1992 ACC/TM13

# System Identification via Weighted Subspace Fitting

A. SWINDLEHURST
*Dept. of Electrical & Computer Eng.*
*Brigham Young University*
*Provo, UT 84602*
*(801) 378-4012*

R. ROY
*Systems Research Associates, Inc.*
*10030 Phar Lap Dr.*
*Cupertino, CA 95014*

B. OTTERSTEN
*Dept. of Electrical Eng.*
*Linköping University*
*S-581 83 Linköping*
*Sweden*

T. KAILATH
*Information Systems Laboratory*
*Stanford University*
*Stanford, CA 94305*

**Abstract**

*This paper presents a new method for the identification of linear systems parameterized by state space models. The method relies on the concept of subspace-fitting, in which the goal is to find a particular input/output data model parameterized by the state matrices that best fits, in the least-squares sense, the dominant subspace of the measured data. Central to this approach is the idea that a weighting may be applied to the observed dominant subspace in order to emphasize certain directions where the signal-to-noise ratio is highest. This has the advantage of making the algorithm robust to systems that are nearly unobservable, or to those whose state space has not been sufficiently excited. Some empirical results are included to validate the algorithm and illustrate its advantages over previous techniques. In addition to presenting the theory and implementation of the new method, this paper also illustrates some interesting connections between state space data models and those encountered in processing the signals received by an array of sensors.*

## 1. Introduction

RESEARCH IN THE IDENTIFICATION of linear, discrete-time systems has been dominated in recent years by work involving input/output system descriptions, including *autoregressive* (AR) and autoregressive *moving-average* (ARMA) data models and their various derivatives (*e.g.*, [1, 2]). In these approaches, the input and output data are assumed to be related by a simple finite difference equation parameterized by a relatively small number of coefficients. Given a particular sequence of input data, the coefficients are typically estimated as those that minimize the squared-error difference between the predicted and measured system output. Algorithms based on this idea are consequently often referred to as *prediction-error methods* (PEMs).

Although linear state space models have become commonplace in estimation and control, they have not been widely used in identification. Implicitly, of course, AR- and ARMA-type models correspond directly with certain canonical state space realizations, and thus a PEM, for example, might be thought of as finding an appropriate state space model for the system. One of the goals of this paper is to demonstrate, however, that there are some important advantages in explicitly considering more general state space forms in identification.

Several early approaches to the general state space identification problem were based on examining the structure of a Hankel matrix composed of samples of the impulse response of the system. Ho and Kalman's algorithm [3], for example, addressed the associated *realization* problem, and exploited the shift structure of the Hankel matrix to determine the system's state matrices. Kung [4] later proposed a least-squares technique based on the singular value decomposition (SVD) of the Hankel matrix that was applicable when measurement noise was present. More recently, De Moor [5, 6] has developed a total-least-squares algorithm that exploits the same shift structure present in a certain input/output data model, and that allows arbitrary input excitations. Related methods have also been recently proposed by Moonen, *et al.*[7], and Verhaegen, *et al.*[8, 9].

The method presented in this paper also uses the shift structure in the input/output model of De Moor, but exploits it in a different way. The motivation for this new algorithm comes from some interesting connections linking the state space identification problem with sensor array signal processing. In particular, it is shown how the identification problem can be cast in the *subspace fitting* framework, where the goal is to find the input/output model which best fits (in the least-squares sense) the dominant subspace of the data. This approach has been successfully applied by Ottersten and Viberg in the context of narrowband direction-of-arrival (DOA) estimation [10, 11]. Of special interest is the fact that the dominant subspace can be weighted to emphasize certain directions in the so-called "signal" subspace where the signal-to-noise ratio is high. This weighting may potentially provide an advantage in cases involving nearly unobservable systems or an insufficiently excited state space.

In the next section, the data model assumed for this problem is introduced, and the similarities of this model with that of the DOA estimation problem are also elucidated. Section 3 partially describes the methods of De Moor and Verhaegen, and shows how they are related to the recently introduced DOA estimation algorithms referred to as ESPRIT [12, 13] and TAM [14]. The subspace fitting approach is then presented in Section 4, and the issue of subspace weighting and identifiability is discussed. A numerical example concludes the paper in Section 5.

## 2. An Input/Output Data Model

Consider the following multiple-input, multiple-output (MIMO) time-invariant linear system in state space form

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k \end{aligned} , \qquad (1)$$

where $\mathbf{x}_k \in R^n$, $\mathbf{u}_k \in R^m$, $\mathbf{y}_k \in R^l$, and the system matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{D}$ are of consistent dimension. If several observations of the input and output vectors are available, they may be grouped together into the single equation (*e.g.*, see [5, 6])

$$\mathbf{Y} = \Gamma\mathbf{X} + \mathbf{H}\mathbf{U} , \qquad (2)$$

[1]This work was supported by the National Science Foundation under grant MIP-9110112.

where

$$\mathbf{Y} = \begin{bmatrix} y_k & y_{k+1} & \cdots & y_{k+j-1} \\ y_{k+1} & y_{k+2} & \cdots & y_{k+j} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k+i-1} & y_{k+i} & \cdots & y_{k+i+j-2} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{i-1} \end{bmatrix}$$

$$\mathbf{X} = [\mathbf{x}_k \ \mathbf{x}_{k+1} \ \cdots \ \mathbf{x}_{k+j-1}]$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{i-2}\mathbf{B} & \mathbf{CA}^{i-3}\mathbf{B} & \mathbf{CA}^{i-4}\mathbf{B} & \cdots & \mathbf{D} \end{bmatrix},$$

and where $\mathbf{U}$ is an $(mi) \times j$ block Hankel matrix constructed exactly as $\mathbf{Y}$, but containing samples of the input sequence. With measurement noise present, $\mathbf{Y}$ would be augmented by an $(li) \times j$ matrix $\mathbf{V}$ containing the disturbance terms.

The *observable* subspace $\mathcal{O}$ and the *excited* subspace $\mathcal{E}$ are defined as

$$\begin{aligned} \mathcal{O} &= \text{span}\{\Gamma_i^T\} \subseteq R^n \\ \mathcal{E} &= \text{span}\{\mathbf{X}\} \subseteq R^n, \end{aligned}$$

where *span* denotes the column space. The dimension of these subspaces will be denoted by $n_o = \dim(\mathcal{O})$ and $n_e = \dim(\mathcal{E})$, and that of their intersection by $n_{oe} = \dim(\mathcal{O} \cap \mathcal{E})$. Note that in general $n_o \geq \text{rank}(\Gamma)$. It is assumed that the block dimension $i$ has been chosen large enough to ensure $n_o = \text{rank}(\Gamma)$, *i.e.*, the dimension $i$ will be assumed to be at least as large as the *observability index* of the system. This will imply necessarily that $li \geq n_o$.

## 2.1. Connections with Sensor Array Data

The method presented in this paper is based on certain similarities between the above data model and that usually assumed for (narrowband) data collected by an array of sensors. To be more precise, assume an $m$-element array of sensors, and suppose that the signals emitted by $d$ narrowband energy sources impinge upon the array. The sensors could represent hydrophones on the ocean floor or towed behind a submarine, or they could be antenna elements in a phased-array radar. Figure 1 depicts a generic scenario for which $m = 3$ and $d = 2$. If we let $\mathbf{g}(\theta_i) \in \mathbb{C}^m$ be the (complex) array response to the $i^{\text{th}}$ signal, then the output of the array at time $t$ is given by

$$\mathbf{z}(t) = \mathbf{G}(\theta_0)\mathbf{s}(t) + \mathbf{n}(t), \tag{3}$$

where $\mathbf{s}(t) \in \mathbb{C}^d$ is a vector containing the amplitude and phase of the signals at time $t$, $\mathbf{n}(t)$ is additive noise, and

$$\mathbf{G}(\theta_0) \stackrel{\text{def}}{=} [\mathbf{g}(\theta_1) \ \cdots \ \mathbf{g}(\theta_d)].$$

If the output of the array is observed at, say, $N$ time instants, equation (3) may be written as

$$\mathbf{Z}_N = \mathbf{G}(\theta_0)\mathbf{S}_N + \mathbf{N}_N, \tag{4}$$

where $\mathbf{Z}_N = [\mathbf{z}(t_1) \ \cdots \ \mathbf{z}(t_N)]$, and $\mathbf{S}_N$ and $\mathbf{N}_N$ are similarly defined.

The quantity $\theta_i$ associated with each emitter is a vector which parameterizes the array response for that emitter; it could include such information as azimuth and elevation angles of arrival (commonly referred to as the direction-of-arrival or DOA), range, wavelength, polarization, or doppler shift of the signal. The set of all possible responses of the array is defined to be the set $\mathcal{G} = \{\mathbf{g}(\theta) : \theta \in \Theta\}$ for some region $\Theta$ in the parameter space. This set is often referred to as the *array manifold*, and is assumed to be known, either analytically or via some calibration procedure, and to be *unambiguous* (*i.e.*, any collection of $d \leq m$ distinct vectors from $\mathcal{G}$ is a linearly independent set). Note that if the array is unambiguous, and if no two signals are perfectly correlated (*i.e.*, if $\text{rank}(\mathbf{S}_N) = d$), then the signal part of the data, $\mathbf{G}(\theta_0)\mathbf{S}_N$, will span a $d$-dimensional subspace of $\mathbb{C}^m$. This subspace is often referred to as the *signal* subspace, and its orthogonal complement is referred to as the *noise* subspace.

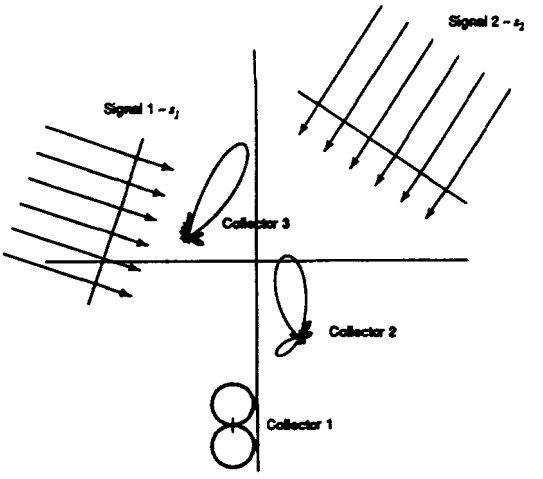


*Figure 1:* Generic Sensor Array Geometry

With these definitions, the generic antenna array parameter estimation problem can be stated as follows:

> *Given $N$ observations $\mathbf{z}(t)$, $t = 1, \cdots, N$, from an antenna array, estimate the number of signals $d$ impinging upon the array, the signal waveforms $\mathbf{s}(t)$, $t = 1, \cdots, N$, and the parameter matrix*
>
> $$\theta_0 = [\theta_1 \ \cdots \ \theta_d]$$
>
> *associated with the emitters.*

Many techniques have been developed over the years to tackle this problem. Of particular relevance to this paper is the class of *subspace based* methods which break the problem into the following two steps: obtain an estimate of the low rank signal subspace, and then find the parameters of the model that in some sense "best match" this estimated subspace.

To see how the linear system data model of (2) might be related to the DOA estimation problem, suppose $j \geq mi$ so that a non-trivial matrix $\mathbf{U}^\perp$ could be found satisfying $\mathbf{U}\mathbf{U}^\perp = \mathbf{0}$. Then, with measurement noise $\mathbf{V}$ present, we have

$$\mathbf{Y}\mathbf{U}^\perp = \Gamma\mathbf{X}\mathbf{U}^\perp + \mathbf{V}\mathbf{U}^\perp. \tag{5}$$

With suitably large choices for $i$ and $j$ (to be made more specific in the next section), it is seen that equation (5) decomposes $\mathbf{Y}\mathbf{U}^\perp$ into the sum of a low rank "signal" term $\Gamma\mathbf{X}\mathbf{U}^\perp$ and a full rank "noise" term $\mathbf{V}\mathbf{U}^\perp$.

Referring back to equation (4), and associating terms as follows:

$$\begin{aligned} \mathbf{Y}\mathbf{U}^\perp &\leftrightarrow \mathbf{Z}_N \\ \Gamma &\leftrightarrow \mathbf{G}(\theta_0) \\ \mathbf{X}\mathbf{U}^\perp &\leftrightarrow \mathbf{S}_N \\ \mathbf{V}\mathbf{U}^\perp &\leftrightarrow \mathbf{N}_N, \end{aligned}$$

it is clear that equation (5) is analogous to the data model for the narrowband sensor array problem. This observation

leads to some interesting and useful parallels that will be important later on. In particular, the signal subspace of (5) will be (nearly) rank deficient if:

1. $\Gamma$ is (nearly) rank-deficient, *i.e.*, if the system is not *observable*;
2. $\mathbf{X}$ is (nearly) rank-deficient, *i.e.*, if the input $\mathbf{U}$ has not excited the full state space;
3. or if $\rho(\mathbf{X}\mathbf{U}^{\perp}) < \rho(\mathbf{X})$, *i.e.*, if there has been a *rank cancellation* due to multiplication by $\mathbf{U}^{\perp}$.

Thus, a *weakly* observable system corresponds to a near ambiguity in the array manifold, or a DOA estimation scenario involving two very closely spaced sources (*i.e.*, $\mathbf{G}(\theta_0)$ nearly rank deficient), while the problems of insufficient excitation and rank cancellation correspond to the coherent signal case (*i.e.*, $\mathbf{S}_N$ nearly rank deficient). Given the similarity of the two models, one would suspect that similar algorithms could be applied to solve both problems. Sections 3 and 4 demonstrate that this is indeed the case.

# 3. Subspace Based Identification

Subspace based methods for state space identification break the problem into three fundamental steps:

1. *Estimate the Low Rank Subspace:* From the input/output data $\mathbf{U}$ and $\mathbf{Y}$ in equation (2), find a subspace that approximately spans the columns of $\Gamma$.
2. *Estimate* $\mathbf{A}$ *and* $\mathbf{C}$*:* Using the approximate shift invariant structure of this subspace, estimate the system matrices $\mathbf{A}$ and $\mathbf{C}$.
3. *Solve for* $\mathbf{B}$ *and* $\mathbf{D}$*:* Using the structure of $\mathbf{H}$, and either the estimates of $\mathbf{A}$ and $\mathbf{C}$ or the estimated low rank subspace, solve a set of linear equations for the matrices $\mathbf{B}$ and $\mathbf{D}$.

The discussion below will focus on steps 1 and 2, and a description of step 3 will not be included. The interested reader may find such a description, for example, in the thesis of De Moor [5].

## 3.1. Estimating the Low Rank Subspace

Two methods have been proposed for estimating the signal subspace of equations (2) and (5). They are the method of De Moor, *et al.*[5, 6], and the MOESP approach of Verhaegen, *et al.*[8, 9]. Although these two techniques have recently been shown to yield essentially identical results [15], we briefly describe them both below.

In De Moor's algorithm, it is assumed that the matrix $\mathbf{U}^{\perp}$ in equation (5) can be found. The existence of such a non-trivial $\mathbf{U}^{\perp}$ requires $j \geq mi$, but since $\mathbf{Y}\mathbf{U}^{\perp}$ must be at least rank $n$, it is usually assumed that $j > mi + n$. Provided the covariance $\Sigma = \mathcal{E}\{(\mathbf{V}\mathbf{U}^{\perp})(\mathbf{V}\mathbf{U}^{\perp})^T\}$ is known, and provided $\mathbf{X}\mathbf{U}^{\perp}$ is rank $n$, a set of vectors that approximately spans the same space as $\Gamma$ can be found from the generalized eigendecomposition of the pair $(\mathbf{R}_{\mathbf{Y}\mathbf{U}^{\perp}}, \Sigma)$, where

$$\mathbf{R}_{\mathbf{Y}\mathbf{U}^{\perp}} = (\mathbf{Y}\mathbf{U}^{\perp})(\mathbf{Y}\mathbf{U}^{\perp})^T .$$

For example, if $\mathbf{E}_s$ denotes the $n$ eigenvectors corresponding to the $n$ largest generalized eigenvalues of $(\mathbf{R}_{\mathbf{Y}\mathbf{U}^{\perp}}, \Sigma)$, then in the limit as $j \to \infty$,

$$\mathbf{E}_s = \Gamma\mathbf{T}$$

for some full rank $n \times n$ matrix $\mathbf{T}$.

The MOESP approach assumes that the state at time $k$ is independent of the input from time $k$ on. Consequently, $\mathbf{X}\mathbf{U}^T \to 0$ as $j \to \infty$, and hence

$$\lim_{j\to\infty} (\Gamma\mathbf{X})(\mathbf{H}\mathbf{U})^T = 0 .$$

Thus, the two components of the sum for $\mathbf{Y}$ in equation (2) are asymptotically orthogonal, and the row space of $\mathbf{Y}$ may be approximately (for finite $j$) decomposed as follows (assuming measurement noise is present):

$$\begin{aligned}\text{span}\{\mathbf{Y}^T\} &= \text{span}\{(\Gamma\mathbf{X} + \mathbf{H}\mathbf{U} + \mathbf{V})^T\} \\ &= \text{span}\{(\mathbf{H}\mathbf{U})^T\} \oplus \text{span}\{\mathbf{P}^{\perp}_{\mathbf{U}^T}(\Gamma\mathbf{X} + \mathbf{V})^T\} \\ &= \text{span}\{(\mathbf{H}\mathbf{U})^T\} \oplus \text{span}\{(\Gamma\mathbf{X})^T + \mathbf{P}^{\perp}_{\mathbf{U}^T}\mathbf{V}^T\} .\end{aligned}$$

Under the above assumption, it is easily seen that the LQ decomposition

$$\begin{bmatrix} \mathbf{U} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} ,$$

leads to

$$\mathbf{L}_{22}\mathbf{Q}_2^T = \Gamma\mathbf{X} + \mathbf{V}\mathbf{P}^{\perp}_{\mathbf{U}^T} ,$$

where $\mathbf{P}^{\perp}_{\mathbf{U}^T}$ is the $j \times j$ projection onto the orthogonal complement of the row space of $\mathbf{U}$. Provided the covariance $\Sigma = \mathcal{E}\{\mathbf{V}\mathbf{P}^{\perp}_{\mathbf{U}^T}\mathbf{V}^T\}$ is known and $\mathbf{X}$ is full rank, a set of vectors that spans the column space of $\Gamma$ may be obtained from the generalized eigendecomposition of $(\mathbf{L}_{22}\mathbf{L}_{22}^T, \Sigma)$ as above.

Note that in both of the above approaches, the covariance $\Sigma$ can be ignored in situations where the input $\mathbf{U}$ is persistently exciting [1], the measurement noise is white, and $j$ is sufficiently large [15]. How large $j$ has to be will depend on the magnitude of the noise. By "ignoring" $\Sigma$, we mean that it can be assumed to be a scaled identity matrix, and the generalized eigendecompositions above can be replaced by simple eigendecompositions. In such cases, one may equivalently choose to perform an SVD on $\mathbf{Y}\mathbf{U}^{\perp}$ or $\mathbf{L}_{22}$ directly.

## 3.2. Estimating the System Matrices

Once a collection of vectors $\mathbf{E}_s$ spanning the column space of $\Gamma$ has been found, the special shift invariance inherent in this subspace may be exploited to estimate the system matrices $\mathbf{A}$ and $\mathbf{C}$. If $\mathbf{E}_s$ is partitioned into the $l \times n$ blocks $\mathbf{E}_s = [\mathbf{E}_0^T \;\; \mathbf{E}_1^T \;\; \cdots \;\; \mathbf{E}_{i-1}^T]^T$, this shift invariant structure is illustrated by the equation

$$\mathbf{E}_s = \begin{bmatrix} \mathbf{E}_0 \\ \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_{i-1} \end{bmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \mathbf{C}\mathbf{A}^2 \\ \vdots \\ \mathbf{C}\mathbf{A}^{i-1} \end{bmatrix} \mathbf{T} , \qquad (6)$$

where $\mathbf{T}$ is a full-rank $n \times n$ matrix. Since in practice we must work with finite $j$, this relationship cannot be directly exploited to obtain the matrices $\mathbf{A}$ and $\mathbf{C}$. Consequently, only an estimate $\hat{\mathbf{E}}_s$ of the low rank subspace is available, equation (6) will not hold exactly, and an estimation procedure is required to determine $\mathbf{A}$ and $\mathbf{C}$.

The first approach to this estimation problem is originally due to Kung [4], and has recently been extended by De Moor *et al.*[5, 6]. The basis of this method is the observation that

$$\mathbf{E}^{\downarrow} = \mathbf{E}^{\uparrow}\mathbf{T}^{-1}\mathbf{A}\mathbf{T} , \qquad (7)$$

where $\mathbf{E}^{\downarrow}$ and $\mathbf{E}^{\uparrow}$ are just shifted versions of the spanning set $\mathbf{E}$:

$$\mathbf{E}^{\uparrow} = \begin{bmatrix} \mathbf{E}_0 \\ \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_{i-2} \end{bmatrix} \quad \text{and} \quad \mathbf{E}^{\downarrow} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \vdots \\ \mathbf{E}_{i-1} \end{bmatrix} .$$

Since the system matrix $\mathbf{A}$ can only be identified up to a similarity transformation (or state coordinate transformation), $\mathbf{E}^{\uparrow}$ and $\mathbf{E}^{\downarrow}$ are related by a viable candidate $\tilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ for the state matrix of the identified system. The particular transformation $\mathbf{T}$ depends on how $\mathbf{E}_s$ was obtained.

Due to the finite observation interval, only the estimates $\hat{\mathbf{E}}^{\uparrow}$ and $\hat{\mathbf{E}}^{\downarrow}$ are available. The *least-squares* (LS) approach of Kung was simply to set $\hat{\mathbf{C}} = \hat{\mathbf{E}}_0$ and estimate $\tilde{\mathbf{A}}$ using a least-squares fit:

$$\hat{\mathbf{A}} = (\hat{\mathbf{E}}^{\uparrow T}\hat{\mathbf{E}}^{\uparrow})^{-1}\hat{\mathbf{E}}^{\uparrow T}\hat{\mathbf{E}}^{\downarrow} .$$

De Moor later observed that since $\hat{\mathbf{E}}^{\uparrow}$ and $\hat{\mathbf{E}}^{\downarrow}$ were both in error, a *total* least-squares (TLS) estimate [16] of $\tilde{\mathbf{A}}$ is more appropriate. The TLS estimate of $\tilde{\mathbf{A}}$ is given by the product $-\mathbf{Q}_{21}\mathbf{Q}_{22}^{-1}$ defined in the SVD

$$\begin{bmatrix} \hat{\mathbf{E}}^{\uparrow T} \\ \hat{\mathbf{E}}^{\downarrow T} \end{bmatrix} [\hat{\mathbf{E}}^{\uparrow} \mid \hat{\mathbf{E}}^{\downarrow}] = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix} \mathbf{\Lambda} \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}^T . \tag{8}$$

Both the LS and TLS approaches described above are closely related to the LS and TLS versions of the ESPRIT algorithm of Roy, Paulraj, and Kailath [12, 13].

While these techniques are straightforward and simple to implement, they treat information from all dimensions of the low rank subspace equally – including the information from less reliable dimensions where the signal contribution is small. The next section describes a subspace fitting technique, similar to the WSF algorithm of Ottersten and Viberg [10, 11] that allows the subspace to be weighted in order to reduce variability in the parameter estimates due to ill-conditioning.

# 4. Weighted Subspace Fitting

It can be shown [10, 17] that the TLS estimate of $\tilde{\mathbf{A}}$ obtained from (8) is equivalent to the estimate obtained from the following *subspace fitting* minimization problem:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A},\mathbf{E}^{\uparrow},\mathbf{T}} \left\| \begin{bmatrix} \hat{\mathbf{E}}^{\uparrow} \\ \hat{\mathbf{E}}^{\downarrow} \end{bmatrix} - \begin{bmatrix} \mathbf{E}^{\uparrow} \\ \mathbf{E}^{\uparrow}\mathbf{A} \end{bmatrix} \mathbf{T} \right\|_F^2 . \tag{9}$$

In this minimization, the matrices $\mathbf{E}^{\uparrow}$ and $\tilde{\mathbf{A}}$ are estimated as those that, in the least-squares sense, most closely fit the shift invariant structure of the estimated subspace. This fact, coupled with equation (6), suggests that the more general subspace fitting problem below be considered:

$$\hat{\mathbf{C}}, \hat{\mathbf{A}} = \arg \min_{\mathbf{A},\mathbf{C},\mathbf{T}} \left\| \begin{bmatrix} \hat{\mathbf{E}}_0 \\ \hat{\mathbf{E}}_1 \\ \hat{\mathbf{E}}_2 \\ \vdots \\ \hat{\mathbf{E}}_{i-1} \end{bmatrix} \mathbf{W}^{1/2} - \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{p-1} \end{bmatrix} \mathbf{T} \right\|_F^2 . \tag{10}$$

In this problem, the full structure of the low rank subspace is exploited in estimating the system matrices $\mathbf{A}$ and $\mathbf{C}$. This minimization problem is identical in form to the weighted subspace fitting (WSF) approach presented in [17] for narrowband DOA estimation problems involving arrays composed of multiple identical subarrays. In fact, once an appropriate parameter vector has been identified for the minimization of (10), an algorithm essentially identical to that described in [17] may be used to estimate $\mathbf{A}$ and $\mathbf{C}$. As in the case of WSF, the more computationally efficient single-shift approaches of Section 3 may be used to provide initial estimates for the gradient search.

## 4.1. Identifiability

In terms of input/output response, a given realization $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ is of course indistinguishable from one defined in another state coordinate system $\{\mathbf{T}^{-1}\mathbf{AT}, \mathbf{T}^{-1}\mathbf{B}, \mathbf{CT}, \mathbf{D}\}$. Consequently, it is impossible in the above problem to uniquely estimate all of the possible $(l+n)n$ elements of $\mathbf{A}$ and $\mathbf{C}$. Instead, a unique parameterization of these matrices must be chosen. There is thus a slight abuse of notation in equation (10); it would more appropriately involve the minimization of the criterion function over elements of the given parameter set.

There are a variety of parameterizations that could be used, depending on any assumptions one is willing to make about the system. For example, if the system is multiple-input, single-output, a particularly relevant choice is to assume that the system is in observer form, and hence that $\mathbf{A}$ is left-companion and $\mathbf{C} = [1 \;\; 0 \;\; \cdots \;\; 0]$. For multiple input, multiple output systems, one might be willing to assume that the state matrix $\mathbf{A}$ is diagonalizable, and implicitly force $\mathbf{T}$ to be the eigenvectors of the estimate $\hat{\mathbf{A}}$. Since this determines $\mathbf{T}$ to within a diagonal scaling, one of the rows of $\hat{\mathbf{C}}$ could then be held fixed in order to guarantee uniqueness. Assuming a diagonalizable $\mathbf{A}$ also allows one to directly impose *a priori* information about the location of the poles of the system into the minimization of (10).

Note that if the system is unobservable ($n_o < n$) or if the state space has not been sufficiently excited ($n_s < n$), then only a lower order state space model may be identified with the system. In fact, the resulting estimated state space will have dimension $n_{oe}$ since the signal subspace of $\mathbf{YU}^{\perp}$ is rank $n_{oe}$ in such cases.

## 4.2. Subspace Weighting

In simple terms, the presence of the weighting matrix $\mathbf{W}$ in the minimization of (10) allows certain directions or dimensions of the low rank subspace to be emphasized over others. For example, in cases where this subspace is nearly rank deficient because the system is nearly unobservable or because the input is not persistently exciting, certain directions in the subspace will have a high variance associated with them due to the additive measurement noise. This fact is evident when the low rank subspace is determined, since certain columns of $\mathbf{E}_s$ will have relatively small generalized eigenvalues associated with them. One would expect that the appropriate weighting would be a function of these eigenvalues.

A possible candidate for the subspace weighting is that proposed by Ottersten and Viberg [10] for the sensor array case:

$$\mathbf{W}_{WSF} = (\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2\mathbf{I})^2\hat{\mathbf{\Lambda}}_s^{-1} , \tag{11}$$

where the diagonal matrix $\hat{\mathbf{\Lambda}}_s$ represents the generalized eigenvalues obtained from the principal component decomposition, and $\hat{\sigma}^2$ is the corresponding minimal eigenvalue. For the approach of (5), $\hat{\mathbf{\Lambda}}_s$ is defined by

$$\hat{\mathbf{R}}_{\mathbf{YU}^{\perp}}\hat{\mathbf{E}}_s = \mathbf{\Sigma}\hat{\mathbf{E}}_s\hat{\mathbf{\Lambda}}_s ,$$

while for MOESP

$$\hat{\mathbf{R}}_{22}\hat{\mathbf{R}}_{22}^T\hat{\mathbf{E}}_s = \mathbf{\Sigma}\hat{\mathbf{E}}_s\hat{\mathbf{\Lambda}}_s .$$

Strictly speaking, this optimal weighting has only been derived for the case where the observations are uncorrelated. In the application considered here, the Hankel structure of the matrices $\mathbf{Y}$ and $\mathbf{V}$ violates this assumption. In principle, exploiting the special structure of the measurement noise matrix $\mathbf{V}$ could significantly improve estimation performance. The primary drawback of all of the subspace based methods presented in this chapter is that there is currently no available method for taking advantage of the special Hankel structure present in the data.

Although the weighting of (11) cannot necessarily be considered optimal for this application, simulation studies have demonstrated that it tends to improve performance relative to using no subspace weighting at all.

# 5. A Simulation Example

As a simple example of the advantage of the subspace fitting approach to this problem, consider the three state SISO system

defined by

$$\mathbf{A} = \begin{bmatrix} 0.30 & -0.40 & 0.60 \\ 0.00 & 0.80 & 0.00 \\ 0.00 & 0.00 & 0.79 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T$$

$$\mathbf{C} = \begin{bmatrix} 1.30 & 0.50 & -0.80 \end{bmatrix}$$

$$\mathbf{D} = 1$$

The poles of this system are located at 0.3, 0.79, and 0.8, and the zeros at 0.82, and $0.68 \pm j0.33$. Note that the lower right $2 \times 2$ block of **A** is diagonal, with nearly equal diagonal elements. Since the last two elements of **B** are identical, the second and third elements of the state vector will essentially track one another, and the state vector sequence **X** will be nearly rank deficient.

The above system was simulated in discrete time using a zero-mean, unit power, white Gaussian random process as its input. White Gaussian measurement noise with a standard deviation of 0.001 was added to the system output, resulting in an effective output signal-to-noise ratio (SNR) of roughly 60 dB. Using the noise-free input and the noisy system output, three methods were used to estimate the system poles and zeros. These were the TLS method of De Moor (essentially identical to TLS ESPRIT), the WSF algorithm presented in the previous section, and the prediction error method (PEM) of [1] based on an *output error* model for the system.

250 Monte Carlo experiments were conducted, with an independent measurement noise and input sequence generated for each trial. The block dimensions of the Hankel matrices were chosen as $i = 12$ and $j = 50$, corresponding to a total of 61 output samples for each trial. It was also assumed that the correct dimension of the state space ($n = 3$) was known. The WSF method was implemented using the weighting of (11), and assuming a diagonal parameterization for the system matrix A. Both WSF and PEM used the true parameter vectors to initialize their respective search routines. The results of the simulations are displayed in Figures 2 through 7. The solid line at the right of each plot represents the unit circle, and the true pole and zero locations are respectively indicated by the symbols '*x*' and 'o'.

The variance of the poles and zeros estimated by WSF is clearly much smaller than that of the other algorithms. This is especially true for the pole at 0.3; in fact, all 250 estimates are so closely bunched together that they are almost indistinguishable from the '*x*' marking the true pole. All three algorithms estimated the complex zeros very accurately; the individual trials are again indistinguishable from the true zero location. However, the variance of the real-valued zero is much smaller for WSF than for either PEM or TLS ESPRIT. In addition, both PEM and TLS ESPRIT estimated complex poles on about 20% of the simulations, while the WSF poles were always purely real (though they were not constrained to be so). One drawback of the subspace fitting methods is that, on a few occasions (5 for WSF and 11 for TLS ESPRIT), they produced unstable pole estimates and non-minimum phase zeros.

## References


[1] L. Ljung, *System Identification - Theory for the User*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1987.

[2] T. Söderström and P. Stoica, *System Identification*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1989.

[3] B. L. Ho and R. E. Kalman, "Efficient Construction of Linear State Variable Models from Input/Output Functions", *Regelungstechnik*, 14:545–548, 1966.

[4] S. Y. Kung, "A New Identification and Model Reduction Algorithm via Singular Value Decomposition", In *Twelfth Asilomar Conf. on Circuits, Systems and Comp.*, pages 705–714, Asilomar, CA., November 1978.

[5] B. De Moor, *Mathematical Concepts and Techniques for Modelling of Static and Dynamic Systems*, PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 1988.

[6] B. De Moor, M. Moonen, L. Vandenberghe, and J. Vandewalle, "A Geometrical Approach for the Identification of State Space Models with Singular Value Decomposition", In *Proc. IEEE ICASSP*, volume 4, pages 2244–2247, New York, NY, 1988.

[7] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle, "On- and Off-line Identification of Linear State Space Models", *Int. J. Control*, 49(1):219–232, 1989.

[8] M. Verhaegen and E. Deprettere, "Subspace Model Identification", In *Algorithms and Parallel VLSI Architectures: Volume B, Proceedings*, E. Deprettere and A. van der Veen, editors, pages 13–32. Elsevier Science Publishers B.V., The Netherlands, 1991.

[9] M. Verhaegen and P. DeWilde, "Subspace Model Identification, Part I: The Output Error State Space Model Identification Class of Algorithms", *submitted to Int. J. Control*, 1991.

[10] M. Viberg and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting", *IEEE Trans. on Sig. Proc.*, 39(5):1110–1121, May 1991.

[11] M. Viberg, B. Ottersten, and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting", *IEEE Trans. on Sig. Proc.*, 39(11):2436–2449, Nov. 1991.

[12] A. Paulraj, R. Roy, and T. Kailath, "A Subspace Rotation Approach to Signal Parameter Estimation", *Proceedings of the IEEE*, pages 1044–1045, July 1986.

[13] R. Roy and T. Kailath, "ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques", *IEEE Trans. on ASSP*, 37(7):984–995, July 1989.

[14] S.Y. Kung, C.K. Lo, and R. Foka, "A Toeplitz Approximation Approach to Coherent Source Direction Finding", In *Proc. ICASSP 86*, Tokyo, Japan, 1986.

[15] M. Viberg, B. Ottersten, B. Wahlberg, and L. Ljung, "A Statistical Perspective on State-Space Modeling Using Subspace Methods", In *Proc. 30th CDC Conf.*, Brighton, England, 1991.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD., 1984.

[17] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple Invariance ESPRIT", *IEEE Trans. on Sig. Proc.*, 40, April 1992.

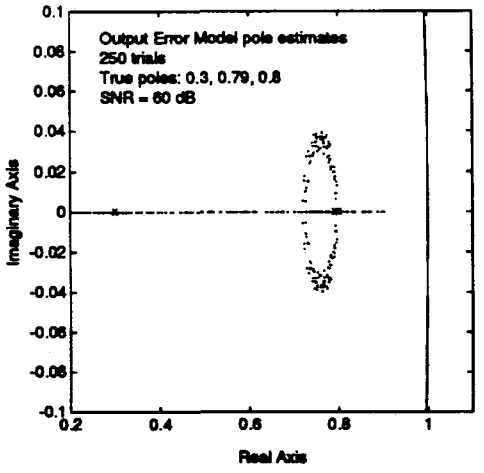


*Figure 2: Pole Estimates for Output Error Model*

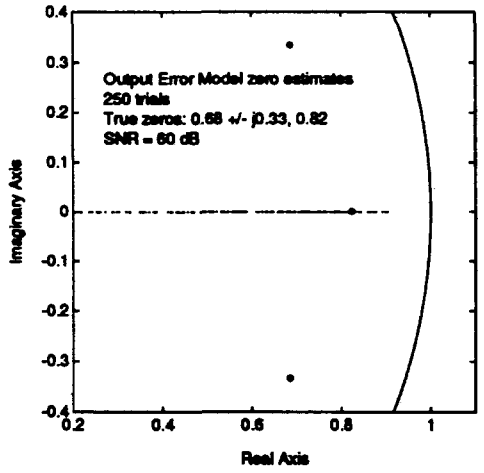


*Figure 3: Zero Estimates for Output Error Model*

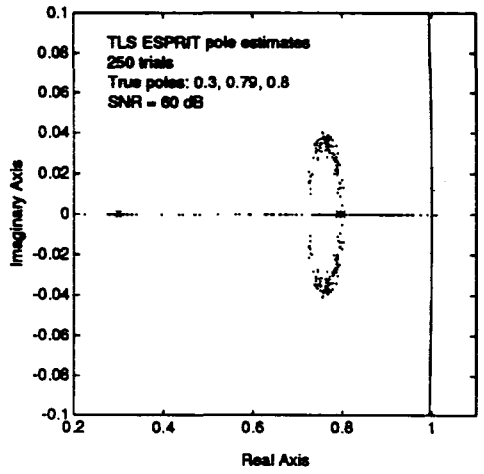


*Figure 4: Pole Estimates for TLS ESPRIT*

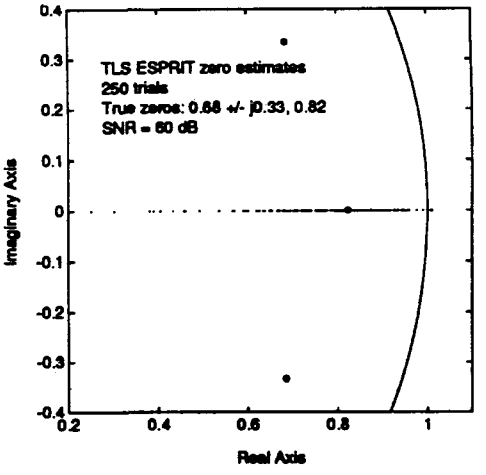


*Figure 5: Zero Estimates for TLS ESPRIT*

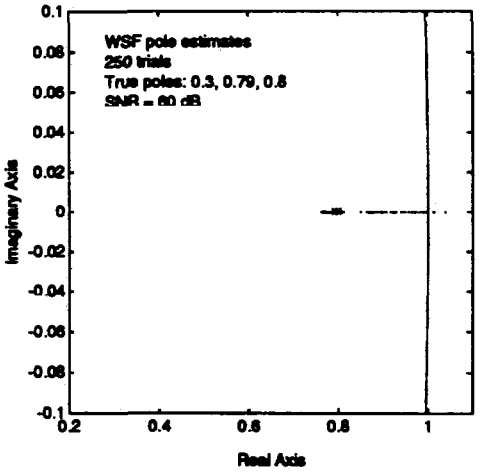


*Figure 6: Pole Estimates for WSF*

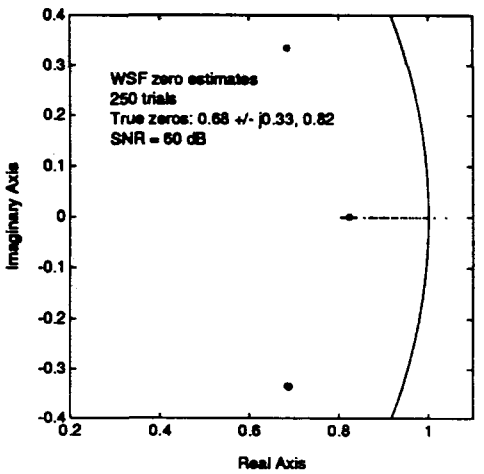


*Figure 7: Zero Estimates for WSF*