subspace sensitivity to noise, and we expect a significant improvement over the FLP model whenever $\sigma_{0p} \ll \alpha_{0_p}$. Our numerical simulations show that reduced subspace sensitivity is indeed correlated with improved frequency estimates.

4. Increasing the Prediction Order. In the previous section we showed how a larger sample size,N, translates to reduced sensitivity of the LP coefficients to noise when the prediction order, L, is fixed. This results in improved frequency estimation. Now we consider the case where N is fixed and L is allowed to vary. This problem is more delicate because both the row and column dimensions of the coefficient matrix are changing, as well as the right hand vector, and we can no longer rely on property that one problem is nicely embedded in the other. To date, the dilemma of choosing an optimal prediction order for given noise variance, frequency locations, and data sample size has not been solved.

The numerical simulations in [14] by Tufts and Kumaresan illustrate that increasing the prediction order may result in improved frequency estimates. However, their simulations (as well as those in [6, 15]) also showed that this phenomenon only holds to a certain degree, because beyond the problem-dependent point the resolution capability deteriorates. This results in a diminished effective SNR. Some typically recommended values of L are $N/3$, $N/2$, $2N/3$, and $3N/4$ (e.g., see [6, 9, 14, 15]), which suggests that despite the vast amount of heuristic arguments surrounding this dilemma a scheme for choosing the prediction order for fixed N remains elusive. However, the vast computational evidence demands two main properties of a purported diagonostic for this dilemma:

(a) Increasing L improves LP parameter and frequency estimation.
(b) Beyond a certain value L^* (problem-dependent) the resolution capability deteriorates.

In the previous section

5. Bounds on $\|g_{LS} - g_{TLS}\|$. We begin this section by deriving a new upper bound for $\|g_{LS} - g_{TLS}\|$ which helps explain the similarities and differences of the LS and TLS methods. From [2] we know

$$\sin\theta \le \|g_{LS} - g_{TLS}\| \le \eta\sin\theta \tag{11}$$

where $\eta = \sqrt{1+\|g_{LS}\|^2}\sqrt{1+\|g_{TLS}\|^2}$ and $\sin\theta$ is the subspace angle between the nullspaces of $[A_p\ b_p]$ and $[\hat{A}\ \hat{b}]$.

The analytical bounds and the numerical experiments in [2] suggest that $\sin\theta$ is of order $(\bar{\sigma}_{p+1}/\sigma_p)^2$. The following result represents an improvement of those bounds and reflects

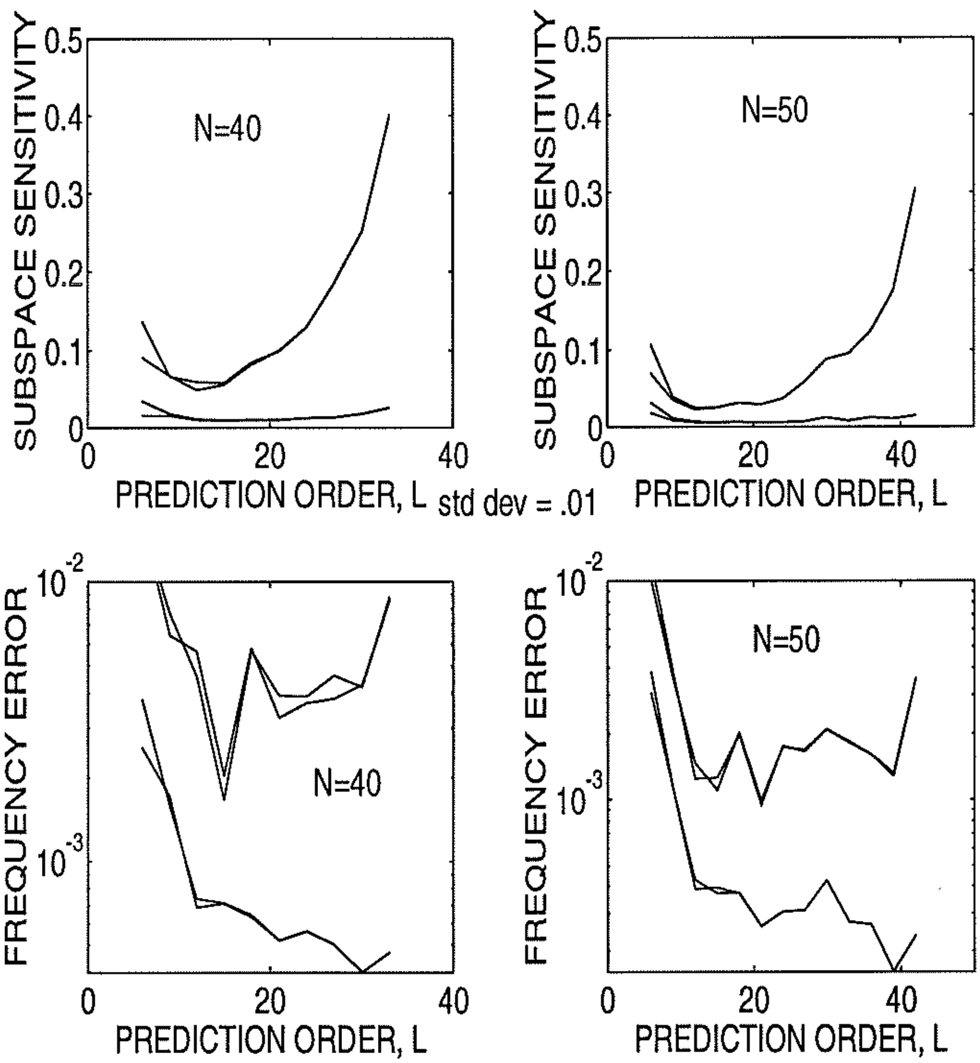

FIG. 1. *The results of subspace sensitivity and error in frequency estimation for $N = 40, 50$ versus prediction order L. The frequencies were $f_1 = 0.34$ Hz and $f_2 = 0.347$ Hz. The standard deviation of the white noise 0.01.*

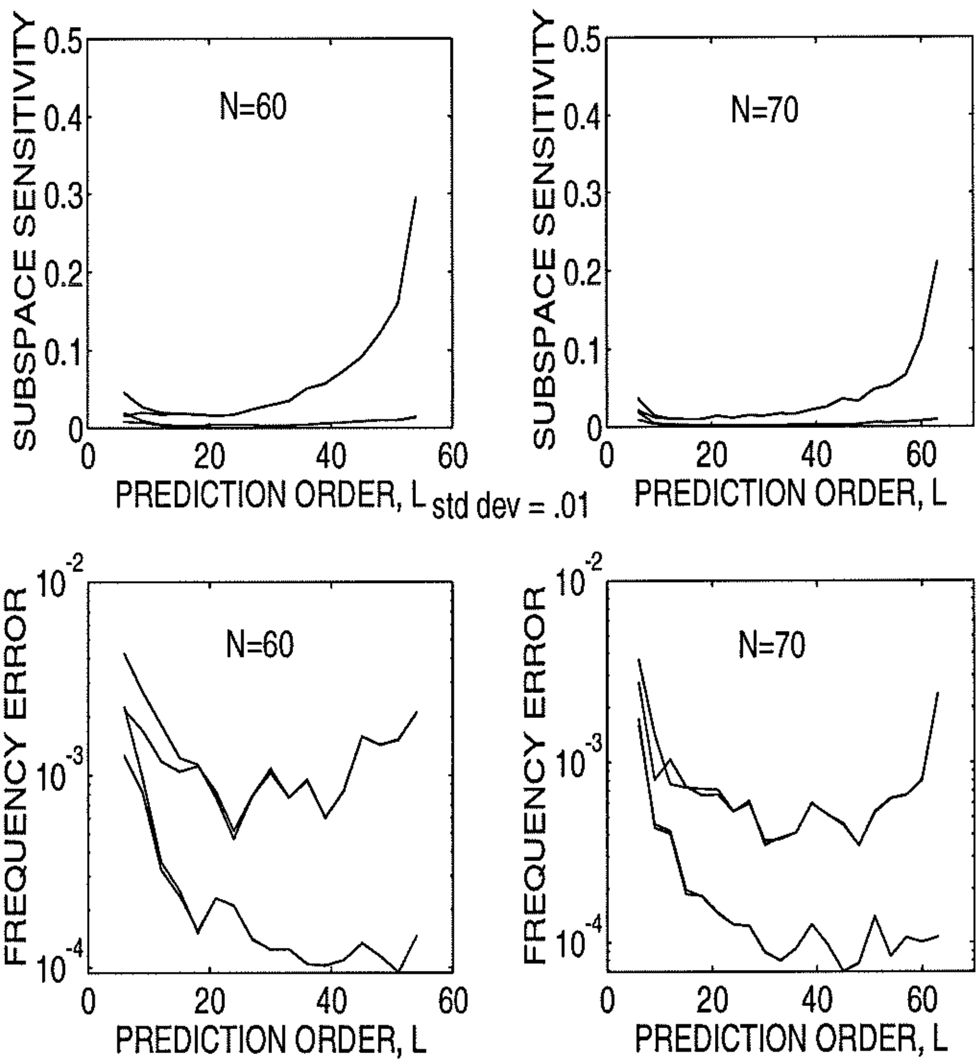

FIG. 2. *The results of subspace sensitivity and error in frequency estimation for $N = 60, 70$ versus prediction order L. The frequencies were $f_1 = 0.34$ Hz and $f_2 = 0.347$ Hz. The standard deviation of the white noise 0.01.*

these experimental observations. Corresponding results hold for the FBLP model.

THEOREM 5.1. *Let g_{LS} and g_{TLS} denote the truncated minimum norm LS and TLS solutions to (1). Let θ denote the subspace angle between $N([A_p\ b_p])$ and $N([\hat{A}\ \hat{b}])$ and define $\eta = \sqrt{1+\|g_{LS}\|^2}\sqrt{1+\|g_{TLS}\|^2}$. If $\frac{1}{2}\|[\delta A_0\ \delta b_0]\| < \sigma_{0_p}$ then*

$$\text{(12)} \qquad \sin\theta \leq \quad \|g_{LS} - g_{TLS}\| \quad \leq \eta\left(\frac{\bar{\sigma}_{p+1}}{\sigma_p}\right)^2$$

$$\text{(13)} \qquad \leq \eta\left(\frac{\|[\delta A_0\ \delta b_0]\|}{\sigma_{0_p} - \|\delta A_0\|}\right)^2.$$

Proof: Form (11) we need to show $\sin\theta \leq (\bar{\sigma}_{p+1}/\sigma_p)^2$. Let A and $[A\ b]$ have the block SVD forms $A = U_1\Sigma_1V_1^T + U_2\Sigma_2V_2^T$ and $[A\ b] = \bar{U}_1\bar{\Sigma}_1\bar{V}_1^T + \bar{U}_2\bar{\Sigma}_2\bar{V}_2^T$, with $A_p = U_1\Sigma_1V_1^T$ and $[\hat{A}\ \hat{b}] = \bar{U}_1\bar{\Sigma}_1\bar{V}_1^T$. Note that $[A_p\ b_p] = U_1U_1^T[A\ b]$. Further, let the rank p matrix $[A_p\ b_p]$ have the block SVD form $[A_p\ b_p] = U'\Sigma'V'^T$, where Σ' is a nonsingular matrix of order p. Since $\sin\theta = \|V'^T\bar{V}_2\|$, we first find an expression for $V'^T\bar{V}_2$.

$$\begin{aligned} V'^T\bar{V}_2 &= \Sigma'^{-1}U'^T[A_p\ b_p]\bar{V}_2 \\ &= \Sigma'^{-1}U'^TU_1U_1^T[A\ b]\bar{V}_2 \\ &= \Sigma'^{-1}U'^TU_1U_1^T\bar{U}_2\bar{\Sigma}_2. \end{aligned}$$

At this point we find an alternative expression for $U_1^T\bar{U}_2$.

$$\begin{aligned} U_1^T\bar{U}_2 &= \Sigma_1^{-1}V_1^TA^T\bar{U}_2 \\ &= \Sigma_1^{-1}[V_1^T\ 0][A\ b]^T\bar{U}_2 \\ &= \Sigma_1^{-1}[V_1^T\ 0]\bar{V}_2\bar{\Sigma}_2. \end{aligned}$$

Substituting, it follows

$$V'^T\bar{V}_2 = \Sigma'^{-1}U'^TU_1(\Sigma_1^{-1}[V_1^T\ 0]\bar{V}_2\bar{\Sigma}_2)\bar{\Sigma}_2.$$

Taking norms, we get

$$\begin{aligned} \|V'^T\bar{V}_2\| &\leq \|\Sigma'^{-1}\|\,\|U'^TU_1\|\,\|\Sigma_1^{-1}\|\,\|[V_1^T\ 0]\bar{V}_2\|\,\|\bar{\Sigma}_2\|^2 \\ &\leq \|\Sigma'^{-1}\|\,\|\Sigma_1^{-1}\|\,\|\bar{\Sigma}_2\|^2 \\ &\leq \|\Sigma_1^{-1}\|^2\,\|\bar{\Sigma}_2\|^2. \end{aligned}$$

The last inequality follows because $\|\Sigma'^{-1}\| \leq \|\Sigma_1^{-1}\|$. Therefore, we have shown

$$\text{(14)} \qquad \sin\theta \leq \left(\frac{\bar{\sigma}_{k+1}}{\sigma_k}\right)^2,$$

and this proves (12). Equation (13) follows immediately from the perturbation property of singular values for $[A\ b]$ and $[A_0\ b_0]$, as well as A_0 and A. This concludes the proof of the theorem.

REFERENCES

[1] R.D. Fierro and J.R. Bunch, *Collinearity and Total Least Squares*, SIAM J. Num. Anal. and Appl., 1993, accepted for publication.

[2] R.D. Fierro and J.R. Bunch, *Perturbation Theory for Orthogonal Projection Methods with Applications to Least Squares and Total Least Squares*, CAM Report 92-44, Department of Mathematics, University of California, Los Angeles, 1992.

[3] R.D. Fierro and K. Yao, *Applications of Perturbation Theory for Othogonal Projection Methods to the Linear Prediction Frequency Estimation Problem.*, CAM Technical Report, Department of Mathematics, University of California, Los Angeles, 1993.

[4] G.H. Golub and C.F. Van Loan, *Matrix Computations (Second Edition)*, John Hopkins University Press, 1989.

[5] R. Kumaresan and D.W. Tufts, *Improved Spectral Resolution III: Efficient Realization*, Proc. IEEE, 68, (1976), pp. 1354-1355.

[6] S.W. Lang and J.H. McLellan, *Frequency Estimation with Maximum Entropy Spectral Estimators*, IEEE Trans. Acoust., Speech, Signal Processing, 28, (1980), pp. 716-724.

[7] A.H. Nuttall, *Spectral Analysis of a Univariate Process with Bad Data Points via Maximum Entropy and Linear Prediction Techniques*, in Naval Underwater Systems Center (NUSC) scientific and engineering studies, Spectral Estimation, NUSC, NewLondon, CT, Mar. 1976.

[8] M.A. Rahman and K. Yu, *Total Least Squares Approach for Frequency Estimation Using Linear Prediction*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 35, (1987), pp. 1442-54.

[9] D.V. Bhaskar Rao, *Perturbation Analysis of a SVD Based Method for the Harmonic Retrieval Problem*, Proc. ICASSP 1985, pp. 624-627.

[10] R. Roy and T. Kailath, *ESPIRIT - Estimation of Signal Parameters via Rotational Invariance Techniques*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 37, (1989), pp. 984-995.

[11] T. Soderstrom and P. Stoica, *On the Accuracy of High-Order Yule-Walker Methods for Cisoids*, Proc. ICASSP 1985, pp. 3573-76.

[12] P. Stoica and T. Soderstrom, *On Spectral and Root Forms of Sinusoidal Frequency Estimators*, Proc. ICASSP 1985, pp. 3257-3260.

[13] P. Stoica, T. Soderstrom and F. Ti, *Asymptotic Properties of the High-Order Yule-Walker Estimates of Sinusoidal Frequencies*, IEEE Transactions on Acoustics, Speech and Signal Processing, 37, (1989), pp. 1721-1734.

[14] D.W. Tufts and R. Kumaresan, *Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood*, Proceedings of the IEEE, 70, (1982), pp. 975-989.

[15] T.J. Ulrych and R.W. Clayton, *Time Series Modelling and Maximum Entropy*, Phys. Earth Planet. Interiors, 12, (1976), pp. 188-200.

[16] S. Van Huffel and J. Vandewalle, *Analysis and Solution of the Nongeneric Total Least Squares Problem*, SIAM J. Matrix Anal. Appl., 3, (1988), pp. 360-372.