E3.3

LEAST SQUARES NULL SPACE VARIATIONAL CHARACTERIZATION FOR NONMINIMUM NORM SOLUTIONS

S. Konyk, Jr.*, M. G. Amin*, M. G. Lagunas**

*Department of Electrical Engineering
Villanova University
Villanova, PA 19085, USA

**Procesado de Senal en Comunicaciones
E.T.S.I. Telecomunicacion
Apdo. 30.002 08080 Barcelona, Spain

ABSTRACT

The least squares estimation problem with nonminimum norm constraints on the unknown model parameters is considered. Contrary to the quadratic constraint least squares solutions the approach presented does not necessarily satisfy the constraint, but rather relies on the nullity of the data matrix to maintain the unconstrained least squares error value while trading off the minimum norm solution by another with the shortest distance from the null space of the constraint. The SVD of the data matrix is used to obtain the necessary information about the minimum norm solution as well as the basis of the null space. Closed form expressions are derived for the case in which the constraint of interest is the smoothness of the model parameters. Examples of sinusoids in white noise are given for illustration.

I. INTRODUCTION

In the least-squares estimation (LSE) problem, the nonzero nullity property of the data matrix may arise mostly due to the nature of the data, as in the case of noise free sinusoids [1]. This property, however also manifests itself in the noisy environment as a consequence of processing the data prior to estimation to increase the signal-to-noise ratio [2]. In the latter case, the data is projected into the signal subspace and results in a reduced rank expression of the nonsingular data matrix. When constraints other than minimum norm exist on the optimum weight vector the null space of the data matrix can be searched in an attempt to satisfy the constraints without increasing the minimum norm solution error value.

In this paper the above LSE problem is considered in the context of a multiple linear regression model implemented by a transversal filter whose tap weights are the unknown model parameters. In this context, cases are considered for which the model parameter estimate is derived from a finite block of data samples whose corresponding data matrix possesses the nonzero nullity property. An optimal solution to this problem is, therefore, any model parameter vector that yields the minimum possible value of the sum of error squares performance criterion. This solution characterizes a parameter subspace whose dimension is equal to that of the data matrix nullity.

For the well-known minimum norm constraint, the result is a unique solution determined by the pseudo inverse of the data matrix and corresponds to selecting a particular member of the subspace of the weight vectors which intersect only the origin of the nullspace. The invariant performance imposed by the span of the null space is used as a reference in searching for and selecting an optimal LSE solution from a set of competitive suboptimal features.

Searching the nullspace is performed by first defining the LSE null space basis via singular value decomposition (SVD) of the data matrix [3]. These bases are then used to determine the additional null space weight components which when added to the minimum norm solution result in a total weight vector which has the shortest distance from the null space of the nonlinear constraints.

The paper presents an insightful approach to solve the above type LSE of problems. Detailed analysis is given to the case where smoothness is chosen as the nonlinear constraint of interest. Smoothness [4] is a measure of how close the filter weights are to a constant mean value. It is viewed as driving the filter towards a rectangular impulse response or equivalently a sinc function in the frequency domain. In the z domain, the smoothness constraint translate into equally spaced zeros of unit value excluding that of zero phase.

Two examples of one step forward linear prediction are presented in which smoothness constraints are considered as the quadratic constraints on the predictor weights. The data in both examples consists of two sinusoids with widely spread frequencies in additive gaussian white noise. The first example represents the noiseless case where the noise power level is set to zero. In this case, with $M \geq 4$, the nonzero nullity of the data matrix is a natural property of the least-squares problem formulation. The nullity in the noisy case, on the other hand, is established through the rank reduction process which invokes setting the noise eigenvalues to zero [5]. Both examples show that by moving in the null space of the data matrix, or its reduced rank form, a smoother filter impulse response can be obtained at no expense of the least-squares error.

II. NULL SPACE LEAST SQUARES

The least squares estimation problem is considered in the context of a linear predictor. The one step predictor output $\hat{u}(i)$ may be expressed as the convolution sum

$$\hat{d}(i) = \sum_{k=1}^{M} w_k u(i-k+1) \quad (1)$$

where M is the filter length, w_k are the filter weights and the $u(\cdot)$ are the tap inputs. Employing the covariance method for windowing the input data $u(i), i = 1, \ldots, N$; the sum of error squares measure is given by

$$J_{\mathcal{E}}(w_1, \ldots, w_M) = \sum_{i=M}^{N} |e(i)|^2 \quad (2)$$

where

$$e(i) = u(i) - \hat{u}(i) \quad (3)$$

Equation (2) may be expressed as

$$\begin{aligned} J_{\mathcal{E}}(\mathbf{w}) &= \mathbf{r}^T\mathbf{r} \\ &= (\mathbf{b} - \mathbf{A}\mathbf{w})^T(\mathbf{b} - \mathbf{A}\mathbf{w}) \end{aligned} \quad (4)$$

where

$$\mathbf{w} = [w_1, w_2, \ldots, w_M]^T, \quad (5)$$

$$\mathbf{r} = [e(M), e(M+1), \ldots, e(N)]^T, \quad (6)$$

$$\mathbf{b} = [u(M+1), u(M+2), \ldots, u(N+1)]^T, \quad (7)$$

and

$$\mathbf{A} = \begin{bmatrix} u(M) & u(M-1) & \ldots & u(1) \\ u(M+1) & u(M) & \ldots & u(2) \\ \vdots & \vdots & \vdots & \vdots \\ u(N) & u(N-1) & \ldots & u(N-M+1) \end{bmatrix} \quad (8)$$

The superscript ^{T} denotes transposition, $\mathbf{w}$ is the $M \times 1$ tap weight vector, $\mathbf{u}(i)$ is the $M \times 1$ input vector, $\mathbf{r}$ is the $M \times 1$ residual vector, and $\mathbf{A}$ is the $(M - N + 1) \times N$ data matrix.

The least squares solution of the filter weights, which minimizes $J_{\mathcal{E}}$ over the data window, satisfies the normal equation

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{w}} = \mathbf{A}^T\mathbf{b} \quad (9)$$

If $rank(\mathbf{A}) = M$, $\mathbf{A}\mathbf{A}^T$ is nonsingular and the tap weight vector is uniquely determined as $\hat{\mathbf{w}} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}^T\mathbf{b}$. On the other hand, for $rank(\mathbf{A}) = r < M$ the nullity of $\mathbf{A}$ is nonzero and the least squares solution ,$\hat{\mathbf{w}}$, is no longer unique. In this case, the pseudo inverse of the data matrix, denoted $\mathbf{A}^\dagger$, characterizes the minimum norm solution that is given by

$$\begin{aligned} \hat{\mathbf{w}} &= \mathbf{A}^\dagger\mathbf{b} \\ &= \mathbf{X}\Sigma^{-1}\mathbf{Y}^T \end{aligned} \quad (10)$$

The orthogonal transformation matrices $\mathbf{Y}$ and $\mathbf{X}$ of the singular value decomposition (SVD) of $\mathbf{A}$ characterise the null space, $N(\mathbf{A})$, and the range space, $R(\mathbf{A})$, of the data matrix, respectively. The first r diagonal entries of Σ satisfy $\sigma_1 \geq \sigma_2 \geq, \ldots, \sigma_M \geq 0$ while the rest of the matrix elements have zero values. The last $(M - r) + 1$ columns of the matrix $\mathbf{X}$ form an orthonormal basis for $N(A)$ and will be used in the next section to move away from the minimum norm solution toward satisfying the nonlinear constraint.

III. SMOOTHNESS CONSTRAINT CASE

The approach to null space least squares is applied to the case where a nonlinear smoothness constraint is imposed on the weight vector. The smoothness constraint provides a simple example which illustrates the fundamental mechanism of null space least squares. The goal here is to constrain the weight vector to the minimum possible smoothness measure which still results in the minimum least squares error. For a given weight vector $\mathbf{w}$, a measure of its smoothness is expressed as

$$\begin{aligned} J_s(\mathbf{w}) &= \sum_{i=1}^{M}(w_i - \overline{w})^2 \\ &= \mathbf{w}^T\mathbf{Q}\mathbf{w} \end{aligned} \quad (11)$$

where $\overline{w}$ is the average value of the weights and $\mathbf{Q}$ is an $M \times M$ symmetric matrix with entries

$$q_{i,j} = \begin{cases} \frac{M-1}{M} & \text{for } i = j \\ \frac{-1}{M} & \text{for } i \neq j \end{cases} \quad (12)$$

The behavior of the smoothness measure for $M = 2$ is illustrated in Figure 1. The null space of $\mathbf{Q}$ is one dimensional and spanned by the $M \times 1$ unit vector

$$\hat{\mathbf{1}} = \begin{bmatrix} \frac{1}{\sqrt{M}} \\ \vdots \\ \frac{1}{\sqrt{M}} \end{bmatrix} \quad (13)$$

A maximally flat impulse response corresponds to zero variance in the weight vector which implies perfect smoothness and thus

$$\mathbf{w} = c\hat{\mathbf{1}} \quad \text{for some } c \in \mathcal{R} \quad (14)$$

The set of column vectors $\mathbf{x}_{r+1}, \ldots, \mathbf{x}_M$ of $\mathbf{X}$ form an orthonormal basis for $N(\mathbf{A})$. Moving a weight vector through $N(\mathbf{A})$ does not change its LSE error, i.e. $J_{\mathcal{E}}(\mathbf{w}) = J_{\mathcal{E}}(\mathbf{w} + \Delta\mathbf{w}), \quad \forall \Delta\mathbf{w} \in N(\mathbf{A})$. Thus the least squares problem in terms of the bases for $N(\mathbf{A})$ and $N(\mathbf{Q})$ is: given the minimum norm least squares weight vector $\hat{\mathbf{w}}$, find the scalars $c, \alpha_{r+1}, \ldots, \alpha_M$ which minimize

$$J_{\mathcal{N}\mathcal{E}}(\mathbf{w}) = \left\| c\hat{\mathbf{1}} - \hat{\mathbf{w}}_{TOT} \right\|^2$$

$$\hat{\mathbf{w}}_{TOT} = \sum_{i=r+1}^{M} \alpha_i \mathbf{x}_i \quad (15)$$

It can be shown that the scalars in (15) which maximize smoothness without increasing the least squares error are

$$\hat{c} = \frac{\hat{\mathbf{1}}^T\hat{\mathbf{w}}}{1 - \sum_{i=r+1}^{M}\hat{\mathbf{1}}^T\mathbf{x}_i} \quad (16)$$

and

$$\hat{\alpha}_i = \hat{c}\hat{\mathbf{1}}\mathbf{x}_i \quad (17)$$

Two simulation examples are given in Section IV which use (15-17) to improve smoothness in the case of sinusoids with additive white noise.

IV. SIMULATIONS

Twenty data samples $x(n)$ $n = 1, \ldots, 20$ are generated with

$$x(n) = cos(\omega_1 n) + cos(\omega_2 n) + \sigma w(n) \qquad (18)$$

where $w(n)$ is a white noise sequence whose power level is chosen to provide infinite SNR in the noiseless case and $20dB$ SNR in the noisy case. In both cases, $\omega_1 = \pi/3$, $\omega_2 = \pi/6$ rad/sec and the filter length $M = 11$.

Noise free case. In this case, the null space of $\mathbf{A}$, $N(\mathbf{A})$, is of dimension 7. The minimum norm solution is $\hat{\mathbf{w}} =$ [0.3415, 0.0000, -0.2500, -0.5000, -0.0915, 0.0000, -0.0915, -0.2500, -0.2500, 0.0000, 0.3415]^T, which yields a smoothness $\hat{\mathbf{w}}^T \mathbf{Q} \hat{\mathbf{w}} = 0.4773$ and zero prediction. The zeros of the z domain weight vector polynomial include $\pm e^{j\pi/3}$, $\pm e^{j\pi/6}$ which correspond to the two sinusoidal frequencies. Incorporating $\hat{\mathbf{w}}$ with the last seven columns of the matrix $\mathbf{X}$ in equation (15) results in $\hat{c} = -3.3166$. From (16) and (17), $\hat{\mathbf{w}}_{TOT} = [-1, \ldots, -1]^T = -\hat{\mathbf{I}}^T$. Accordingly $\hat{\mathbf{w}}^T \mathbf{Q} \hat{\mathbf{w}} = 0$, i.e. maximum smoothness (flat response) is obtained while maintaining $J_\xi = 0$. As shown in Figure 2, the new set of zeros still include $\pm e^{j\pi/3}$, $\pm e^{j\pi/6}$, i.e., the polynomial roots which correspond to the input signal did not exhibit any displacement as a result of operating in $N(\mathbf{A})$. This property is satisfied for any set of values of ω_1 and ω_2. Figure 3 shows the weight vector before and after smoothing.

Noisy case. In this case, the SVD of $\mathbf{A}$ in first performed and followed by setting the last 7 values of the diagonal elements of Σ to zero. With the new Σ the reduced rank form of $\mathbf{A}$ is obtained using the same orthogonal matrices $\mathbf{X}$ and $\mathbf{Y}$. The rank reduction of $\mathbf{A}$ provides and an improvement of smoothness over the full rank original data matrix and reduces $\hat{\mathbf{w}}^T \mathbf{Q} \hat{\mathbf{w}}$ from 3.3911 to 0.4954. The rank reduction however causes an increase in the LSE error from approximately zero to 0.2533. After reducing r from 11 to 4 the problem becomes similar to the noiseless case. The corresponding value of $\hat{c}$ is found to be -3.8085 and the total weight vector obtained from equation (15) has kept $J_\xi = .2533$ while further reducing the smoothness to 0.0246. Figure 4 shows a plot of the three different values of the optimum weight vector results from the above process. While the weight vector, based on the original data matrix is highly unsmooth, the nonminimum norm weight vector of the reduced rank data matrix is almost flat. This smoothness improvement is also shown in Figure 5 which illustrates the immigration of the z domain roots of the weight vector polynomial towards the unit circle.

V. CONCLUSIONS

The paper presented an insightful approach to the least squares estimation problem in which 1) the data matrix has nonzero nullity, 2) it is desired to minimize a nonlinear quadratic function in the estimator weight vector. Contrary to the solution of the constraint least squares problem [6,7] which satisfies the constraint by trading off the unconstrained least squares error value, the weight vector in the underlying problem is set to be as close as possible to the constraint null space without an increase in the estimation error. The key to this approach is to define the basis of the null space via the SVD of the data matrix which also provides the minimum norm solution. There bases are then used to drive the weight vector away from its minimum norm value to an optimum solution which maintains the LSE error and has the shortest distance from the null space of the constraint. A theoretical frame work is developed with closed form expression for the case where smoothness represents the quadratic function. Verification is given using examples of sinusoid in white noise.

REFERENCES

[1] S. Marple, Digital Spectral Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[2] S. Haykin, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1986.

[3] G. Golub and C. Van Loan, Matrix Computations, Johns Hopkins University Press, 1983.

[4] S. Konyk and M. Amin,"A new class of nonlinearly constrained linear estimators," IEEE International Conference on Acoustics, Speech and Signal Procss., New York, NY, April, 1988.

[5] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids making linear prediction perform like maximum likelihood," Proceedings of the IEEE, pp 975-989, 1982.

[6] O. Frost, "An algorithm for linearly constrained adaptive array processing," Proceedings of the IEEE, August, 1972.

[7] T. Mu and L. Griffiths, "A solution space approach to achieving partially adaptive arrays," IEEE International Conference on Acoustics, Speech and Signal Procss., New York, NY, April, 1988.

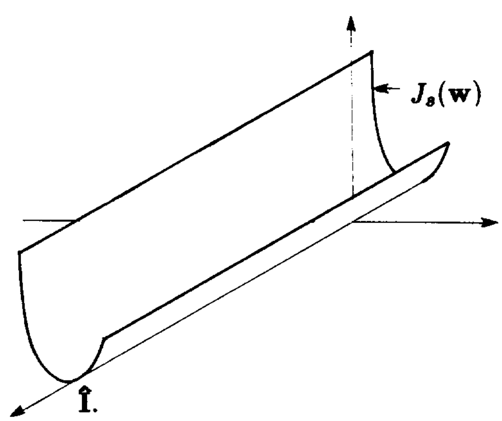

Fig. 1 Illustration of the invariance of the smoothness measure, $J_s(\mathbf{w})$, with respect to its null space $\hat{\mathbf{I}}$.

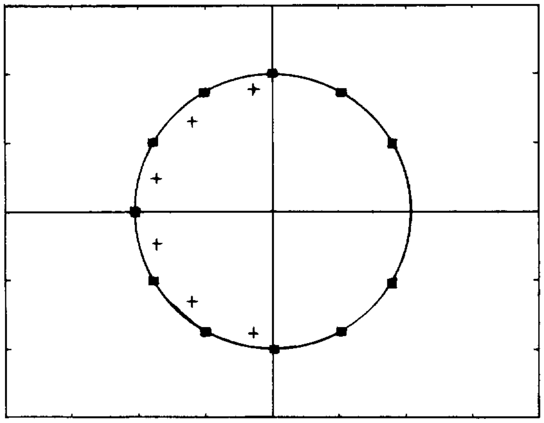

Fig. 2 Zero diagram of the LSE weight vector in the noise free example, (×) for $\hat{\mathbf{w}}$, (■) for $\mathbf{w}_{TOT}$.

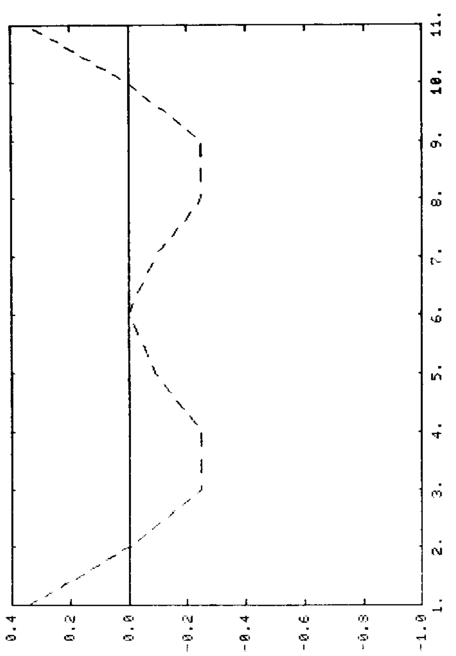

Fig. 3 The weight vector of pre- and post-smoothing in the noise free case.

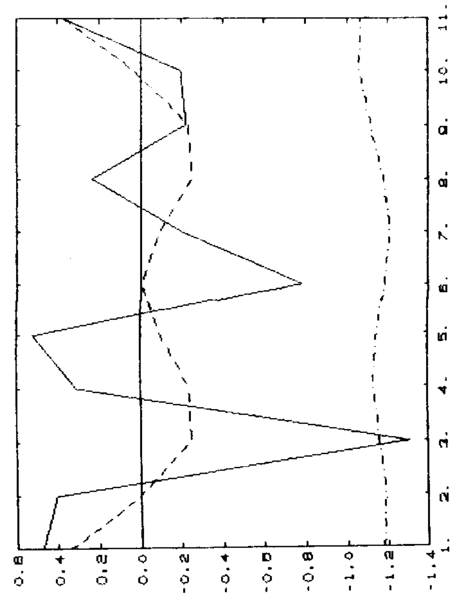

Fig. 4 The weight vector of pre- and post-smoothing in the noisy data case.

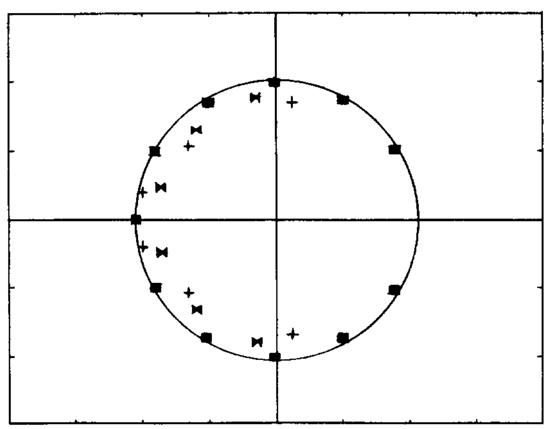

Fig. 5 Zero diagram of the LSE weight vector in the noisy data example, (×) for $\hat{\mathbf{w}}$, (■) for $\mathbf{w}_{TOT}$, (⋈) for $\hat{\mathbf{w}}$ of the reduced rank data matrix.