Theorem 1: If the spectral uncertainty classes $\mathcal{S}$ and $\mathfrak{N}$ are such that the following two conditions hold:

i) $\sup_{\mathcal{S}} 1/2\pi \int_{-\infty}^{\infty} \sigma(\omega)\, d\lambda(\omega) \le w_s < \infty$

ii) Let $C = \{\omega | \sigma(\omega) > 0 \text{ or } \nu(\omega) > 0, \text{ for some } \sigma \in \mathcal{S} \text{ or } \nu \in \mathfrak{N}\}$, then there is a pair of PSD's $(\bar{\sigma}, \bar{\nu}) \in \mathcal{S} \times \mathfrak{N}$ such that $\bar{\sigma}(\omega) + \bar{\nu}(\omega) > 0$ a.e. on C and $\int_{-\infty}^{\infty} (\bar{\sigma}(\omega) + \bar{\nu}(\omega))\, d\lambda(\omega) < \infty$;

then there exists H_R^+ achieving the infimum in (5). Furthermore, this infimum is finite.

Proof: A careful examination of the proof of Theorem 1 in [3] (or the version in [4]) shows that it does not depend on the restriction in the frequency domain to $[-\pi, \pi]$. That is, if the signal and noise have finite power, then the spectral densities can be taken with respect to any finite Borel measure μ on $\mathbb{R}$. Since the entire proof takes place in the frequency domain, condition ii) above allows us to reformulate the problem (5) in terms of μ and then apply the results of [3]. Specifically, let μ be a finite Borel measure on $\mathbb{R}$ having $\bar{\sigma} + \bar{\nu}$ as its density (Radon-Nikodym derivative) with respect to λ, i.e., let

$$\mu(A) = \int_A [\bar{\sigma}(\omega) + \bar{\nu}(\omega)]\, d\lambda(\omega)$$

for every Borel set A. If we let

$$f_S(\omega) = \frac{\sigma(\omega)}{\bar{\sigma}(\omega) + \bar{\nu}(\omega)}$$

and

$$f_N(\omega) = \frac{\nu(\omega)}{\bar{\sigma}(\omega) + \bar{\nu}(\omega)} \tag{6}$$

then f_S and f_N are the PSD's, with respect to μ, corresponding to σ and ν for any $\sigma \in \mathcal{S}$ and $\nu \in \mathfrak{N}$. So, for example,

$$\int_A f_S(\omega)\, d\mu(\omega) = \int_A \sigma(\omega)\, d\lambda(\omega)$$

for any A. Note that f_S and f_N are finite (a.e. μ or λ) because of condition ii). We can now apply Theorem 1 in [3], [4] to the new classes of f_S's and f_N's defined from the elements of $\mathcal{S}$ and $\mathfrak{N}$ via (6) to yield the existence of a most robust causal transfer function H_R^+. This concludes the proof of Theorem 1.

We now turn our attention to the problem of finding H_R^+, the most robust transfer function. We begin with a definition.

Definition 1: A pair of PSD's (σ_L, ν_L) is *least favorable for causal estimation* for the uncertainty classes $\mathcal{S}$ and $\mathfrak{N}$ if

$$e_D^+(\sigma_L, \nu_L) = \max_{\mathcal{S}\times\mathfrak{N}} e_D^+(\sigma, \nu) \tag{7}$$

where $e_D^+(\sigma, \nu)$, the minimum-MSE for (σ, ν), is defined in (4).

Note that (7) means that (σ_L, ν_L) solves the *maximin* game

$$\max_{\mathcal{S}\times\mathfrak{N}} \min_{\mathcal{K}} e_D(\sigma, \nu; H). \tag{8}$$

Hence, if the minimax equality holds here [i.e., if (8) equals (5)] then (σ_L, ν_L) is a least favorable pair if and only if (σ_L, ν_L) and its optimal causal transfer function H_L^+ form a saddle point solution to the game (5) [or equivalently, the game (8)], that is (σ_L, ν_L) and H_L^+ satisfy

$$e_D(\sigma, \nu; H_L^+) \le e_D(\sigma_L, \nu_L; H_L^+) \le e_D(\sigma_L, \nu_L; H). \tag{9}$$

Clearly, if (9) holds then H_L^+ is a most robust causal transfer function.

Our next theorem gives conditions under which the optimal transfer function for a least favorable pair is most robust. This is useful because it is often easier to solve the maximization problem (7) than it is to solve the minimax game (5).

Theorem 2: If the spectral uncertainty classes $\mathcal{S}$ and $\mathfrak{N}$ are such that conditions i) and ii) of Theorem 1 hold and, furthermore, $\mathcal{S}$ and $\mathfrak{N}$ are convex, then a pair of PSD's (σ_L, ν_L) in $\mathcal{S} \times \mathfrak{N}$ and its optimal causal transfer function H_L^+ form a saddle point solution to the minimax game (5) if and only if (σ_L, ν_L) is least favorable for causal estimation, i.e., solves (7) (note that in this case H_L^+ is a most robust causal transfer function).

The proof follows from that of Theorem 2 in [3], [4] via an argument similar to Theorem 1 above.

IV. Conclusion

We have shown for the continuous time, linear, signal estimation problems of prediction, filtering, and smoothing that the minimax robust estimator is just the optimal estimator for a least favorable signal and noise spectral pair. As in [2]–[4], this result greatly simplifies the process of obtaining robust estimators in those cases where an expression is available for the minimum mean square error (see [6]–[11] for examples and methodologies).

References

[1] K. S. Vastola and H. V. Poor, "An analysis of the effects of spectral uncertainty on Wiener filtering," *Automatica*, vol. 19, pp. 289–293, May 1983.

[2] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 531–536, June 1980.

[3] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 316–327, Mar. 1984.

[4] J. Franke and H. V. Poor, "Minimax-robust filtering and finite-length robust predictors," in *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Härdle, and R. D. Martin, Eds. New York: Springer-Verlag, 1984, pp. 87–126.

[5] K. S. Vastola, "Topics in robust statistical signal processing," Ph.D. thesis, Dep. Elec. Eng., Univ. Illinois, 1982; (also, Coordinated Sci. Lab. Tech. Rep. R-965).

[6] J. Snyders, "Error expressions for optimal linear filtering of stationary processes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 574–582, 1972.

[7] J. Snyders, "On optimal linear estimation of signals with general spectral distribution," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 654–658, 1974.

[8] K. Yao, "An alternative approach to the linear casual least-square filtering theory," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 232–240, 1971.

[9] K. Yao, "On the direct calculation of mmse of linear realizable estimator by Toeplitz form method," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 95–97, 1971.

[10] D. L. Snyder, "Some useful expressions for optimal linear filtering in white noise," *Proc. IEEE*, vol. 53, pp. 629–630, 1965.

[11] A. J. Viterbi, "On the minimum mean square error resulting from linear filtering of stationary signals in white noise," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 594–595, 1965.

Identification of Multivariable Errors in Variable Models with Dynamics

MICHAEL GREEN AND BRIAN D. O. ANDERSON

Abstract—**This note considers the task of identifying a causal, linear, dynamic, multivariable system excited by stationary, zero-mean noise of unknown spectrum, and given measurements of the system inputs and outputs contaminated by independent, additive noise also of unknown spectra. Although the solution is in general not unique, finite-dimensional parameterizations of the solution set are given, even though the various spectra may not be rational.**

I. Introduction

Consider the problem of identifying a linear, time-invariant, dynamic, multivariable system given noisy measurements of it. In contrast to the common situation, the input as well as the output is contaminated with an unknown amount of noise.

More specifically, we postulate the existence of three random vector

Manuscript received February 25, 1985; revised August 9, 1985. Paper recommended by Associate Editor, H. L. Weinert.

The authors are with the Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, Australia.

IEEE Log Number 8406542.

0018-9286/86/0500-0467$01.00

sequences $\{\hat{x}_k\}$, $\{u_k\}$, $\{v_k\}$ of the same dimension n, mutually independent and stationary, together with a time-invariant, linear, multivariable, system defined by a bounded, linear, causal, convolution operator $\{W_k, k \geqslant 0\}$ mapping $\{\hat{x}_k\}$ into a vector sequence $\{\hat{y}_k\}$ also of dimension n according to

$$\hat{y}_k = \sum_{-\infty}^{k} W_{k-l}\hat{x}_l. \tag{1.1}$$

The processes $\{\hat{x}_k\}$, $\{\hat{y}_k\}$ are not available for measurement, but rather we can measure, for $k \in (\infty, \infty)$

$$x_k = \hat{x}_k + u_k \tag{1.2a}$$

$$y_k = \hat{y}_k + v_k. \tag{1.2b}$$

Our concern is not to identify a particular $\{W_k\}$, nor to give conditions for a unique solution to exist, but to characterize the class of $\{W_k\}$ which fit the data. For scalar systems ($n = 1$), this approach in the nondynamic case goes back to [1]–[3]. The scalar dynamic case has been treated in [4] and [5]. Although we state the problem for square systems, the nonsquare rational case can be reduced to the square case by pre- or postmultiplication by unimodular matrices (i.e., row/column operations). This possibility is explored after the square problem is solved in Section IV.

Since it is the aim of this note to extend the results of [5] to multivariable systems, we now briefly review the main results of [5].

Let us recall first the following static result (see, e.g., [6], [7]). Suppose (1.2) holds and

$$\hat{y}_k = w\hat{x}_k \tag{1.3}$$

with w a real scalar to be identified, and $\{\hat{x}_k\}$, $\{u_k,\}$ $\{v_k\}$ are discrete-time, zero mean, white noise Gaussian processes. We are given the matrix

$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} = E\left\{ \begin{bmatrix} x_k \\ y_k \end{bmatrix} [x_k \quad y_k] \right\} \tag{1.4}$$

and we assume that $E[\hat{x}_k^2] > 0$. The range of possible w is

$$\left[\frac{\sigma_{xy}}{\sigma_{xx}}, \frac{\sigma_{yy}}{\sigma_{xy}}\right] \quad \text{if } \sigma_{xy} > 0 \tag{1.5a}$$

$$\left[\frac{\sigma_{yy}}{\sigma_{xy}}, \frac{\sigma_{xy}}{\sigma_{xx}}\right] \quad \text{if } \sigma_{xy} < 0 \tag{1.5b}$$

$$0 \quad \text{if } \sigma_{xy} = 0. \tag{1.5c}$$

For the dynamic case, let $[\begin{smallmatrix} \sigma_{xx}(\omega) & \sigma_{xy}(\omega) \\ \sigma_{yx}(\omega) & \sigma_{yy}(\omega) \end{smallmatrix}]$ be the power spectrum matrix of $[x\ y]'$, and $w(z)$ the transfer function from $\hat{x}$ to $\hat{y}$, with z denoting the delay operator, as in [5]. Under certain reasonable assumptions, called the standing assumptions, a similar result is obtained in [5], namely

$$\arg\left[\frac{\sigma_{xy}(\omega)}{\sigma_{xx}(\omega)}\right] = \arg\,[w(e^{j\omega})] = \arg\left[\frac{\sigma_{yy}(\omega)}{\sigma_{yx}(\omega)}\right] \tag{1.6a}$$

and

$$\left|\frac{\sigma_{xy}(\omega)}{\sigma_{xx}(\omega)}\right| \leqslant |w(e^{j\omega})| \leqslant \left|\frac{\sigma_{yy}(\omega)}{\sigma_{yx}(\omega)}\right|. \tag{1.6b}$$

The remainder of [5] is devoted to constructing $w(e^{j\omega})$ satisfying (1.6) and parameterizing the solution set of (1.6). The key idea is to construct the magnitude of $w(e^{j\omega})$ from the phase, which is known from (1.6a), by using a formula from analytic function theory relating the real and imaginary parts of analytic functions. The principle of the argument allows one to determine the number, N say, of unstable (nonminimum phase) zeros of $w(z)$. The solution set is then shown to be an $N + 1$ parameter family. Indeed, N of the parameters are just the positions of the unstable zeros of $w(z)$, and the remaining parameter is a scaling constant which must be chosen to satisfy (1.6b). (Note that this may not be possible for an arbitrary choice of the zeros of $w(z)$, but for at least one choice, a suitable scaling constant must exist.) Thus, in the minimum-phase (no unstable zeros) case, or if the unstable zeros are given, the solution is as for the static case, i.e., uniquely determined up to a scaling constant confined to a finite interval.

In the multivariable case it is still possible to obtain formulas analogous to (1.6) when $\Sigma_{xx}(\omega) > 0\ \forall\ \omega$, namely

$$W(e^{j\omega}) = \Sigma_{yx}(\omega)(\Sigma_{xx}(\omega) - A(\omega))^{-1} \tag{1.7a}$$

where $A(\omega)$ is an arbitrary Hermitian matrix valued function of ω satisfying

$$\Sigma_{xx}(\omega) - \Sigma_{xy}(\omega)\,\Sigma_{yy}^{\#}(\omega)\,\Sigma_{yx}(\omega) \geqslant A(\omega) \geqslant 0 \tag{1.7b}$$

and $\Sigma_{yy}^{\#}(\omega)$ denotes the Moore–Penrose pseudoinverse of Σ_{yy} [5].

However, this does not ensure that W is causal and there are major difficulties involved in applying the scalar solution technique just outlined to solving (1.7). The scalar technique proceeds from phase information, but what is the phase of a matrix? Second, given some definition of the phase of a matrix, is it possible to reconstruct the complete matrix from knowledge of the phase, perhaps under a minimum-phase assumption? Even if one knew the phase of every entry of W, one cannot determine the zero structure of the entries from the zero structure of W (e.g., if W is known to be minimum phase). Thus, an entry-by-entry solution is not possible.

The solution technique presented in this note is based on the factorization of matrix valued functions, a special case of which is the better known spectral factorization. The factorization theory is well developed and is used extensively in the theory of integral equations. After formally stating the problem, our assumptions, and notation, we will introduce this factorization theory, and hence proceed to solve the errors-in-variables identification problem at hand.

II. Formal Problem Statement

We now introduce some basic assumptions on the errors-in-variables (EIV) system, as well as some notation, which will apply throughout this note.

The vector notation used is standard, with all vectors being complex column vectors. If A is a matrix or vector, A^* will denote the Hermitian conjugate of A (i.e., complex conjugate transpose). A matrix A is Hermitian if $A^* = A$. A matrix function $A(\omega)$, $0 \leqslant \omega < 2\pi$ will be called positive (nonnegative) if the quadratic form $x^*A(\omega)x$, where x is an arbitrary nonzero vector, has only real positive (nonnegative) values. The notation $A(\omega) > B(\omega)$ ($A(\omega) \geqslant B(\omega)$) will mean $A(\omega) - B(\omega)$ is positive (nonnegative). This will often be abbreviated $A > B$ ($A \geqslant B$).

The EIV system consists of six random vector sequences $\{\hat{x}_k\}$, $\{\hat{y}_k\}$, $\{u_k\}$, $\{v_k\}$, $\{x_k\}$, $\{y_k\}$ of dimension n related by (1.1) and (1.2). The causal impulse response $\{W_k, k \geqslant 0\}$ is assumed to satisfy the stability requirement

$$\sum_{k=0}^{\infty} \|W_k\| < \infty. \tag{2.1a}$$

We furthermore assume that $\{\hat{x}_k\}$, $\{u_k\}$, $\{v_k\}$ are mutually independent, stationary, zero mean processes and that their power spectrum matrices are bounded and, respectively, positive, nonnegative, nonnegative.

For the factorization theory, and also to be consistent with [5], it is more convenient to use the mathematical literature notation where z denotes the *backward* shift operator, rather than z^{-1} which is used in engineering literature. The transfer matrix associated with the sequence $\{W_k, k \geqslant 0\}$, $W(z)$, is defined by

$$W(z) = \sum_{0}^{\infty} W_k z^k. \tag{2.2}$$

For technical reasons, we also assume $\det W(z) \neq 0$ for $|z| = 1$. We will discuss how this assumption can be removed later (see Remark 7 of Section IV). The assumptions in the preceding two paragraphs will be called the *standing assumptions*.

The standing assumptions ensure that $W(z)$ is analytic in $|z| < 1$. Note also that det $W(z)$ can have only a finite number of zeros in $|z| < 1$, and no zeros on $|z| = 1$.

A *standard data matrix* will be a $2n \times 2n$, bounded, nonnegative, Hermitian matrix

$$\Sigma(\omega) = \begin{bmatrix} \Sigma_{11}(\omega) & \Sigma_{21}(\omega)^* \\ \Sigma_{21}(\omega) & \Sigma_{22}(\omega) \end{bmatrix}$$

with Σ_{ij} $n \times n$ satisfying Σ_{11}, Σ_{22} positive and $\Sigma_{21}(\omega)$ nonsingular for all ω.

If x, y come from an EIV system satisfying the standing assumptions and $\Sigma(\omega)$ is the power spectrum matrix of $[x^*, y^*]^*$, then $\Sigma(\omega)$ is a standard data matrix since

$$\Sigma(\omega) = \begin{bmatrix} \Sigma_{xx}(\omega) & \Sigma_{xy}(\omega) \\ \Sigma_{yx}(\omega) & \Sigma_{yy}(\omega) \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{\hat{x}\hat{x}}(\omega) + \Sigma_{uu}(\omega) & \Sigma_{\hat{x}\hat{x}}(\omega) W(e^{j\omega})^* \\ W(e^{j\omega}) \Sigma_{\hat{x}\hat{x}}(\omega) & W(e^{j\omega}) \Sigma_{\hat{x}\hat{x}}(\omega) W(e^{j\omega})^* + \Sigma_{vv}(\omega) \end{bmatrix}. \tag{2.3}$$

As we shall see, the converse is not true.

Given a standard data matrix $\Sigma(\omega)$, our problem is to determine the class of EIV systems which satisfy the standing assumptions and are such that $\Sigma(\omega)$ is the power spectrum matrix of $[x^*, y^*]^*$.

The pair $W(e^{j\omega})$, $\Sigma_{\hat{x}\hat{x}}(\omega)$ corresponding to an EIV system which satisfies the standing assumptions and has $\Sigma(\omega)$ as the power spectrum matrix of $[x^*, y^*]^*$ will be called a *solution of the EIV problem.*

III. Summary of Matrix Factorization Theory

The matrix factorization theory we will be concerned with in this section has a long history, going back to Hilbert, Wiener and Hopf, Paley and Wiener and others, and has been closely associated with the solution of singular integral equations. More recently, the theory has been exposed by Gohberg, Krein, and Clancey, [8], [9] from which we draw the material in this section. The theory is better known to the linear systems and stochastic process communities in the specialized context of spectral factorization, or the slightly more general canonical factorization of [10]. We will, however, need the general theory of [9].

Throughout this note we will be concerned with the factorization, relative to the unit circle, of matrix functions with entries in the Wiener algebra (defined below), and will always state the factorization results for this algebra, although this is not necessary (see [9]).

Let C denote the unit circle $|z| = 1$ in the Riemann sphere $C \cup \{\infty\}$, and let C_+, C_- be, respectively, the regions $\{|z| < 1\}$, $\{|z| > 1\} \cup \{\infty\}$. The Wiener algebra of complex $n \times n$ matrix functions, which we denote $\mathcal{W}_n$, consists of all $n \times n$ complex matrices F on C of the form

$$F(e^{j\omega}) = \sum_{k=-\infty}^{\infty} F_k e^{jk\omega} \tag{3.1}$$

for which the norm

$$\|F\|_\omega = \sum_{k=-\infty}^{\infty} \|F\|_k \tag{3.2}$$

is finite. Note that $F(e^{j\omega})^* = \Sigma_{-\infty}^{\infty} F_k^* e^{-jk\omega}$.

We will denote by $\mathcal{W}_n^+$, $\mathcal{W}_n^-$, respectively, the subalgebras of matrix functions F_-, F_- of the form

$$F_-(e^{j\omega}) = \sum_{k=0}^{\infty} F_k e^{jk\omega} \tag{3.3a}$$

$$F_-(e^{j\omega}) = \sum_{k=-\infty}^{-1} F_k e^{jk\omega} \tag{3.3b}$$

$\mathcal{W}_n^+$ ($\mathcal{W}_n^-$) can also be thought of as the space of matrix functions with absolutely convergent Fourier series which are analytic in C_+ (C_-). Also observe that $F \in \mathcal{W}_n^+$ if and only if $F^* \in \mathcal{W}_n^- \oplus I$.

Further, let $G[\mathcal{W}_n]$, $G[\mathcal{W}_n^+]$, $G[\mathcal{W}_n^-]$ denote, respectively, the group of elements in $\mathcal{W}_n$, $\mathcal{W}_n^+$, $\mathcal{W}_n^- \oplus I$ which are invertible; that is, their inverses exist and are in $\mathcal{W}_n$, $\mathcal{W}_n^+$, $\mathcal{W}_n^- \oplus I$, respectively. Note that the space of rational $n \times n$ matrix functions is dense in $\mathcal{W}_n$.

Theorem 3.1 (Existence): Every element $F \in G[\mathcal{W}_n]$ admits a factorization $F = F_+ D F_-$ relative to C, where $F_+ \in G[\mathcal{W}_n^+]$, $F_- \in G[\mathcal{W}_n^-]$, and

$$D(z) = \operatorname{diag}\left[\left[\frac{z - z_+}{z - z_-}\right]^{\kappa_1}, \cdots, \left[\frac{z - z_+}{z - z_-}\right]^{\kappa_n}\right], \qquad z \in C \tag{3.4}$$

where z_+, z_- are arbitrary points in C_+, C_-, respectively, and $\kappa_1, \cdots, \kappa_n$ are integers, with $\kappa_i \geqslant \kappa_{i+1}$, $i = 1, \cdots, n-1$.

Remarks:

1) The factorization in the *theorem* is a left factorization of F. In the right factorization, the positions of F_+ and F_- are reversed. The integers κ_i, $i = 1, \cdots, n$ are called the left (right) partial indexes of the factorization, or since we will only be dealing with left factorizations, simply the *partial indexes*. The integer $\kappa = \kappa_1 + \cdots + \kappa_n$ is called the total index, and is equal to 2π times the change in arg det $F(e^{j\omega})$ around C. In general, no components of the factorization (i.e., F_+, D, F_-) are the same for left and right factorizations.

2) $F \in G[\mathcal{W}_n]$ is equivalent to $F \in \mathcal{W}_n$ and det $F(e^{j\omega}) \neq 0$ for all ω.

3) Since $0 \in C_+$ and $\infty \in C_-$, it proves possible to choose $D(z) =$ diag $[z^{\kappa_1}, \cdots, z^{\kappa_n}]$ for $z \in C$. In this case, the factorization will be called *standard.*

Theorem 3.2: The partial indexes in any two left factorizations of F are the same. In particular, if two factorizations use the same z_+, z the two D matrices must be the same; any two standard left factorizations of F have the same diagonal factor.

Remark: The factors $F_\pm$ are not unique, but can be characterized (see [8], [9]). As we will not require this characterization, we will not give any details.

Theorem 3.3 (Spectral Factorization): There exists $F_- \in G[\mathcal{W}_n^-]$ such that

$$F = F_-^* F_-, \qquad F_- \in G[\mathcal{W}_n^-]$$

if and only if F is positive and Hermitian.

With the notation of this section, observe that *the standing assumptions on W, $\Sigma_{\hat{x}\hat{x}}$ are equivalent to $W \in \mathcal{W}_n^+$ and $\Sigma_{\hat{x}\hat{x}}$ has a spectral factorization of the type in Theorem 3.3.*

IV. The General Solution

In this section we completely characterize the solution set of the EIV problem in terms of the factors of the cross-spectrum matrix $\Sigma_{yx}(\omega)$. This result extends the results of [5] to the multivariable case as promised. Before stating the main result, we define a set of $\mathcal{H}(D)$, and derive a few simple properties of it.

Definition: Let D be any diagonal matrix function of the form (3.4). Then

$$\mathcal{H}(D) = \{H_- \in G[\mathcal{W}_n^-] : DH_- \in \mathcal{W}_n^+\}. \tag{4.1}$$

Lemma 4.1: Let $H = H_+ DH_-$ be a left factorization of $H \in G[\mathcal{W}_n]$. Then $H \in \mathcal{W}_n^+$ if and only if $H_- \in \mathcal{H}(D)$.

Proof: Suppose $H \in \mathcal{W}_n^+$. Then $DH_- = H_+^{-1} H \in \mathcal{W}_n^+$, since $H_+^{-1} \in G[\mathcal{W}_n^+]$. Conversely, if $H_- \in \mathcal{H}(D)$, we have $DH_- \in \mathcal{W}_n^+$ and so $H = H_+ DH_- \in \mathcal{W}_n^+$ $\nabla\nabla\nabla$

Lemma 4.2: Let $\kappa_1, \cdots, \kappa_n$ be the partial indexes of D, where D is as in (3.4). Then $H_- \in \mathcal{H}(D)$ if and only if $\kappa_i \geqslant 0$, $i = 1, \cdots, n$ and $(H_-)_{ij}$ is a polynomial of degree $\leqslant \kappa_i$ in $1/(z - z_+)$ such that det $H_-(z) \neq 0$, $|z| \geqslant 1$.

Proof: If H_- is such that $(H_-)_{ij}$ is a polynomial of degree $\leqslant \kappa_i$ in $1/(z - z_+)$, with det $H(z) \neq 0$, $|z| \geqslant 1$ and $\kappa_i \geqslant 0$ for all i, then it is obvious that $H_- \in G[\mathcal{W}_n^-]$ and $DH_- \in \mathcal{W}_n^+$, so $H \in \mathcal{H}(D)$.

Conversely, if $H_- \in \mathcal{H}(D)$, then $H_- \in G[\mathcal{W}_n^-]$, so det $H(z) \neq 0$, $|z| \geqslant 1$. Let D be as in (3.4). Then

$$(DH_-)_{ij} = \left[\frac{z - z_+}{z - z_-}\right]^{\kappa_i} (H_-)_{ij} = H_{ij} \in \mathcal{W}_i^+, \qquad \text{as } H_- \in \mathcal{H}(D).$$

Hence,

$$(H_-)_{ij} = \left[\frac{z-z_-}{z-z_-}\right]^{\kappa_i} H_{ij}. \tag{4.2}$$

If $\kappa_i < 0$ for some i, (4.2) implies $(H_-)_{ij}$, which is analytic in C_-, admits an analytic continuation into C_+ (namely $[(z - z_-)/(z - z_+)]^{\kappa_i}H_{ij}$) and hence must be constant (recall $\infty \in C_-$). In view of the fact that $(H_-)_{ij} = 0$ for $z = z_+$ by (4.2), we have $(H_-)_{ij} = 0$. However, this implies H has a zero column, and so it is not invertible, contradicting $H_- \in G[\mathcal{W}_n]$. Thus, $\mathcal{K}(D)$ is empty if $\kappa_i < 0$ for any i.

If $\kappa_i \geq 0$ for all i, then it follows from (4.2) that $(H_-)_{ij}$ is analytic in C_- and may be continued analytically into C_+, with the exception of the point z_-, at which it has a pole of order $\leq \kappa_i$. Thus, $(H_-)_{ij}$ must be a polynomial in $1/(z - z_+)$ of degree $\leq \kappa_i$. ▽▽▽

Theorem 4.1: Suppose $\Sigma(\omega)$ is a standard data matrix and let $\Sigma_{21}(\omega) = F_+(e^{j\omega})D(e^{j\omega})F_-(e^{j\omega})$ be a fixed but arbitrary factorization of Σ_{21}, $F_+ \in G[\mathcal{W}_n^+]$, $F_- \in G[\mathcal{W}_n^-]$, D as in (3.4). For a solution to the EIV problem to exist, it is necessary and sufficient that there exists an $H_- \in \mathcal{K}(D)$ such that

$$H_-(e^{j\omega})H_-(e^{j\omega})^* \leq D(e^{j\omega})^{-1}F_+(e^{j\omega})^{-1}\Sigma_{22}(\omega)F_+(e^{j\omega})^{-*}D(e^{j\omega})^{-*} \tag{4.3a}$$

$$[H_-(e^{j\omega})H_-(e^{j\omega})^*]^{-1} \leq F_-(e^{j\omega})^{-*}\Sigma_{11}(\omega)F_-(e^{j\omega})^{-1}. \tag{4.3b}$$

In this case $W(e^{j\omega})$, $\Sigma_{\hat{x}\hat{x}}(\omega)$ is a solution of the EIV problem if and only if

$$W(e^{j\omega}) = F_+(e^{j\omega})D(e^{j\omega})H_-(e^{j\omega})H_-(^{j\omega})^*F_-(e^{j\omega})^{-*} \tag{4.4a}$$

$$\Sigma_{\hat{x}\hat{x}}(\omega) = F_-(e^{j\omega})^*(H_-(e^{j\omega})H_-(e^{j\omega})^*)^{-1}F_-(e^{j\omega}) \tag{4.4b}$$

where $H_- \in \mathcal{K}(D)$ is *any* solution of (4.3).

Proof: Note that since Σ is a standard data matrix, $\Sigma_{21} \in G[\mathcal{W}_n]$, and thus has a factorization by Theorem 3.1.

Part 1: We show that W, $\Sigma_{\hat{x}\hat{x}}$ satisfy the standing assumptions with $\Sigma_{21} \equiv \Sigma_{yx} = W\Sigma_{\hat{x}\hat{x}}$ if and only if (4.4) holds.

Let W, $\Sigma_{\hat{x}\hat{x}}$ satisfy (4.4) for some $H_- \in \mathcal{K}(D)$. Then

$$\begin{aligned} W\Sigma_{\hat{x}\hat{x}} &= F_+DH_-H_-^*F_-^{-*}F_-^*(H_-H_-^*)^{-1}F_- \\ &= F_+DF_- \\ &= \Sigma_{21}. \end{aligned}$$

Clearly, as $H_- \in \mathcal{K}(D)$, we have $DH_- \in \mathcal{W}_n^+$. Also F_+, H_-^*, $F_-^{-*} \in G[\mathcal{W}_n^+]$, so $W \in \mathcal{W}_n^+$. Since $\Sigma_{\hat{x}\hat{x}}$ has a spectral factorization, namely $(H_-^{-1}F_-)^*(H_-^{-1}F_-)$, it also satisfies the standing assumptions.

Conversely, let W, $\Sigma_{\hat{x}\hat{x}}$ satisfy the standing assumptions, with $\Sigma_{21} \equiv \Sigma_{yx} = W\Sigma_{\hat{x}\hat{x}}$. Let $\Sigma_{\hat{x}\hat{x}} = M_-^*M_-$. Hence,

$$W = \Sigma_{21}\Sigma_{\hat{x}\hat{x}}^{-1} = F_+DF_-M_-^{-1}M_-^{*} \in \mathcal{W}_n^-.$$

Since F_+, $M_-^{-*} \in G[\mathcal{W}_n^+]$, this implies $DF_-M_-^{-1} \in \mathcal{W}_n^-$. That is,

$$F_-M_-^{-1} = H_- \in \mathcal{K}(D) \text{ by Lemma 4.1.}$$

This gives $M_- = H_-^{-1}F_-$, and thus

$$\Sigma_{\hat{x}\hat{x}} = M_-^*M_- = F_-^*H_-^{-*}H_-^{-1}F_- = F_-^*(H_-H_-^*)^{-1}F_-$$

and

$$W = \Sigma_{yx}\Sigma_{\hat{x}\hat{x}}^{-1} = F_+DF_-F_-^{-1}H_-H_-^*F_-^{-*} = F_-DH_-H^*F_-^{-*}$$

and we are done with Part 1.

Part 2: We show any W, $\Sigma_{\hat{x}\hat{x}}$ satisfying (4.4) are compatible with the standard data if and only if (4.3) is satisfied.

Let W, $\Sigma_{\hat{x}\hat{x}}$ satisfying (4.4) also be compatible with the standard data matrix.

Thus,

$$W\Sigma_{\hat{x}\hat{x}}W^* + \Sigma_{vv} = \Sigma_{yy} \equiv \Sigma_{22},$$

i.e.,

$$F_+DH_-H_-^*D^*F_+^* = \Sigma_{22} - \Sigma_{vv}$$

so

$$\begin{aligned} H_-H_-^* &= D^{-1}F_+^{-1}(\Sigma_{22} - \Sigma_{vv})F_+^{-*}D^{-*} \\ &\leq D^{-1}F_+^{-1}\Sigma_{22}F_+^{-*}D^{-*}. \end{aligned}$$

Also,

$$\Sigma_{\hat{x}\hat{x}} + \Sigma_{uu} = \Sigma_{xx} \equiv \Sigma_{11},$$

i.e.,

$$F_-^*(H_-H_-^*)^{-1}F_- = \Sigma_{11} - \Sigma_{uu}$$

so

$$\begin{aligned} (H_-H_-^*)^{-1} &= F_-^{-*}(\Sigma_{11} - \Sigma_{uu})F_-^{-1} \\ &\leq F_-^{-*}\Sigma_{11}F_-^{-1}. \end{aligned}$$

Conversely, if $H_-H_-^*$ satisfies (4.3), define

$$\Sigma_{uu} = \Sigma_{11} - \Sigma_{\hat{x}\hat{x}}$$

$$\Sigma_{vv} = \Sigma_{22} - W\Sigma_{\hat{x}\hat{x}}W^*$$

and it is trivial to verify that they are nonnegative, Hermitian, bounded and are compatible with the standard data matrix. ▽▽▽

Remarks:

1) It is natural to ask whether minimum-phase solutions always exist to the problem, and if not, when they exist. This can easily be derived from Theorem 4.1 and the result is the following.

Definition: A matrix function $W \in \mathcal{W}_n$ will be called minimum-phase if W is causal, stable and has causal stable inverse, i.e., $W \in G[\mathcal{W}_n^+]$.

Corollary 4.1: Hypotheses as for Theorem 4.1. The standard data matrix is produced by a minimum-phase plant W if and only if all the left partial indexes of $\Sigma_{yx}(\omega)$ are zero and there is a nonsingular constant matrix satisfying

$$HH^* \leq F_+^{-1}\Sigma_{yy}F_+^{-*} \tag{4.5a}$$

$$(HH^*)^{-1} \leq F_-^{-*}\Sigma_{xx}F_-^{-1}. \tag{4.5b}$$

In this case, $W(e^{j\omega})$, $\Sigma_{\hat{x}\hat{x}}(\omega)$ is a solution of the EIV problem if and only if

$$W(e^{j\omega}) = F_+(e^{j\omega})HH^*F_-(e^{j\omega})^{-*}$$

$$\Sigma_{\hat{x}\hat{x}}(\omega) = F_-(e^{j\omega})^*(HH^*)^{-1}F_-(e^{j\omega})$$

where H is any constant nonsingular matrix satisfying (4.5).

In the scalar case, HH^* is just a real scalar confined to a finite interval. With H $n \times n$, we have $n(n+1)/2$ parameters forming the lower triangular portion of HH^* to adjust within the restrictions imposed by Corollary 4.1.

2) If Σ_{yx} is rational, there exists a rational factorization, i.e., $\Sigma_{yx} = F_+DF_-$, F_+, F_- rational [9, p. 14], and since $\mathcal{K}(D)$ has only rational elements (Lemma 4.2), we see that a rational spectrum Σ_{yx} can only come from rational W, $\Sigma_{\hat{x}\hat{x}}$. Indeed, it is possible to calculate the factorization F_+DF_- only when Σ_{yx} is rational. Likewise, if the standard data Σ is constant, Corollary 4.1 implies that all solutions W, $\Sigma_{\hat{x}\hat{x}}$ of the EIV problem are constant. Thus, rational (respectively, constant) data can only arise from a rational (respectively, static) problem.

3) Lemma 4.2 implies that the set $\mathcal{K}(D)$ can be parameterized finite dimensionally the coefficients of the polynomials being the parameters. Clearly then, the total number of parameters is not more than $n(\kappa + n)$, where $\kappa = \Sigma_i^n \kappa_i = 2\pi \times$ (change in arg det $\Sigma_{21}(\omega)$ around C) is the total index of $\Sigma_{21}(\omega)$.

4) Consider the scalar case $n = 1$. The technique of [5] gives rise to solutions of the form

$$W(z) = \mu U_A(z)\bar{W}_A(z) \tag{4.6}$$

where

$$U_A(z) = \frac{\prod_1^{\kappa}(z-\alpha_i)}{\prod_{i=1}^{\kappa}(1-\bar{\alpha}_i z)} \tag{4.7}$$

is an all-pass function and $\bar{W}_A(z)$ is minimum-phase, dependent on $A = \{\alpha_i: |\alpha_i| < 1, i = 1, \cdots, \kappa\}$.

To connect this with Theorem 4.1, let F_-DF_- be a standard factorization of Σ_{yx}. Now observe that an arbitrary $H_- \in \mathcal{K}(D)$ has the form

$$H_-(z) = \frac{h\prod_1^{\kappa}(z-\alpha_i)}{z^{\kappa}}, \quad |\alpha_i|<1, \; i=1, \cdots \kappa \tag{4.8}$$

with h an arbitrary scalar and κ the total index. Observe also that

$$H_-^*(z) = h^* \prod_1^{\kappa}(1-\bar{\alpha}z), \quad |z|=1 \tag{4.9}$$

so that the all-pass factor (4.7) may be written

$$U_A(z) = \frac{D(z)H_-(z)}{H_-^*(z)}, \quad |z|=1. \tag{4.10}$$

Now from (4.4a) W is $F_+DF_-^{-*}|H_-|^2$, and this can be written in the form of (4.6) with the scaling constant $\mu = |h|^2$, and

$$\bar{W}_A = \frac{F_-F_-^{-*}(H_-^*)^2}{|h|^2}. \tag{4.11}$$

5) Theorem 4.1 and Lemma 4.2 imply that *the solution set of the EIV problem is empty if* Σ_{yx} *has a negative left partial index.* If Σ_{yx} does have a negative partial index, we can find $W \in \mathcal{W}_n$ compatible with the standard data matrix, but such a W cannot be causal, and hence does not satisfy the standing assumptions. Of course, even if all the left partial indexes of Σ_{yx} are nonnegative, noncausal W compatible with the standard data matrix can exist (simply relax the condition $H \in \mathcal{K}(D)$ to $H \in G[\mathcal{W}_n^-]$ in Theorem 4.1). Inverse-causal W (i.e., W^{-1} causal) which satisfy the standard data matrix will exist only when all the left partial indexes of Σ_{21} are nonpositive (to see this, interchange x and y and observe that the left partial indexes of Σ_{xy} are the negative of the left partial indexes of Σ_{yx}). Although not surprising since multivariable transfer matrices can have both causal and inverse causal channels, this situation is in contrast to the scalar case, where we can always find W compatible with the standard data and either W or W^{-1} satisfying the standing assumptions. To recover a corresponding property for the multivariable case, one should perhaps interchange some of the input and output components, rather than the entire vectors.

6) The proof of Theorem 4.1, in particular Part 2, shows that if any one of the four unknowns W, $\Sigma_{\hat{x}\hat{x}}$, Σ_{uu}, Σ_{vv} is available, the identification problem is uniquely solvable. Indeed, any information about these four quantities, in particular upper bounds on Σ_{uu}, Σ_{vv}, will reduce the solution set.

7) Part of the assumptions on the standard data was $\det \Sigma_{21}(\omega) \neq 0$ for $\omega \in [0, 2\pi]$. If $\det \Sigma_{21}(\omega) = 0$ at ω_i, $i = 1, \cdots, N < \infty$ and is nonzero otherwise, the problem can be solved similarly to the scalar case (see [5]) as follows.

Form a contour C_ϵ by perturbing the unit circle C towards the origin by circular arcs of radius ϵ centered at the points ω_i, $i = 1, \cdots, N$. Choose ϵ such that $\det \Sigma_{21}(z) \neq 0$, $z \in C_\epsilon$ and such that $\Sigma_{11}(z)$ is positive on C_ϵ. Then perform all factorizations relative to C_ϵ instead of C and proceed as before.

8) It is natural to ask whether the procedure developed in this note can be extended to nonsquare systems, where W, Σ_{yx} are $n \times m$ and $\Sigma_{\hat{x}\hat{x}}$ is $m \times m$. If Σ_{yx} is rational with $n > m$ $(n < m)$, we can reduce it by premultiplication by a unimodular polynomial matrix in z (postmultiplication by a unimodular polynomial matrix in $1/z$) to a matrix where the last $n - m$ rows $(m - n$ columns) are zero. This reduces the problem to a square one which can be solved using Theorem 4.1. There is however an important distinction between the cases $m \leqslant n$ and $m > n$. The first case is the same in its essentials as the square case, with a finite-dimensional parameterization of the solution set. In the second case (more inputs than outputs) the solution set is no longer finite dimensional. This is because the set $\{H_- \in G[\mathcal{W}_m^-] : [D\,0]H_-$ is analytic in $|z| < 1$, with D $n \times n$ as in (3.4)$\}$ is not finite dimensional. In the nonrational case, the reduction by unimodular polynomial matrices is not possible, and identifying the solution set remains an open problem.

9) Theorem 4.1 provides necessary and sufficient conditions for a power spectrum matrix to arise from an $n \times n$ (or $n \times m$ as in previous remark) EIV system. However, we assume that n (or n & m) are known *a priori*. Theorem 4.1 does not provide any procedure for finding n (n & m) such that a solution must exist, apart from trial and error. In other words, it will tell when one has chosen the wrong set of inputs and outputs, but will not (as yet) help decide which inputs/outputs should be chosen. See also Remark 5. This is currently under investigation.

V. Conclusion

By hypothesizing causality of a transfer matrix appearing in a dynamic, multivariable, errors-in-variable model and making certain other reasonable assumptions, it is possible to parameterize the class of transfer matrices consistent with the available data in a finite-dimensional way.

Despite this complete characterization of the solution set of the EIV problem, a number of questions remain. It is of particular importance to establish the robustness properties of solution set, since the standard data power spectrum matrix will in practice be only an approximation of the true spectrum. Such would be clearly the case if the spectrum were constructed from a finite data sample, or if data were not available within a certain frequency range. Furthermore, to construct the solution set of the EIV problem, it is necessary to reduce the "in principle" constructive procedure for obtaining factorizations of rational matrices in [9] to a practical numerical algorithm.

References

[1] C. Gini, "Sull interpolazione di una retta quando i valori della variable independente sono affetti da errori accidentali," *Metron*, vol. 1, pp. 63-82, 1921.

[2] R. Frisch, "Statistical confluence analysis by means of complete regression systems," Econom. Inst., Univ. Oslo, Pub. 5, 1934.

[3] R. E. Kalman, "System identification from noisy data," in *Proc. Int. Symp. Dynam. Syst.*, Gainsville, FL, 1981.

[4] B. D. O. Anderson and M. Deistler, "Identifiability in dynamic errors-in-variables models," *J. Time Series Anal.*, vol. 5, pp. 1-13, 1984.

[5] B. D. O. Anderson, "Identification of scalar errors-in-variables models with dynamics," *Automatica*, to be published.

[6] A. Madansky, "The fitting of straight lines when both variables are subject to error," *J. Amer. Statist. Assoc.*, vol. 54, pp. 173-205, 1959.

[7] P. A. P. Moran, "Estimating structural and functional relationships," *J. Multivariable Anal.*, vol. 1, pp. 232-255, 1971.

[8] I. C. Gohberg, and M. G. Krein, "Systems of integral equations on a half line with kernels depending on the difference of arguments," *Uspehi Mat. Nauk*, vol. 13, pp. 3-72, 1958, pp. 3-72; English trans., *Amer. Math. Soc. transl.*, vol. 2, pp. 217-287, 1960.

[9] K. Clancey and I. C. Gohberg, *Factorization of Matrix Functions and Singular Integral Operators.* New York: Birkhäuser Verlag, Basel.

[10] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties.* New York: Academic, 1975.

On Almost Sure Convergence of Adaptive Algorithms

D. H. SHI AND F. KOZIN

Abstract—**We present an extension of the Furstenberg–Kesten theorem on the convergence of random matrices. This extension is applied to the study of almost sure convergence of certain adaptive algorithms. In particular, we establish that the NLMS algorithm is almost surely convergent under extremely weak necessary and sufficient conditions. We**

Manuscript received August 6, 1985; revised October 11, 1985. This work was supported in part by the National Science Foundation under Grant CEE-8311190.

D. H. Shi is with Xiamen University, China.

F. Kozin is with the Department of Electrical Engineering and Computer Science, Polytechnic Institute of New York, Brooklyn, NY 11201.

IEEE Log Number 8406887.