- [6] —, "An optimal control depending on the conditional density of the unobserved state," in *Applied Stochastic Analysis, Lecture Notes in Control and Information Sciences* 177, I. Karatzas and D. Ocone, Eds. New York: Springer-Verlag, 1992.
- [7] R. L. Moose et al., "Modeling and estimation for tracking maneuvering targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, pp. 448–455, May 1979.
- [8] H. L. Pastrick, S. M. Seltzer, and M. E. Warren, "Guidance laws for short-range tactical missiles," *J. Guid. Contr.*, vol. 4, no. 2, pp. 98–108, 1981.
- [9] E. Y. Rodin, "A pursuit evasion bibliography version 2," *Comput. Math. Appl.*, vol. 18, no. 1–3, pp. 245–320, 1989.
- [10] A. S. Willsky, "Detection of abrupt changes in dynamical systems," in *Lecture Notes in Control and Information Sciences*, vol. 77, *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, Eds. Berlin: Springer-Verlag, 1986.

A Subspace Fitting Method for Identification of Linear State-Space Models

A. Swindlehurst, R. Roy, B. Ottersten, and T. Kailath

Abstract—A new method is presented for the identification of systems parameterized by linear state-space models. The method relies on the concept of *subspace fitting*, wherein an input/output data model parameterized by the state matrices is found that best fits, in the least-squares sense, the dominant subspace of the measured data. Some empirical results are included to illustrate the performance advantage of the algorithm compared to standard techniques.

I. INTRODUCTION

Research in the identification of discrete-time linear systems has focused in recent years on prediction error methods (PEM's) based on autoregressive and autoregressive moving-average (ARMA) data models and their various derivatives (e.g., [1], [2]). Although linear state-space models are common in estimation and control, they have not been widely used in identification. Implicitly, of course, a given input/output model can be associated with various canonical state-space realizations, and thus a PEM might be thought of as finding a state-space model for the system. One of the goals of this note 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 [3]–[5]. More recently, De Moor [6], [7] 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.

Manuscript received May 21, 1993; revised February 1, 1994. This work was supported by the National Science Foundation under Grant MIP-9110112.

A. Swindlehurst is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602 USA.

R. Roy is with ArrayComm, Inc., Santa Clara, CA 95054 USA.

B. Ottersten is with the Department of Signals, Sensors, and Systems, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

T. Kailath is with the Information Systems Lab, Stanford University, Stanford, CA 94305 USA.

IEEE Log Number 9406993.

Related methods have also been recently proposed by Moonen *et al.* [8] and Verhaegen *et al.* [9].

The method presented in this note also exploits the inherent shift structure in the data, but in a different way. The motivation for this new algorithm comes from some recent results in 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 estimation [10]. Of special interest is the fact that a weighting can be applied to the dominant subspace to emphasize certain directions where the signal-to-noise ratio is high. This weighting can provide an advantage in cases involving nearly unobservable systems or an insufficiently excited state space.

II. DATA MODEL AND ASSUMPTIONS

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

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{u}_k \in \mathbb{R}^m$, $\mathbf{y}_k \in \mathbb{R}^l$, $\mathbf{v}_k \in \mathbb{R}^l$, and the system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are of consistent dimension. The system input $\mathbf{u}_k$ is assumed known and the output, $\mathbf{y}_k$, is corrupted by additive measurement noise, $\mathbf{v}_k$. If several observations of the input and output vectors are available, they may be grouped together into the single equation (e.g., see [6] and [11])

$$\mathbf{Y} = \mathbf{\Gamma}\mathbf{X} + \mathbf{H}\mathbf{U} + \mathbf{V} \quad (2)$$

where

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

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

$$\mathbf{H} = \begin{bmatrix} \mathbf{D} & 0 & 0 & \cdots & 0 \\ \mathbf{C}\mathbf{B} & \mathbf{D} & 0 & \cdots & 0 \\ \mathbf{C}\mathbf{A}\mathbf{B} & \mathbf{C}\mathbf{B} & \mathbf{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}\mathbf{A}^{i-2}\mathbf{B} & \mathbf{C}\mathbf{A}^{i-3}\mathbf{B} & \mathbf{C}\mathbf{A}^{i-4}\mathbf{B} & \cdots & \mathbf{D} \end{bmatrix}$$

$\mathbf{x} = [\mathbf{x}_k \mathbf{x}_{k+1} \cdots \mathbf{x}_{k+j-1}]$, and where $\mathbf{U}(mi \times j)$ and $\mathbf{V}(li \times j)$ are block Hankel matrices constructed exactly as $\mathbf{Y}$, but containing samples of the input and disturbance sequences, respectively.

We will make the following assumptions concerning the data model of (2):

- The system is observable, and the block dimension i is chosen to be large enough so that $\text{rank}(\mathbf{\Gamma}) = n$. This implies $li \geq n$.
- The input sequence $\mathbf{u}_k$ has *sufficiently excited* the system and $j \geq n$, so that $\text{rank}(\mathbf{X}) = n$.

- The input is uncorrelated with the measurement noise (open-loop operation).
- A nontrivial matrix $U^\perp$ can be found to satisfy $UU^\perp = 0$. This requires $j > mi$, and leads to

$$YU^\perp = \Gamma XU^\perp + VU^\perp. \quad (3)$$

- rank $(XU^\perp) = n$, so that $j \geq mi + n$. This rank condition is satisfied for most choices of the input (e.g., with probability one if the input is zero-mean white noise) [6], [11].

With these assumptions, $YU^\perp$ is an $li \times p$ matrix, with $n \leq p \leq j - mi$. In most situations, $p = j - mi > n$, so that $YU^\perp$ is composed of a (low) rank n "signal" term $\Gamma XU^\perp$ that depends on the parameters of interest, and a "noise" term $VU^\perp$ that is (with probability one) full rank.

It is convenient for us to also define here the "sample covariance" matrix

$$\hat{R}_{YU^\perp} = \frac{1}{j}(YU^\perp)(YU^\perp)^T$$

and note that it converges to its limiting expected value with probability one

$$R_{YU^\perp} = \lim_{j \rightarrow \infty} \mathcal{E} \{ \hat{R}_{YU^\perp} \} = \Gamma R_{XU^\perp} \Gamma^T + \sigma^2 \Sigma$$

where

$$R_{XU^\perp} = \frac{1}{j}(XU^\perp)(XU^\perp)^T, \quad \sigma^2 \Sigma = \frac{1}{j} \mathcal{E} \{ (VU^\perp)(VU^\perp)^T \}$$

and Σ is normalized so that $\det(\Sigma) = 1$. It is easy to show that the generalized eigenvalues $\{\lambda_k\}$ of the pair $(R_{YU^\perp}, \Sigma)$ satisfy $\lambda_1 \geq \dots \geq \lambda_n > \lambda_{n+1} = \dots = \lambda_l = \sigma^2$. Furthermore, the eigenvectors $E = [e_1 \dots e_n]$ associated with $\lambda_1, \dots, \lambda_n$ satisfy

$$\Sigma E = \Gamma(\eta)T \quad (4)$$

for some $n \times n$ matrix T . We have written Γ as a function of a vector η , which contains the elements of the parameterization chosen for A and C . Just as there are many realizations or coordinate systems that can be used to describe the state space, there are many identifiable parameterizations η that can be chosen, each yielding a different T that satisfies (4). The subspace fitting method presented in this note is based on the relationship of (4).

Note that with no measurement noise or infinite data, the model order n is revealed by simply counting how many of the smallest generalized eigenvalues are equal. With a finite collection of noisy data, a statistical test is required to estimate this quantity. This problem has been extensively studied in very general contexts, and many such tests have been developed (e.g., see [12]–[14]). We will thus assume throughout the remainder of the note that n has been correctly determined.

III. A SUBSPACE FITTING APPROACH

Because of the effects of noise, only an estimate of the generalized eigenvectors $\hat{E}$ can be obtained from $\hat{R}_{YU^\perp}$ (or equivalently from an SVD of $YU^\perp$). Consequently, assuming an appropriate identifiable parameterization η has been chosen, we propose the following two-step identification procedure based on (4).

1) Estimate η by means of the following weighted least-squares problem:

$$\begin{aligned} \hat{\eta} &= \arg \min_{\eta} \| \Sigma \hat{E} W^{1/2} - \Gamma(\eta)T \|_F^2 \\ &= \arg \min_{\eta} \text{Tr} (P_{\hat{\eta}}^\perp \Sigma \hat{E} W E^T \Sigma^T) \end{aligned} \quad (5)$$

where $P_{\hat{\eta}}^\perp = I - \Gamma(\eta)[\Gamma(\eta)^T \Gamma(\eta)]^{-1} \Gamma(\eta)^T$. The estimates of the system matrices are thus given by $A(\hat{\eta}), C(\hat{\eta})$.

2) Estimate B and C using a noise-free version of (2) and the estimates $A(\hat{\eta}), C(\hat{\eta})$.

Some comments are in order concerning the two steps of the Weighted Subspace Fitting (WSF) algorithm described above.

Step 1: The choice of the weighting matrix W will be discussed below in Subsection A. A variety of possible parameterizations for η exist, depending on what (if any) prior information about the system is available. For example, if the system is multiple-input single-output, a particularly appropriate choice is to assume that the system is in observer form, and hence that A is left-companion and $C = [1 \ 0 \dots 0]$. Alternatively, A could be assumed to be diagonalizable, in which case C is solved for directly, and the minimization of (5) can be simplified to involve only a search over the n pole locations.

One of the advantages of the problem formulation considered herein is that any *a priori* information about the structure of A and C can be directly incorporated into the problem. In some instances, the dynamical model underlying the system may be well understood, but may depend on certain parameters that are imprecisely known. These parameters appear as unknown constants in the matrices A and C of the system model, and could serve as the parameters in the minimization of (5) just as easily as the poles or the coefficients of the characteristic polynomial. Whether or not the set of unknown parameters can be uniquely identified is then problem dependent, and must be determined on a case-by-case basis.

The minimization of (5) is identical in form to a problem in antenna array signal processing considered in [15]. In fact, once an appropriate parameter vector has been identified, an algorithm essentially identical to that described in [15] may be used to estimate A and C . Additional information on the connection between state-space system identification and array signal processing can be found in [16] and [17].

We note here that, even though $\Sigma \neq I$, consistent parameter estimates can be obtained by setting $\Sigma = I$ if the noise is white and independent with equal variance among the l outputs, and if j is sufficiently large.¹ In other cases, however, ignoring Σ when $\Sigma \neq I$ (as is done in [6], [7], [9]) can lead to biased parameter estimates.

Step 2: This step is also used by De Moor in his identification method [6], [7]. A description of the mechanics required to solve for B and D can also be found in [19], and will not be given here.

A. Subspace Weighting

In simple terms, the presence of the weighting matrix W in the minimization of (5) allows certain directions or dimensions of the low rank subspace to be emphasized over others. For example, the generalized eigenvector e_1 corresponding to the largest eigenvalue represents the component of the measured data where the signal energy is the strongest. The eigenvector e_n gives the direction where the signal energy is weakest but still nonzero. A measure of how "strong" or "weak" the signal energy is in a given direction can be based on how much bigger than σ^2 the eigenvalues $\lambda_1, \dots, \lambda_n$ are. The difference $\lambda_n - \sigma^2$ can be quite small in cases where Γ or X are ill-conditioned (e.g., because the system is nearly unobservable or the input is not sufficiently exciting), and in such cases one would probably wish to give much less weight to e_n in (5) than e_1 .

The following "optimal" weighting has been derived for subspace fitting problems of the form (5) in [10]:

$$W_{WSF} = (\hat{\Lambda}_s - \sigma^2 I)^2 \hat{\Lambda}_s^{-1} \quad (6)$$

¹ How large j has to be for this to hold depends on the magnitude of the noise.

Fig. 1. Pole and zero estimates for simulation example 1.

where the diagonal matrix $\hat{\Lambda}_s$ contains $\lambda_1, \dots, \lambda_n$ as its diagonal elements, and $\hat{\sigma}^2$ is an appropriate estimate of σ^2 (obtained, for example, as an average of the $li-n$ smallest generalized eigenvalues). The term "optimal" here means that, under certain conditions, the weighting of (6) will lead to minimum variance parameter estimates. The Hankel structure of Y and V in our problem violates one of these conditions, namely, that the columns of the data matrix be independent. However, using W_{WSF} in (5) has been empirically shown to yield better results than using no weighting at all ($W = I$).

B. Comparison to Other Techniques

The principle difference between the WSF algorithm presented above and the methods described in [6], [7], and [9] is that the WSF approach can exploit the complete structure of F rather than just a single "shift-invariance." This advantage is especially evident in situations where *a priori* information is available about the structure of A and C . On the other hand, the methods of [6], [7], and [9] admit a simple "closed-form" solution, whereas (5) can only be solved by means of a multidimensional search.

Fig. 1. Continued.

Unlike standard ARMA-based PEM approaches, the WSF algorithm is just as easily applied to MIMO systems as in the single-input single-output case, has a built-in mechanism for model order determination, and can exploit structured system models with more compact parameterizations. Even in the unstructured case, WSF tends to yield pole and zero estimates with much lower variance than PEM's, as illustrated by the simulation examples of the next section. Recent work in identification and controller design for rapid thermal processing of semiconductor wafers [19], [20] has also borne out some of the advantages of using subspace-based identification methods over PEM's.

IV. SIMULATION EXAMPLES

In this section, we consider two simple simulation examples that illustrate the advantage of the subspace fitting approach. In the first example, the parameters of the following three state SISO discrete time system will be estimated:

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

$$B = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$C = [1.30 \quad 0.40 \quad -0.80]$$

$$D = 1.$$

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 for this system, X will be ill-conditioned since the lower right 2×2 block of A is diagonal, with nearly equal diagonal elements, and the last two elements of B are identical. This is also evidenced by the fact that there is a near pole-zero cancellation in the transfer function.

This system was simulated using a zero-mean unit power, white Gaussian process as its input, and assuming a zero initial state.² White Gaussian measurement noise with a standard deviation of 0.001 was added to the system output. Using the noise-free input and the noisy system output, three methods were used to estimate the system poles and zeros. These were the method of De Moor [6], [7] (which, for brevity, we refer to as the S4ID technique), the WSF algorithm, and a PEM based on an *output error* model [1]. The correct model order was assumed to be known in each case.

We conducted 250 Monte Carlo experiments, 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. The WSF method was implemented using the weighting of (6), 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 Fig. 1. The solid line at the right of each plot represents the unit circle, and the true pole and zero locations are indicated by the symbols $\times$ and $\circ$, respectively.

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 $\times$ marking the true pole. All three algorithms estimated the complex zeros very accurately; the individual trials are again indistinguishable from the true zero locations. However, the variance of the real-valued zero is much smaller for WSF than for either PEM or S4ID. In addition, both PEM and S4ID estimated complex poles on about 20% of the simulations, while the WSF poles were always purely real (although they were not constrained to be so). One drawback of the subspace-based methods is that, on a few occasions (5 for WSF and 11 for S4ID), they produced unstable pole estimates.

²The effects of a nonzero initial state can be handled by assuming the presence of an additional input that is a unit Dirac impulse [6].

Fig. 2. Pole estimates for output error model, example 2.

Fig. 3. Pole estimates for WSF, example 2.

In the second example, we consider the following system used in the simulation studies of [9]

$$A = \begin{bmatrix} 1.92 & -0.9316 \\ 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$C = [0.05 \quad 0.025] \\ D = 0.$$

The poles of this system are at $0.96 \pm j0.01$, and there is a single zero at -0.5 . The same type of input and measurement disturbance as above were used in this case, except the variance of the disturbance was increased to one. This resulted in an average effective output SNR of only 4.4 dB. The dimensions of the block Hankel matrices were chosen as $i = 40$ and $j = 161$, for a total of 200 samples per trial. As before, both WSF and the output error PEM were initialized with the true parameters, and 250 independent trials were conducted. The results of the simulation are plotted in Figs. 2 and 3.

Fig. 2 shows the pole estimates for the output error model, with the true pole locations indicated by the crosshairs, and Fig. 3 shows

the same for the WSF algorithm. No separate plot is shown for S4ID, since it gave results essentially equivalent to WSF in this case (with the exception of one outlier). The WSF approach again yields pole estimates with a much smaller variance than the corresponding PEM. The average mean-square prediction error for WSF in this case was 0.985, compared to 1.403 for the output error PEM.

Because of the various optimality properties of PEM's in general [1], one may be surprised by the superior performance of WSF in the above two examples. Since both algorithms used essentially the same minimization procedure (Gauss-Newton iterations with identical initial conditions and termination criteria), the relatively poor performance of the output error PEM in these examples may be attributed to a propensity for convergence to local minima.

V. CONCLUSIONS

We have presented a new method for identification of linear systems parameterized by state-space models. The method is based on the notion of weighted subspace-fitting (WSF), a problem-solving philosophy most recently applied to parameter estimation problems in sensor array signal processing. Like other state-space identification schemes, WSF is easily applied in the MIMO case, allows for direct incorporation of *a priori* physical constraints into the problem, and appears to have better numerical properties than methods based on input/output models. In addition, the ability to appropriately weight the signal subspace vectors used in the WSF minimization provides a degree of robustness over previous subspace-based methods when the system is nearly unobservable or not sufficiently excited. Although implementation of WSF requires a multidimensional parameter search, accurate initial conditions for the search are readily obtained.

REFERENCES

- [1] L. Ljung, *System Identification—Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [2] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [3] B. L. Ho and R. E. Kalman, "Efficient construction of linear state variable models from input/output functions," *Regelungstechnik*, vol. 14, pp. 545–548, 1966.
- [4] H. P. Zeiger and A. J. McEwen, "Approximate linear realization of given dimension via Ho's algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-19, p. 153, 1974.
- [5] S. Y. Kung, "A new identification and model reduction algorithm via singular value decomposition," In *Proc. 12th Asilomar Conf. Circuits, Syst., Comp.*, Asilomar, CA, Nov. 1978, pp. 705–714.
- [6] B. De Moor, "Mathematical concepts and techniques for modelling of static and dynamic systems," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, 1988.
- [7] 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*, vol. 4, New York, 1988, pp. 2244–2247.
- [8] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle, "On- and off-line identification of linear state space models," *Int. J. Contr.*, vol. 49, no. 1, pp. 219–232, 1989.
- [9] M. Verhaegen and P. DeWilde, "Subspace model identification, Parts 1–2," *Int. J. Contr.*, vol. 56 no. 5, pp. 1187–1241, 1992.
- [10] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.
- [11] B. Gopinath, "On the identification of linear time-invariant systems from input-output data," *Bell Syst. Tech. J.*, vol. 48, pp. 1101–1113, 1969.
- [12] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, pp. 122–148, 1963.
- [13] H. Akaike, "A new look at statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.
- [14] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.
- [15] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Processing*, vol. 40, pp. 867–881, Apr. 1992.
- [16] A. Swindlehurst, R. Roy, B. Ottersten, and T. Kailath, "System identification via weighted subspace fitting," in *Proc. Amer. Contr. Conf.*, June 1992, pp. 2158–2163.
- [17] A. van der Veen, E. Deprettere, and A. Swindlehurst, "Subspace based signal analysis using singular value decomposition," *Proc. IEEE*, vol. 81, pp. 1277–1308, Sept. 1993.
- [18] 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.
- [19] P. Gyugyi, Y. M. Cho, G. Franklin, T. Kailath, and R. Roy, "A model-based control of rapid thermal processing systems," in *Proc. 1st IEEE Conf. Contr. Appl.*, Dayton, OH, Sept. 1992.
- [20] P. Gyugyi, Y. M. Cho, G. Franklin, and T. Kailath, "Control of wafer temperature in rapid thermal processing: Part I—State space model identification and control," submitted to *Automatica*, 1992.

Consistency of Modified LS Estimation Method for Identifying 2-D Noncausal SAR Model Parameters

Ping-Ya Zhao and John Litva

Abstract—Least squares (LS) and maximum likelihood (ML) are the two main methods for parameter estimation of two-dimensional (2-D) noncausal simultaneous autoregressive (SAR) models. ML is asymptotically consistent and unbiased but computationally unattractive. On the other hand, conventional LS is computationally efficient but does not produce accurate parameter estimates for noncausal models. Recently, in [17], a modified LS estimation method was proposed and shown to be unbiased. In this note we prove that, under certain assumptions, the method introduced in [17] is also consistent.

I. INTRODUCTION

Two-dimensional (2-D) signal and system modeling and parameter estimation have many applications such as 2-D Kalman filtering [1], image estimation and identification [2]–[4], image restoration [5], [6], multidimensional spectrum estimation [7], [8], direction finding [9], texture analysis and synthesis [10], [11], multidimensional system identification [12], [13], etc. Several kinds of models are used, to cite a few: Gauss-Markov random field (GMRF) models which are sometimes called conditional Markov (CM) models [14]–[16], simultaneous autoregressive (SAR) models [16], [17], autoregressive moving average (ARMA) models [18], [19], etc. In this note, we will concentrate on the problem of estimating the parameters $\{g_{k,l}\}$ of the 2-D noncausal SAR model in the form of

$$y(i, j) = - \sum_{(k,l) \in N} g_{k,l} y(i-k, j-l) + e(i, j) \quad (1)$$

Manuscript received June 18, 1993; revised February 20, 1994.
The authors are with Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada, L8S 4K1.
IEEE Log Number 9406994.