Invertibility of Higher-Order Moment Matrices

Robert D. Nowak, *Student Member, IEEE*, and Barry D. Van Veen, *Member, IEEE*

Abstract— **Signal processing algorithms based on Volterra filter structures often require that a matrix composed of higher-order moments of the underlying process be invertible. Previously, this problem has been studied for uncorrelated random variables. This paper establishes conditions under which a wide class of correlated processes have invertible higher-order moment matrices.**

I. Introduction

Volterra filters have received increasing attention in the recent signal processing literature. Algorithms for Volterra filtering applications, such as signal detection [12], estimation [3,4,12] and system identification [9,10], often require inversion of a matrix composed of higher-order moments or statistics of the underlying random process. Hence, the existence of the inverse is an important issue. The non-singularity of moment matrices is also important in the study of stability and convergence of adaptive algorithms [7,8]. Previous authors have considered the invertibility of moment matrices corresponding to uncorrelated random variables [2,4,10].

This paper generalizes earlier work to include a much wider class of correlated random processes. First, it is shown that the moment matrix corresponding to absolutely continuous random variables is invertible. Second, necessary and sufficient conditions are established for the invertibilty of the moment matrix corresponding to a linear transformation of a given random process.

Notation and previous work are reviewed in section II. In section III, the case of jointly absolutely continuous random variables is examined. In section IV, the invertibility of a general class of moment matrices corresponding to linear transformations of random variables is studied. These results are generalized to polynomial transformations in section V. Section VI includes several applications relevant to adaptive filtering and nonlinear system identification.

II. Notation and Previous Work

Consider a collection of real-valued random variables $\{X_i : 1 \leq i \leq m\}$. Througout the paper, let the integer $N \geq 1$ be fixed. If $\{X_i : 1 \leq i \leq m\}$ is the input to an N-th order Volterra filter, then the output is a linear combination of the random variables [11]

$$\{ \prod_{i=1}^{m} X_i^{n_i} \ : \ \sum_{i=1}^{m} n_i \leq N, \ n_i \in \{0,1,\ldots,N\} \}. \quad (1)$$

There are exactly $r_m = \binom{m+N}{m}$ random variables in the collection above. The number r_m is the number of N-selections from an $m+1$ set [5,11]. Define the $r_m \times 1$ random vector $\bar{\boldsymbol{\mathcal{X}}}$ whose elements are the random variables in (1).

Let $E\{\cdot\}$ denote the expectation operator. The random variables $\{X_i : 1 \leq i \leq m\}$ are said to be *uncorrelated up to order n* if $E|X_i|^n < \infty$ for $1 \leq i \leq m$ and

$$E\{\prod_{i=1}^{m} X_i^{p_i}\} \ = \ \prod_{i=1}^{m} E\{X_i^{p_i}\}$$

for $\sum_{i=1}^{m} p_i \leq n, \ 0 \leq p_i \leq n$. The condition $E|X_i|^n < \infty$, $1 \leq i \leq m$ is necessary and sufficient to guarantee that the cross moments up to order n exist [4]. Throughout the paper we assume, without explicitly stating, $E|X_i|^{2N} < \infty$, $1 \leq i \leq m$, so that all moments involved exist.

Previous Work

Begin with the single random variable case.

Lemma 1.1 [13]: Let $m = 1$. The matrix $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is singular if and only if X_1 is a discrete random variable that takes on at most N distinct values with positive probability. □

The following results have been obtained for the case $m \geq 1$.

Lemma 1.2 [4]: The matrix $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is singular if one or more of the random variables $\{X_i\}_{i=1}^m$ is a discrete random variable that takes at most N distinct values with positive probability. □

If the random variables are sufficiently uncorrelated, the last result can be strengthened.

Lemma 1.3 [4]: If the random variables $\{X_i\}_{i=1}^m$ are uncorrelated up to order $2N$, then the matrix $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is singular if and only if one or more of the random variables is a discrete random variable that takes at most N distinct values with positive probability. □

Generalizations of these results and an excellent survey of applications are found in [4]. Similar results have been obtained for i.i.d. sequences and a special class of deterministic signals known as pseudo-random multilevel sequences [10].

R. Nowak is supported by Rockwell International Doctoral Fellowship Program. B. Van Veen is supported in part by the National Science Foundation under Award MIP-895 8559, Army Research Office under Grant DAAH04-93-G-0208, and by the National Institute of Health under Grant NINCDS #NSR 0116436.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin, 1415 Johnson Drive, Madison, WI 53706.

III. Jointly Continuous Random Variables

Define the random vector $\mathbf{X} \triangleq (X_1, \ldots, X_m)^T$.

Theorem 3.1: If $\mathbf{X}$ has a density with respect to m-dimensional Lebesgue measure, then $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is invertible.

A proof is given in Appendix A. Note that no assumptions are made on the correlation between the random variables. Hence, Theorem 3.1 can be used to establish the invertibility of $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ even when the random variables are correlated. Throughout the remainder of the paper we assume that all densities are with respect to Lebesgue measure and simply say that a random variable/vector has a density.

IV. Linear Transformations of Random Variables

Consider the linear transformation $\mathbf{Y} \triangleq \mathbf{HX}$, where $\mathbf{H}$ is a constant $q \times m$ matrix and we assume $q \leq m$. Let $\bar{\mathcal{Y}}$ be a $r_q \times 1$ random vector ($r_q = \binom{q+N}{q}$) whose elements are the random variables

$$\{ \prod_{i=1}^{q} Y_i^{n_i} \ : \ \sum_{i=1}^{q} n_i \leq N, \ n_i \in \{0, 1, \ldots, N\} \ \}. \qquad (2)$$

Theorem 4.1: If $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is invertible, then $E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\}$ is invertible if and only if $\mathbf{H}$ is full rank. If $q = m$, then the invertibility of $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is also a necessary condition.

The proof of Theorem 4.1 is given in Appendix B. Recall that Lemma 1.1, Lemma 1.3, or Theorem 3.1 can be used to establish the invertibility of $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$. Also note that, in general, $Y_1, \ldots, Y_q$ are correlated.

Next, Corollary 4.1 gives conditions for the invertibility of the $2n$-th order homogeneous moment matrix and follows directly from Theorem 4.1. Here we only require that $\mathbf{X}$ has finite cross moments of order $2n$. For $n \geq 1$ define the vectors $\bar{\mathbf{X}}^{(n)}$ and $\bar{\mathbf{Y}}^{(n)}$ whose elements are the random variables,

$$\{ \prod_{i=1}^{m} X_i^{n_i} \ : \ \sum_{i=1}^{m} n_i = n, \ n_i \in \{0, 1, \ldots, n\} \ \}$$

and

$$\{ \prod_{i=1}^{q} Y_i^{n_i} \ : \ \sum_{i=1}^{q} n_i = n, \ n_i \in \{0, 1, \ldots, n\} \ \}$$

respectively.

Corollary 4.1: If $E\{\bar{\mathbf{X}}^{(n)}\bar{\mathbf{X}}^{(n)^T}\}$ is invertible, then $E\{\bar{\mathbf{Y}}^{(n)}\bar{\mathbf{Y}}^{(n)^T}\}$ is invertible if and only if $\text{rank}\mathbf{H} = q$. If $q = m$, then the invertibility of $E\{\bar{\mathbf{X}}^{(n)}\bar{\mathbf{X}}^{(n)^T}\}$ is a necessary condition. □

V. Extension to Polynomial Transformations

In the previous section we considered linear transformations of random variables. The results easily generalize to polynomial transformations. Let $\{Z_j : 1 \leq j \leq r\}$ be real-valued random variables. Set $X_i = \prod_{1 \leq j \leq r} Z_j^{p_{i,j}}$, $1 \leq i \leq m$ with $p_{i,j} \in \{0, 1, \ldots, n\}$, $\sum_{j=1}^{r} p_{i,j} \leq n$ for each i, and assume that the $\{X_i\}_{i=1}^{m}$ are unique products of the $\{Z_j\}_{j=1}^{r}$. Now $\mathbf{Y} = \mathbf{HX}$ defines a polynomial transformation of the underlying $\{Z_j\}_{j=1}^{r}$. Note that in this case $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is composed of cross moments of $\{Z_j\}_{j=1}^{r}$ up to order $2Nn$. Provided these moments exist, Theorem 4.1 is applicable. If $\mathbf{Z} \triangleq (Z_1, \ldots, Z_r)^T$ has a density and $\mathbf{H}$ is full rank, then Theorems 3.1 and 4.1 indicate that $E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\}$ is invertible. However, note that in general $\mathbf{X}$ defined in this manner does not have a density.

VI. Application to Adaptive Filtering and System Identification

Let $\mathbb{Z}$ denote the set of integers and let $\{X_k : k \in \mathbb{Z}\}$ be a stationary random process. Let

$$\mathbf{X} = [X_k, \ X_{k-1}, \ \ldots, \ X_{k-m+1}]^T.$$

Technically, the process need only be $2N$-th order stationary, for appropriate N, where $2N$-*th* order stationary is the obvious generalization of wide sense stationarity ($N = 1$).

Definition 5.1 [10]: If the $2N$-th order correlation matrix $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ exists and is invertible then $\{X_k\}_{k \in \mathbb{Z}}$ is called *persistently exciting of order* N *and degree* m or $p.e.(N, m)$ for short.

Remark 5.1: Theorem 3.1 implies that if

$$\mathbf{X} = [X_k, \ X_{k-1}, \ \ldots, \ X_{k-m+1}]^T$$

has a density, then $\{X_k\}_{k \in \mathbb{Z}}$ is $p.e.(N, m)$ for any finite N provided the required moments exist.

Remark 5.2: For practical system identification experiments, one needs to consider the sample moment matrix obtained from finite set of observations. If $\mathbf{X}$ has a density, then satisfaction of the $p.e.$ condition implies the sample moment matrix is invertible[1] *with probability 1* provided at least r_m observations of $\bar{\mathcal{X}}$ are used to form the sample moment matrix. However, in general, there is a non-zero probability that the sample moment matrix is singular.

Example 1 - Moving Average Processes: Theorem 4.1 implies that a p-*th* order moving average (MA) of a $p.e.(N, p+q)$ process is $p.e.(N, q)$. Suppose

$$Y_k = \sum_{i=1}^{p} h_k X_{k-i}$$

with $h_i \in \mathbb{R}$, $1 \leq i \leq p$. Let $\mathbf{Y} = [Y_k, \ Y_{k-1}, \ \ldots, \ Y_{k-q+1}]^T$. Define the $q \times (p+q)$ matrix

$$\mathbf{H} \ = \ \begin{bmatrix} h_1 & h_2 & \cdots & h_p & & \\ & \ddots & \ddots & & \ddots & \\ & & h_1 & h_2 & \cdots & h_p \end{bmatrix}. \qquad (3)$$

Provided $h_i \neq 0$ for some $1 \leq i \leq p$ the matrix $\mathbf{H}$ is full rank. Thus if $\{X_k\}_{k \in \mathbb{Z}}$ is $p.e.(N, p+q)$, then $\{Y_k\}_{k \in \mathbb{Z}}$ is $p.e.(N, q)$. □

[1] This follows directly from Remark 3.1 in Appendix A.

Example 2 - Static Polynomial Transformations: Let $n \geq 1$ be an integer and suppose $Y_k = \sum_{i=1}^{n} \gamma_i X_k^i$, with $\gamma_i \in \mathbb{R},\ 1 \leq i \leq n$. Let

$$\mathbf{X} = [X_k,\ X_k^2,\ \ldots,\ X_k^n,\ X_{k-1},\ X_{k-1}^2,\ \ldots,\ X_{k-m+1}^n]^T,$$

$\mathbf{Y} = [Y_k,\ Y_{k-1},\ \ldots,\ Y_{k-m+1}]^T$, and define the matrix $\mathbf{H}$ by

$$\mathbf{H} = \begin{bmatrix} \gamma_1 \cdots \gamma_n & & & \\ & \gamma_1 \cdots \gamma_n & & \\ & & \ddots & \\ & & & \gamma_1 \cdots \gamma_n \end{bmatrix}. \quad (4)$$

Provided $\gamma_i \neq 0$ for some $1 \leq i \leq n$ the matrix $\mathbf{H}$ is full rank. Thus, by Theorem 4.1, if $\{X_k\}_{k\in\mathbb{Z}}$ is $p.e.(nN, m)$, then $\{Y_k\}_{k\in\mathbb{Z}}$ is $p.e.(N, m)$.□

The results in Examples 1 and 2 can be further generalized to establish conditions under which the output of a Volterra filter driven by $\{X_k\}$ is $p.e.$ Following from the discussion in section V, if a $p.e.(nN, m+q)$ process $\{X_k\}$ is the input to a non-zero n-th order Volterra filter with memory q, then the output process is guaranteed to be $p.e.(N, m)$.

Remark 5.3: Lemma 1.3 implies that a Gaussian white noise (GWN) is $p.e.(N, m)$ for any positive integers N, m [10]. Hence, polynomial transformations (*i.e.,* Volterra filters) of GWN processes are $p.e.(N, m)$ for any positive integers N, m.

In fact, GWN driven ARMA processes are also $p.e.(N, m)$ for any positive integers N, m.

Example 3 - ARMA Processes: Let $\{Y_k\}_{k\in\mathbb{Z}}$ be a stationary ARMA process driven by GWN. Let

$$\mathbf{Y} = [Y_k,\ \ldots, Y_{k-m+1}]^T$$

and $\mathbf{R} = E\{\mathbf{Y}\mathbf{Y}^T\}$, the second-order correlation matrix of $\{Y_k\}_{k\in\mathbb{Z}}$. It is well known that $\mathbf{R}$ is invertible [14]. Let $\mathbf{H}$ be an $m \times m$ matrix whose columns are the orthonormal eigenvectors of $\mathbf{R}$. If $\mathbf{X} \triangleq \mathbf{H}^{-1}\mathbf{Y}$, then the elements of $\mathbf{X}$ are uncorrelated Gaussian random variables and hence are independent. Since $\mathbf{X}$ is vector of independent Gaussian random variables, Lemma 1.3 indicates that $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is invertible. $\mathbf{Y} = \mathbf{H}\mathbf{X}$ and $\mathbf{H}$ is invertible so, by Theorem 4.1, $\{Y_k\}_{k\in\mathbb{Z}}$ is $p.e.(N, m)$ for any positive integers N, m. Alternatively, it can be shown that an ARMA process has a density and hence Theorem 3.1 applies. □

VII. Conclusions

It is shown that a large class of correlated random variables lead to invertible moment matrices. If $\mathbf{X}$ has a density with respect to Lebesgue measure, then the corresponding moment matrix $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is invertible. Linear and polynomial transformations of random variables are also considered. Necessary and sufficient conditions are given for the invertibility of the corresponding moment matrices. In the context of adaptive filtering and system identification, invertibility of moment matrices implies that the underlying random process is persistently exciting. It is shown that many common types of linear and polynomial transformations preserve the excitation properties of stationary random processes.

Appendix A

The following well-known fact is used to prove Theorem 3.1.

Remark 3.1 [15, pp. 28-29]: Let $p : \mathbb{R}^m \rightarrow \mathbb{R}$ be a polynomial, $p \neq 0$, and λ denote Lebesgue measure on $\mathbb{R}^m$. The set of real zeros of p has zero Lebesgue measure; that is

$$\lambda(\{\mathbf{x} \in \mathbb{R}^m : p(\mathbf{x}) = 0\}) = 0.$$

Proof of Theorem 3.1: We proceed by establishing that $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is a positive definite matrix. Let $\mathbf{a} \in \mathbb{R}^{r_m}$ be arbitrary. Positive definiteness requires $\mathbf{a}^T E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}\mathbf{a} > 0$ or $E\{(\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}})^2\} > 0$ for all $\mathbf{a} \neq \mathbf{0}$. In general, $E\{(\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}})^2\} \geq 0$ with equality if and only if $\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}} = 0$ almost surely, that is $P(\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}} = 0) = 1$, where P is the probability measure on the underlying probability space. However, $\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}}$ is a polynomial in the elements of $\mathbf{X}$ and since $\mathbf{X}$ has a density f, Remark 3.1 implies $P(\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}} = 0) = 0$ for all $\mathbf{a} \neq \mathbf{0}$. Hence, $E\{(\mathbf{a}^T\bar{\boldsymbol{\mathcal{X}}})^2\} > 0$ for all $\mathbf{a} \neq \mathbf{0}$ and $E\{\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T\}$ is positive definite. □

Appendix B

The proof of Theorem 4.1 involves Kronecker products. Let $\mathbf{X}^{(1)} \triangleq \mathbf{X}$, $\mathbf{Y}^{(1)} \triangleq \mathbf{Y}$, $\mathbf{H}^{(1)} \triangleq \mathbf{H}$, and for $n \geq 2$ recursively define the n-fold Kronecker products $\mathbf{X}^{(n)} \triangleq \mathbf{X}^{(n-1)} \otimes \mathbf{X}$, $\mathbf{Y}^{(n)} \triangleq \mathbf{Y}^{(n-1)} \otimes \mathbf{Y}$, and $\mathbf{H}^{(n)} \triangleq \mathbf{H}^{(n-1)} \otimes \mathbf{H}$. The identity $\mathbf{Y}^{(n)} = \mathbf{H}^{(n)}\mathbf{X}^{(n)}$ is easily derived using $\mathbf{Y} = \mathbf{H}\mathbf{X}$ and Kronecker product identities [6]. Also, it can be shown that $\mathrm{rank}\mathbf{H}^{(n)} = (\mathrm{rank}\mathbf{H})^n$.

For $N \geq 1$ define $\boldsymbol{\mathcal{X}} \triangleq [1, \mathbf{X}^{(1)^T}, \ldots, \mathbf{X}^{(N)^T}]^T$ and $\boldsymbol{\mathcal{Y}} \triangleq [1, \mathbf{Y}^{(1)^T}, \ldots, \mathbf{Y}^{(N)^T}]^T$. The vectors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ contain all random variables in (1) and (3) respectively. However, due to the Kronecker product construction, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ contain redundant random variables, *e.g.,* X_1X_2 and X_2X_1. Define full rank matrices $\mathbf{P}_m$ and $\mathbf{P}_q$ that eliminate the redundant products so that $\mathbf{P}_m\boldsymbol{\mathcal{X}} = \bar{\boldsymbol{\mathcal{X}}}$ and $\mathbf{P}_q\boldsymbol{\mathcal{Y}} = \bar{\boldsymbol{\mathcal{Y}}}$, where $\bar{\boldsymbol{\mathcal{X}}}$ and $\bar{\boldsymbol{\mathcal{Y}}}$ are the vectors defined from (1) and (3). Finally define the block diagonal matrix $\boldsymbol{\mathcal{H}} \triangleq \mathrm{diag}\{1, \mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(N)}\}$ where the operator diag is defined such that the rows and columns of each rectangular block matrix are nonoverlapping. The next two properties follow directly from the previous definitions.

Property B.1: $\mathrm{rank}\boldsymbol{\mathcal{H}} = \sum_{i=0}^{N}(\mathrm{rank}\mathbf{H})^i$. □

Property B.2: $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{X}}$. □

We now prove Theorem 4.1.

Proof: First consider the case $q = m$. Using the definition of $\mathbf{P}_m$ and Property B.2,

$$E\{\bar{\boldsymbol{\mathcal{Y}}}\bar{\boldsymbol{\mathcal{Y}}}^T\} = E\{\mathbf{P}_m\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^T\mathbf{P}_m^T\} = \mathbf{P}_m\boldsymbol{\mathcal{H}}E\{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^T\}\boldsymbol{\mathcal{H}}^T\mathbf{P}_m^T.$$

Sufficiency is established as follows. Note that

$$\mathrm{rank}E\{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^T\} \leq r_m,$$

with equality if and only if $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is invertible. Also, by Property B.1, $\mathcal{H}$ is invertible if and only if $\mathbf{H}$ is invertible. Hence, if $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ and $\mathbf{H}$ are invertible, then $E\{\mathcal{Y}\mathcal{Y}^T\} = \mathcal{H}E\{\mathcal{X}\mathcal{X}^T\}\mathcal{H}^T$ has rank r_m. The action of $\mathbf{P}_m$ and $\mathbf{P}_m^T$ removing duplicate rows and columns does not affect the rank so $E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\}$ has rank r_m and is invertible.

On the other hand, if $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is singular, then $E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\}$ is singular since

$$\begin{aligned} r_m &> \operatorname{rank}E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}, \\ &= \operatorname{rank}E\{\mathcal{X}\mathcal{X}^T\}, \\ &\geq \operatorname{rank}\mathbf{P}_m\mathcal{H}E\{\mathcal{X}\mathcal{X}^T\}\mathcal{H}^T\mathbf{P}_m^T, \\ &= \operatorname{rank}E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\}. \end{aligned}$$

Therefore, in this case the invertibility of $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is a necessary condition for $E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\}$ to be invertible. Also note that if $\mathbf{H}$ is not invertible and has rank $s < m$, then $\mathbf{HX}$ lies in an s-dimensional subspace of $\mathbb{R}^m$. It follows that $\operatorname{rank}\mathcal{H}E\{\mathcal{X}\mathcal{X}^T\}\mathcal{H}^T \leq \binom{s+N}{s} < r_m$. Thus, the invertibility of $\mathbf{H}$ is necessary as well.

For the case $q < m$, the sufficiency is established in a similar fashion. Assume $\operatorname{rank}\mathbf{H} = q$ and let $\hat{\mathbf{H}}$ be an $m \times m$ matrix with the first q rows identical to $\mathbf{H}$ and with the remaining $m-q$ rows chosen such that $\hat{\mathbf{H}}$ is invertible. Define $\hat{\mathcal{H}} = \operatorname{diag}\{\hat{\mathbf{H}}^{(0)}, \hat{\mathbf{H}}^{(1)}, \ldots, \hat{\mathbf{H}}^{(N)}\}$. Let $\hat{\mathbf{Y}} = \hat{\mathbf{H}}\mathbf{X}$ and $\hat{\mathcal{Y}} = \mathbf{P}_m\hat{\mathcal{H}}\mathcal{X}$. $\hat{\mathbf{Y}}$ contains all random variables in $\bar{\mathcal{Y}}$ and additional random variables involving the last $m-q$ random variables of $\hat{\mathbf{Y}}$. Define a full rank matrix $\hat{\mathbf{P}}$ such that $\hat{\mathbf{P}}\hat{\mathcal{Y}} = \bar{\mathcal{Y}}$. If $E\{\bar{\mathcal{X}}\bar{\mathcal{X}}^T\}$ is invertible, then the result for the $q = m$ case implies $\hat{\mathbf{R}} = E\{\mathbf{P}_m\hat{\mathcal{H}}\mathcal{X}\mathcal{X}^T\hat{\mathcal{H}}^T\mathbf{P}_m^T\}$ is invertible. Hence, $E\{\bar{\mathcal{Y}}\bar{\mathcal{Y}}^T\} = \hat{\mathbf{P}}\hat{\mathbf{R}}\hat{\mathbf{P}}^T$ is invertible since $\hat{\mathbf{P}}$ is full rank. Necessity of the condition $\operatorname{rank}\mathbf{H} = q$ is established in the same manner as given for the case $q = m$. □

References

[1] P. Billingsley, *Probability and Measure*, New York: John Wiley & Sons, 1986.

[2] P. Bondon, "Invertibility conditions of a class of moment matrices and applications," *Cahiers de la Recherche Scientifique et Technique*, vol. 1, no. 1, April 1993.

[3] P. Bondon, M. Benidir and B. Picinbono, "A nonlinear approach to estimate the amplitude of a signal," *Proc. ICASSP-92*, pp. V301-V304 San Francisco, 1992.

[4] P. Bondon, M. Benidir and B. Picinbono, "Sur l'inversibilité de la matrice de corrélation d'un vecteur aléatoire à composantes polynomiales," Colloque GRETSI, Juan-les-Pins, France, pp.13-16, 1991.

[5] R. C. Bose and B. Manvel, *Introduction to Combinatorical Theory*, New York, John Wiley & Sons, pp. 8-9, 1984.

[6] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Cir. and Sys.*, vol. 25, pp. 772-781, 1978.

[7] J. A. Bucklew, T. G. Kurtz, and W. A. Sethares, "Weak Convergence and Local Stability Properties of Fixed Step Size Recursive Algorithms," *IEEE Trans. Inform. Theory*, vol. 39, pp. 966-978, 1993.

[8] S. Dasgupta, Y. Shrivastava, and G. Krenzer, "Persistent Excitation in Bilinear Systems," *IEEE Trans. Auto. Control*, vol. 36, no. 3, pp. 305-313, 1991.

[9] V. J. Mathews, "Adaptive polynomial filters," *IEEE SP Magazine*, vol. 8, no. 3, pp. 10-26, 1991.

[10] R. Nowak and B. Van Veen, "Random and pseudorandom inputs for Volterra filter identification," *IEEE Trans. Signal Proc.*, vol. 42, No. 8, August 1994.

[11] R. Nowak and B. Van Veen, "Efficient methods for identification of Volterra filter models," *Signal Processing*, vol. 38, No. 3, pp. 417-428, August 1994.

[12] B. Picinbono and P. Duvaut, "Optimal linear-quadratic systems for detection and estimation," *IEEE Trans. Inform. Theory*, vol. 34, pp. 304-311, 1988.

[13] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society, pp. 5-6, 1950.

[14] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International, Cambridge UK, 1989.

[15] W. Murray Wonham, *Linear Multivariable Control*, Springer-Verlag, New York, 1979.