Low Rank Estimation of Higher Order Statistics

Thomas F. André[†], Robert D. Nowak[*], *Member, IEEE*, Barry D. Van Veen[‡], *Member, IEEE*,

[†‡]Department of Electrical and Computer Engineering
University of Wisconsin-Madison
1415 Engineering Drive, Madison, WI 53706 USA

[*] Department of Electrical & Computer Engineering
Rice University, Houston, TX 77005 USA

EDICS SP3.2

Abstract — Low rank estimators for higher order statistics are considered in this paper. Rank reduction methods offer a general principle for trading estimator bias for reduced estimator variance. The bias-variance tradeoff is analyzed for low rank estimators of higher order statistics using a tensor product formulation for the moments and cumulants. In general the low rank estimators have a larger bias and smaller variance than the corresponding full rank estimator. Often a tremendous reduction in variance is obtained in exchange for a slight increase in bias. This makes the low rank estimators extremely useful for signal processing algorithms based on sample estimates of the higher order statistics. The low rank estimators also offer considerable reductions in the computational complexity of such algorithms. The design of subspaces to optimize the tradeoffs between bias, variance, and computation is discussed and a noisy input, noisy output system identification problem is used to illustrate the results.

I. Introduction

Cumulants have received significant attention recently in the signal processing community because of their insensitivity to Gaussian noise. The third and higher order cumulants of a Gaussian random variable are zero, so cumulants offer a mechanism for designing algorithms that are not affected by Gaussian signals. Examples of applications include system identification [2, 3], signal reconstruction [15, 16], detection and parameter estimation [6, 7], bearing estimation [19, 11], blur identification [17], geophysics [13], and acoustics [20]. However, in general, very large data records are required to obtain accurate estimates of cumulants and the computational burden of cumulant based algorithms is often severe. These difficulties have limited the practical application of cumulant based methods.

[†] Supported by the Army Research Office under Grant DAAH04-93-G-0208. [‡] Supported in part by the National Science Foundation under Award MIP-895 8559, Army Research Office under Grant DAAH04-93-G-0208.

In this paper we discuss the application of rank reduction and subspace methods to higher order statistics based algorithms. Rank reduction methods have been applied extensively to signal processing problems based on second order statistics and provide a general principle for trading increased estimator bias for lower estimator variance [4]. While this is a general principle, the details of its application vary since each problem brings its own definitions for bias and variance. Here we focus on low rank estimators for higher order statistics. The variance of the low rank estimator can be considerably less than the variance of the corresponding full rank estimator. However, the cost of this reduced variability is an estimator bias associated with the component of the true higher order statistic that does not lie in the subspace corresponding to the low rank estimator. Hence, we trade bias for reduced variance in the estimated statistics. This is advantageous when the variance is the dominant effect, which occurs with short data records and low signal to noise ratio. Furthermore, by performing the signal processing in the subspace associated with the low rank estimator we attain significant reductions in computational burden.

We begin by developing a matrix framework for moments and cumulants that is based on the tensor or Kronecker product. The algebra of tensor products offers an elegant and mathematically tractable framework for treating the transformations associated with subspaces and low rank estimators. After developing low rank estimators using linear transformations, we address the bias-variance tradeoff. The variance of low rank estimators is compared to that of conventional moment and cumulant estimators. When the data consists of a weak non-Gaussian signal in additive Gaussian noise, we show that the ratio of the variance of the low rank (rank Q) n^{th} order moment estimator to that of the full rank (rank M) estimator is $\mathcal{O}(\frac{Q^n}{M^n})$. Hence, the reduction in variance can be quite large. Next, we discuss several strategies for choosing the subspace based on prior information about the signal characteristics. The goal of subspace design is to optimize the tradeoff between bias, variance, and computation. Lastly, these results are applied to the noisy input, noisy output, linear system identification problem studied by Giannakis and Dantawate [2].

II. Vector and Matrix Representation of Moments and Cumulants

Boldface upper and lower case symbols are used to represent matrix and vector quantities, respectively. Superscript T denotes transposition. The notation $x(k)$ represents the k^{th} value in the zero mean, real valued, stationary time series x. Define the M-dimensional random vector $\mathbf{x} = [x(0)\ x(1)\ \ldots\ x(M-1)]^T$. Wherever necessary we assume that all required moments exist and are finite. In general, we assume $\mathrm{E}|x(k)|^{2n} < \infty$.

We adopt a time series notation for simplicity. The extension of our results to array processing and other applications is particularly straightforward using the vector and matrix representation.

The Kronecker (tensor) product is used extensively in this paper. The Kronecker product of matrices (or vectors) $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$. Define the n-fold Kronecker product of $\mathbf{x}$ as $\mathbf{x}^{(n)} = \underbrace{\mathbf{x} \otimes \mathbf{x} \ldots \otimes \mathbf{x}}_{n \text{ times}}$. Properties of the Kronecker product are explored in depth in [1].

A. Vector and matrix representation for higher order moments

The second order moments of $x(k)$ and $\mathbf{x}$ are written :

$$\begin{aligned} \mathrm{Cov}(x,l) = \quad m_{2,x}(l) &= \mathrm{E}\{x(k)\ x(k+l)\}, \\ \mathrm{Cov}(\mathbf{x}) = \quad \mathbf{M}_{2,\mathbf{x}} &= \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\}. \end{aligned} \tag{1}$$

We define a vector version of $\mathbf{M}_{2,\mathbf{x}}$ using the Vec operator [1]:

$$\mathbf{m}_{2,\mathbf{x}} = \mathrm{Vec}(\mathbf{M}_{2,\mathbf{x}}) = \mathrm{E}\{\mathbf{x}^{(2)}\}. \tag{2}$$

The vector and matrix representations for higher moments are generalizations of the second order case. The n^{th} order moments of x are written :

$$m_{n,x}(l_1, l_2, \ldots, l_{n-1}) = \mathrm{E}\{x(k)\ x(k+l_1)\ x(k+l_2) \ldots x(k+l_{n-1})\}. \tag{3}$$

In matrix and vector notation we have :

$$\begin{aligned} \mathbf{M}_{n,\mathbf{x}} &= \mathrm{E}\{\mathbf{x}^{(n-1)} \otimes \mathbf{x}^T\}, \\ \mathbf{m}_{n,\mathbf{x}} &= \mathrm{Vec}(\mathbf{M}_{n,\mathbf{x}}) = \mathrm{E}\{\mathbf{x}^{(n)}\}. \end{aligned} \tag{4}$$

In certain cases, it is more convenient to use the matrix form rather than the vector form and vice-versa. This is the reason for introducing both forms here.

B. Vector and matrix representation for higher order cumulants

The second and third order cumulants are equal to the second and third order moments. The fourth order cumulant is expressed in terms of the fourth and second order moments as

$$\begin{aligned} c_{4,x}(l_1, l_2, l_3) &= \mathrm{E}\{x(k)\ x(k+l_1)\ x(k+l_2)\ x(k+l_3)\} \\ &\quad -\mathrm{E}\{x(k)\ x(k+l_1)\} \cdot \mathrm{E}\{x(k+l_2)\ x(k+l_3)\} \\ &\quad -\mathrm{E}\{x(k)\ x(k+l_2)\} \cdot \mathrm{E}\{x(k+l_3)\ x(k+l_1)\} \\ &\quad -\mathrm{E}\{x(k)\ x(k+l_3)\} \cdot \mathrm{E}\{x(k+l_1)\ x(k+l_2)\}. \end{aligned} \tag{5}$$

n^{th} order cumulants of x can be defined in terms of the joint moments of x up to order n. The following proposition establishes matrix and vector representations for the fourth order cumulant.

Proposition 1:

- The fourth order cumulant matrix of $\mathbf{x}$, $\mathbf{C}_{4,\mathbf{x}}$, is given by :

$$\mathbf{C}_{4,\mathbf{x}} = \mathrm{E}\{\mathbf{x}^{(3)} \otimes \mathbf{x}^T\} + \mathbf{\Gamma}_M \cdot \mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\}, \tag{6}$$

  where $\mathbf{\Gamma}_M$ is the sum of three $M^3 \times M^3$ permutation matrices and depends only on M, the size of $\mathbf{x}$.

- The fourth order cumulant vector of $\mathbf{x}$, $\mathbf{c}_{4,\mathbf{x}}$, is given by :

$$\mathbf{c}_{4,\mathbf{x}} = \mathrm{E}\{\mathbf{x}^{(4)}\} + \mathbf{\Pi}_M \cdot \mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x}^{(2)}\}, \tag{7}$$

  where $\mathbf{\Pi}_M$ is the sum of three $M^4 \times M^4$ permutation matrices and depends only on M, the size of $\mathbf{x}$.

$\square$

The proof of the proposition and the definition of matrices $\mathbf{\Pi}_M$ and $\mathbf{\Gamma}_M$ are given in Appendix A. Proposition 1 relates the fourth order cumulant to the moment matrices and vectors as :

$$\begin{aligned} \mathbf{C}_{4,\mathbf{x}} &= \mathbf{M}_{4,\mathbf{x}} + \mathbf{\Gamma}_M \cdot \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{M}_{2,\mathbf{x}}, \\ \mathbf{c}_{4,\mathbf{x}} &= \mathbf{m}_{4,\mathbf{x}} + \mathbf{\Pi}_M \cdot \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{m}_{2,\mathbf{x}}. \end{aligned} \tag{8}$$

Similar representations can be derived for fifth and higher order cumulants. These derivations are not presented here due to space constraints and the limited applicability of such high order cumulants in signal processing. Throughout the paper we often leave the moment or cumulant order arbitrary with the understanding that we are only interested in the second through fourth order cases. Vector and matrix representations for the moments and cumulants of complex valued random data can also be derived, but are not given here.

Note that the matrix and vector statistics defined above have a very large number of elements. For example, $\mathbf{c}_{n,\mathbf{x}}$ has M^n elements. If $M = 20$, then $\mathbf{c}_{4,\mathbf{x}}$ has 160,000 elements. Due to the inherent symmetry of the Kronecker product, the number of unique elements in $\mathbf{c}_{n,\mathbf{x}}$ is $\binom{n+M-1}{n} = \frac{(n+M-1)!}{n!(M-1)!}$. This number still grows extremely rapidly with dimension. Hence, for $M = 20$ there are 8,855 unique elements in $\mathbf{c}_{4,\mathbf{x}}$. The large dimensionality associated with higher order statistics presents computational difficulties and is strong motivation for the subspace methods described in this paper.

The strength of the matrix and vector moment and cumulant representations given here is analytic tractability. The powerful properties of Kronecker products are available for manipulating and analysing the higher order statistics. Alternate definitions for moment and cumulant matrices are given in [2, 11]. These are chosen specifically for the problem at hand and may be reconstructed from slices or subsets of the more general definitions given here.

C. Linear transformations

Proposition 2: Let $\mathbf{A}$ be a $Q \times M$ $(Q < M)$ deterministic matrix. Define the zero mean, real valued random variable $\mathbf{z}$ as the product of $\mathbf{A}$ and $\mathbf{x}$:

$$\mathbf{z} = \mathbf{A}\ \mathbf{x}. \tag{9}$$

Then the following holds :

$$\begin{aligned} \mathbf{m}_{n,\mathbf{z}} &= \mathbf{A}^{(n)}\ \mathbf{m}_{n,\mathbf{x}}, \\ \mathbf{M}_{n,\mathbf{z}} &= \mathbf{A}^{(n-1)}\ \mathbf{m}_{n,\mathbf{x}}\ \mathbf{A}^T, \\ \mathbf{c}_{n,\mathbf{z}} &= \mathbf{A}^{(n)}\ \mathbf{c}_{n,\mathbf{x}}, \\ \mathbf{C}_{n,\mathbf{z}} &= \mathbf{A}^{(n-1)}\ \mathbf{C}_{n,\mathbf{x}}\ \mathbf{A}^T. \end{aligned} \tag{10}$$

$\square$

The proof is given in Appendix B. It is possible to write these equations because of the properties of the tensor product.

D. Sample Estimators of Moments and Cumulants

In practice moments and cumulants are often unknown and must be estimated from data. The matrix and vector representations simplify identification of sample estimators.

1. Moment Estimators

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_N$ be independent identically distributed observed random vectors. A sample estimate of $\mathbf{m}_{n,\mathbf{x}}$ is

$$\widehat{\mathbf{m}}_{n,\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i^{(n)}. \tag{11}$$

It is easy to verify that $\widehat{\mathbf{m}}_{n,\mathbf{x}}$ is an unbiased estimator of $\mathbf{m}_{n,\mathbf{x}}$. The matrix sample estimate $\widehat{\mathbf{M}}_{n,\mathbf{x}}$ is defined analogously.

2. Cumulant Estimators

A natural estimator of the fourth order cumulant is obtained by replacing the expected values in (8) by the corresponding sample averages :

$$\widehat{\mathbf{c}}_{4,\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i^{(4)} + \mathbf{\Pi}_M \cdot (\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i^{(2)}) \otimes (\frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j^{(2)}). \tag{12}$$

However, we find that this estimator is biased for finite N. Taking the expected value and performing some algebra gives

$$\begin{aligned} \mathrm{E}\{\widehat{\mathbf{c}}_{4,\mathbf{x}}\} &= \mathbf{m}_{4,\mathbf{x}} + \mathbf{\Pi}_M \cdot (\frac{N-1}{N}\mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{m}_{2,\mathbf{x}} + \frac{1}{N}\mathbf{m}_{4,\mathbf{x}}), \\ &= \mathbf{c}_{4,\mathbf{x}} + \frac{1}{N}\mathbf{\Pi}_M \cdot (\mathbf{m}_{4,\mathbf{x}} - \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{m}_{2,\mathbf{x}}). \end{aligned} \tag{13}$$

The bias is $\mathcal{O}(N^{-1})$ and thus the estimator is asymptotically unbiased. An unbiased estimator is easily derived using the vector notation. Specifically let :

$$\widehat{\mathbf{c}}'_{4,\mathbf{x}} = \frac{1}{N}(\mathbf{I} - \frac{1}{N-1} \cdot \mathbf{\Pi}_M\,) \cdot \sum_{i=1}^{N}\mathbf{x}_i^{(4)} + \frac{1}{N^2-N} \cdot \mathbf{\Pi}_M \cdot \sum_{i,j=1}^{N}\mathbf{x}_i^{(2)} \otimes \mathbf{x}_j^{(2)}. \tag{14}$$

It is easily verified that $\mathrm{E}\{\widehat{\mathbf{c}}'_{4,\mathbf{x}}\} = \mathbf{c}_{4,\mathbf{x}}$.

E. Subspace Notation and Algebra

Let $\mathbf{T}$ be an $M \times Q$ $(Q < M)$ matrix whose Q orthonormal columns span an Q dimensional approximation subspace $\mathcal{U} \subset \mathbb{R}^M$. Define the zero-mean real-valued random variable $\mathbf{z}$ as the coordinates of $\mathbf{x}$ in $\mathcal{U}$:

$$\mathbf{z} = \mathbf{T}^T\,\mathbf{x}. \tag{15}$$

$\mathbf{z}$ is termed the subspace data, since it is the representation of $\mathbf{x}$ in the subspace $\mathcal{U}$. The goal is to use $\mathbf{z}$ in the higher order signal processing algorithm of interest instead of $\mathbf{x}$. The subspace moments and cumulants are defined as :

$$\begin{aligned} \mathbf{m}_{n,\mathbf{z}} &= \mathrm{E}\{\mathbf{z}^{(n)}\}, \\ \mathbf{M}_{n,\mathbf{z}} &= \mathrm{E}\{\mathbf{z}^{(n-1)} \otimes \mathbf{z}^T\}, \\ \mathbf{c}_{4,\mathbf{z}} &= \mathrm{E}\{\mathbf{z}^{(4)}\} + \mathbf{\Pi}_Q \cdot \mathrm{E}\{\mathbf{z}^{(2)}\} \otimes \mathrm{E}\{\mathbf{z}^{(2)}\}, \\ \mathbf{C}_{4,\mathbf{z}} &= \mathrm{E}\{\mathbf{z}^{(3)} \otimes \mathbf{z}^T\} + \mathbf{\Gamma}_Q \cdot \mathrm{E}\{\mathbf{z}^{(2)}\} \otimes \mathrm{E}\{\mathbf{z} \otimes \mathbf{z}^T\}. \end{aligned} \tag{16}$$

It is useful for analysis and algorithm development to express the subspace data, $\mathbf{z}$, in the original full dimensional space. This is accomplished by projecting $\mathbf{x}$ onto the subspace $\mathcal{U}$. Let $\mathbf{P} = \mathbf{T}\;\mathbf{T}^T$ be a projection matrix onto $\mathcal{U}$ and define $\tilde{\mathbf{x}}$:

$$\tilde{\mathbf{x}} = \mathbf{P} \cdot \mathbf{x} = \mathbf{T} \cdot \mathbf{z}. \tag{17}$$

The moments and cumulants of $\tilde{\mathbf{x}}$: $\mathbf{m}_{n,\tilde{\mathbf{x}}}$, $\mathbf{M}_{n,\tilde{\mathbf{x}}}$, $\mathbf{c}_{n,\tilde{\mathbf{x}}}$, $\mathbf{C}_{n,\tilde{\mathbf{x}}}$, are all defined analogously to (16). We refer to $\tilde{\mathbf{x}}$ as low rank data and its statistics as low rank moments or cumulants to distinguish it from the subspace data and statistics associated with $\mathbf{z}$.

The moments and cumulants of $\mathbf{x}$ are now related to the moments and cumulants of $\mathbf{z}$ and $\tilde{\mathbf{x}}$. Let $\mathbf{G}_{n,\mathbf{x}}$ denote either the cumulant matrix $\mathbf{C}_{n,\mathbf{x}}$ or moment matrix $\mathbf{M}_{n,\mathbf{x}}$ and $\mathbf{g}_{n,\mathbf{x}}$ the corresponding cumulant or moment vector. Application of Proposition 2 yields the following :

- The vector moments or cumulants of $\mathbf{x}$, $\mathbf{z}$ and $\tilde{\mathbf{x}}$ are related by :

$$\begin{aligned} \mathbf{g}_{n,\mathbf{z}} &= \mathbf{T}^{(n)^T} \cdot \mathbf{g}_{n,\mathbf{x}}, \\ \mathbf{g}_{n,\tilde{\mathbf{x}}} &= \mathbf{T}^{(n)} \cdot \mathbf{g}_{n,\mathbf{z}}, \\ &= \mathbf{P}^{(n)} \cdot \mathbf{g}_{n,\mathbf{x}}. \end{aligned} \tag{18}$$

- The matrix moments or cumulants of $\mathbf{x}$, $\mathbf{z}$ and $\tilde{\mathbf{x}}$ are related by :

$$\begin{aligned} \mathbf{G}_{n,\mathbf{z}} &= \mathbf{T}^{(n-1)^T} \cdot \mathbf{G}_{n,\mathbf{x}} \cdot \mathbf{T}, \\ \mathbf{G}_{n,\tilde{\mathbf{x}}} &= \mathbf{T}^{(n-1)} \cdot \mathbf{G}_{n,\mathbf{z}} \cdot \mathbf{T}^T, \\ &= \mathbf{P}^{(n-1)} \cdot \mathbf{G}_{n,\mathbf{x}} \cdot \mathbf{P}. \end{aligned} \tag{19}$$

III. Motivations for Low Rank Estimation in Higher Order Signal Processing

There are several advantages to implementing a signal processing algorithm in terms of subspace data, $\mathbf{z}$, rather than the full space data $\mathbf{x}$. Perhaps the most obvious advantage is a tremendous reduction in computational complexity. There are $\binom{n+M-1}{n}$ unique statistics in $\mathbf{m}_{n,\mathbf{x}}$, but only $\binom{n+Q-1}{n}$ unique statistics in $\mathbf{m}_{n,\mathbf{z}}$. The ratio of unique subspace to full space statistics is approximately $(\frac{Q}{M})^n$. If $\frac{Q}{M} = 0.1$ and $n = 4$ then the subspace offers a factor of 10,000 fewer unique statistics. The exact computational savings associated with a signal processing algorithm dependent on how the algorithm utilizes the higher order statistics.

A second very significant advantage of subspace processing is the reduction in variance associated with estimating the much smaller dimensioned subspace moment or cumulant matrices from finite data records. Reduction in the variability of the moment or cumulant estimates results in reduced variance of the signal processing algorithm. The reductions in variance and computation are obtained at the expense of a bias. The low rank or subspace moment or cumulant estimates are biased because information is discarded when the data is mapped into the subspace. The impact of both bias and variance on the signal processing algorithm depends on the specific algorithm employed. Thus, subspace processing leads to a tradeoff between bias and variance. As in the second order case [4], each problem brings its own manifestation of bias and variance effects. Ideally, the subspace is chosen to optimize the tradeoff between bias, variance, and computation for each signal processing algorithm. In this paper we take a more general view and address the bias and variance effects on the estimators of higher order statistics themselves. This provides motivation for the subspace philosophy and a starting point for analysis of these effects in specific algorithms.

In this section we compare the variance of the low rank and full rank moment estimators. The third order cumulant is equal to the third order moment for zero-mean signals. The fourth order cumulant is a function of both the fourth and second order moments and the variance calculation for the fourth order cumulant estimator is complicated because of this. However, understanding the variance reduction associated with the moment estimators that are used in the fourth order cumulant estimators provides insight into the performance of the cumulant estimators as well.

Evaluation of the bias and variance associated with subspace processing is performed by comparing the unbiased estimate $\widehat{\mathbf{m}}_{n,\mathbf{x}}$ to the corresponding low rank estimate $\mathbf{P}^{(n)}\ \widehat{\mathbf{m}}_{n,\mathbf{x}}$. The mean of $\mathbf{P}^{(n)}\ \widehat{\mathbf{m}}_{n,\mathbf{x}}$ is $\mathbf{P}^{(n)}\ \mathbf{m}_{n,\mathbf{x}}$. Hence, the subspace estimate is biased if $\mathbf{m}_{n,\mathbf{x}}$ does not lie in the space spanned by the columns of $\mathbf{P}^{(n)}$. This is a fairly restrictive condition. However it is often possible to choose $\mathbf{P}$ *a priori* so that the bias is insignificant, that is, so that $\mathbf{P}^{(n)}\ \mathbf{m}_{n,\mathbf{x}}$ is very close to $\mathbf{m}_{n,\mathbf{x}}$. The advantage of using $\mathbf{P}^{(n)}\ \widehat{\mathbf{m}}_{n,\mathbf{x}}$ over $\widehat{\mathbf{m}}_{n,\mathbf{x}}$ is that the low rank estimator has lower variance. Methods for choosing the subspace to minimize this bias are discussed in the following section. Here we derive expressions for the bias and the variance, demonstrating the reduced variance of the low rank estimator.

Begin with the mean squared error (MSE) of the low rank estimator $\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}}$

$$\| \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}} \|^2 \triangleq \mathrm{tr}(\mathrm{E}\{(\mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}})(\mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}})^T\}). \tag{20}$$

The MSE is expressed as the sum of a squared bias term and variance term as follows.

$$\begin{aligned} \| \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}} \|^2 &= \| \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\} + \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\} - \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}} \|^2, \\ &= \| \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\} \|^2 + \| \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\} \|^2, \end{aligned} \tag{21}$$

since

$$\mathrm{tr}(\mathrm{E}\{(\mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\})(\mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\} - \mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}})^T\}) = 0.$$

The first term in (21), $\| \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\} \|^2$, is the squared bias of the estimator. The norm here simply becomes the Euclidean norm because $\mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathrm{E}\{\widehat{\mathbf{m}}_{n,\mathbf{x}}\}$ is deterministic. The sample moment estimator $\widehat{\mathbf{m}}_{n,\mathbf{x}}$ is unbiased so we have

$$\mathrm{Bias}^2(\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}}) = \| \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\mathbf{m}_{n,\mathbf{x}} \|_2^2 . \quad (22)$$

The variance of the estimator is the second term of (21),

$$\mathrm{var}(\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}}) = \| \mathbf{P}^{(n)}\ \mathbf{m}_{n,\mathbf{x}} - \mathbf{P}^{(n)}\ \widehat{\mathbf{m}}_{n,\mathbf{x}} \|^2 . \quad (23)$$

Assuming $\widehat{\mathbf{m}}_{n,\mathbf{x}}$ is given by the sample estimate formed from iid data vectors, (23) simplifies to :

$$\mathrm{var}(\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}}) = \frac{1}{N}(\mathrm{E}\{(\mathbf{x}^T\ \mathbf{P}\ \mathbf{x})^{(n)}\} - \mathbf{m}_{n,\mathbf{x}}^T\ \mathbf{P}^{(n)}\ \mathbf{m}_{n,\mathbf{x}}). \quad (24)$$

In the full rank case, $\mathbf{P} = \mathbf{I}$ and we have

$$\mathrm{var}(\widehat{\mathbf{m}}_{n,\mathbf{x}}) = \frac{1}{N}(\mathrm{E}\{(\mathbf{x}^T\ \mathbf{x})^{(n)}\} - \mathbf{m}_{n,\mathbf{x}}^T\ \mathbf{m}_{n,\mathbf{x}}). \quad (25)$$

Further comparison of full space and subspace variances requires specification of the statistics of $\mathbf{x}$.

We now examine a common special case in which $\mathbf{x} = \mathbf{s} + \mathbf{w}$. Here $\mathbf{s}$ is a zero-mean signal with non-zero n^{th} moment and $\mathbf{w}$ is a zero-mean Gaussian process that is independent of the signal. The third order and fourth order cumulant are insensitive to the Gaussian noise and thus may be used to separate the signal $\mathbf{s}$ from noise $\mathbf{w}$. However, the variance of the sample estimators is dependent on $\mathbf{w}$, a fact that may prevent separation of signal and noise in practice.

Assume the subspace $\mathcal{U}$ is well chosen so that the signal lies in the subspace spanned by the columns of $\mathbf{P}$, that is $\mathbf{P}\ \mathbf{s} = \mathbf{s}$. At high signal to noise ratio, the second and higher moments of the signal $\mathbf{s}$ dominate those of $\mathbf{w}$ and we have:

$$\begin{aligned} \mathrm{var}(\mathbf{P}^{(n)}\ \widehat{\mathbf{m}}_{n,\mathbf{x}}) &\approx \frac{1}{N}(\mathrm{E}\{(\mathbf{s}^T\ \mathbf{P}\ \mathbf{s})^{(n)}\} - \mathbf{m}_{n,\mathbf{s}}^T\ \mathbf{P}\ \mathbf{m}_{n,\mathbf{s}}), \\ &= \frac{1}{N}(\mathrm{E}\{(\mathbf{s}^T\ \mathbf{s})^{(n)}\} - \mathbf{m}_{n,\mathbf{s}}^T\ \mathbf{m}_{n,\mathbf{s}}). \end{aligned} \quad (26)$$

Hence, the low rank and full rank moment estimates have equivalent variance at high SNR. At low SNR the moments of the noise dominate those of the signal and we have:

$$\mathrm{var}(\mathbf{P}^{(n)}\ \widehat{\mathbf{m}}_{n,\mathbf{x}}) \approx \frac{1}{N}(\mathrm{E}\{(\mathbf{w}^T\ \mathbf{P}\ \mathbf{w})^{(n)}\} - \mathbf{m}_{n,\mathbf{w}}^T\ \mathbf{P}\ \mathbf{m}_{n,\mathbf{w}}), \quad (27)$$

where $\mathbf{m}_{n,\mathbf{w}}$ is the n^{th} order moment vector of the noise $\mathbf{w}$ (*i.e.*, $\mathbf{m}_{n,\mathbf{w}} = \mathrm{E}\{\mathbf{w}^{(n)}\}$). Now consider the ratio

$$\frac{\mathrm{var}(\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}})}{\mathrm{var}(\widehat{\mathbf{m}}_{n,\mathbf{x}})} \approx \frac{\mathrm{E}\{(\mathbf{w}^T\mathbf{P}\mathbf{w})^n\} - \mathbf{m}_{n,\mathbf{w}}^T\mathbf{P}^{(n)}\mathbf{m}_{n,\mathbf{w}}}{\mathrm{E}\{(\mathbf{w}^T\mathbf{w})^n\} - \mathbf{m}_{n,\mathbf{w}}^T\mathbf{m}_{n,\mathbf{w}}}. \tag{28}$$

The quantities $\mathrm{E}\{(\mathbf{w}^T\mathbf{P}\mathbf{w})^n\}$ and $\mathrm{E}\{(\mathbf{w}^T\mathbf{w})^n\}$ involve the $2n^{\text{th}}$ moment of $\mathbf{w}$. $\mathbf{m}_{n,\mathbf{w}}^T\mathbf{P}^{(n)}\mathbf{m}_{n,\mathbf{w}}$ and $\mathbf{m}_{n,\mathbf{w}}^T\mathbf{m}_{n,\mathbf{w}}$ involve the square of the n^{th} order moment of $\mathbf{w}$. For a Gaussian random vector, the $2n^{\text{th}}$ order moment is much larger than the square of the n^{th} order moment[1]. From this, it follows that $E\{(\mathbf{w}^T\mathbf{P}\mathbf{w})^n\} \gg \mathbf{m}_{n,\mathbf{w}}^T\mathbf{P}^{(n)}\mathbf{m}_{n,\mathbf{w}}$, $E\{(\mathbf{w}^T\mathbf{w})^n\} \gg \mathbf{m}_{n,\mathbf{w}}^T\mathbf{m}_{n,\mathbf{w}}$ and we have

$$\frac{\mathrm{var}(\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}})}{\mathrm{var}(\widehat{\mathbf{m}}_{n,\mathbf{x}})} \approx \frac{\mathrm{E}\{(\mathbf{w}^T\mathbf{P}\mathbf{w})^n\}}{\mathrm{E}\{(\mathbf{w}^T\mathbf{w})^n\}}. \tag{29}$$

Under the assumption that $\mathbf{w}$ is a zero-mean Gaussian noise, $\mathrm{E}\{(\mathbf{w}^T\mathbf{P}\mathbf{w})^{(n)}\}$ is computed using the following result.

Theorem 1: For $\mathbf{w} \sim N(\mathbf{0}, \mathbf{\Sigma})$,

$$\mathrm{E}\{(\mathbf{w}\mathbf{P}\mathbf{w}^T)^n\} = \sum_{r_1=0}^{n-1}\left[\binom{n-1}{r_1} g_{(n-1-r_1)} \sum_{r_2=0}^{r_1-1}\left[\binom{r_1-1}{r_2} g_{(r_1-1-r_2)} \cdots\right]\right] \triangleq G(\mathbf{P}, \mathbf{\Sigma}, n),$$

where

$$g_{(k)} = 2^k k!\,\mathrm{tr}(\{\mathbf{P}\ \mathbf{\Sigma}\}^{k+1}),\ \ k = 0, 1, 2, \ldots$$

□

Theorem 1 is a direct application of Theorem 3.2b.2 in [12]. Combining Theorem 1 and (29) we have the following result.

Theorem 2: If $\mathbf{w} \sim N(\mathbf{0}, \mathbf{\Sigma})$, at low SNRs

$$\frac{\mathrm{var}(\mathbf{P}^{(n)}\widehat{\mathbf{m}}_{n,\mathbf{x}})}{\mathrm{var}(\widehat{\mathbf{m}}_{n,\mathbf{x}})} \approx \frac{G(\mathbf{P}, \mathbf{\Sigma}, n)}{G(\mathbf{I}, \mathbf{\Sigma}, n)}. \tag{30}$$

□

Notice that the reduction in variance is proportional to the $g_{(k)}$ defined in Theorem 1. The $g_{(k)}$ in the numerator depend on the projection matrix and noise statistics through the term $\mathrm{tr}(\{\mathbf{P}\mathbf{\Sigma}\}^{k+1}) = \mathrm{tr}(\{\mathbf{P}\mathbf{\Sigma}\mathbf{P}\}^{k+1})$. Let λ_l^P, $l = 1, 2, \ldots, Q$ denote the Q non zero eigenvalues of $\mathbf{P}\mathbf{\Sigma}\mathbf{P}$ ordered so that $\lambda_1^P \geq \lambda_2^P \geq \ldots \geq \lambda_Q^P \geq 0$. We have

$$\mathrm{tr}(\{\mathbf{P}\mathbf{\Sigma}\mathbf{P}\}^{k+1}) = \sum_{l=1}^{Q}(\lambda_l^P)^{k+1}. \tag{31}$$

[1]For example, if $n = 3$ the third order moment is zero whereas the sixth moment is non-zero. If $n = 4$, the 8^{th} order moment is more than 10 times the square of the 4^{th} order moment.

The $g_{(k)}$ involved in the denominator of the variance ratio in Theorem 2 are dependent on $\mathrm{tr}(\mathbf{\Sigma}^{k+1})$. Letting λ_l, $l = 1, 2, \ldots M$ be the M eigenvalues of $\mathbf{\Sigma}$ ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M \geq 0$. We have

$$\mathrm{tr}(\mathbf{\Sigma}^{k+1}) = \sum_{l=1}^{M} (\lambda_l)^{k+1}. \tag{32}$$

There are only Q eigenvalues in the $g_{(k)}$ associated with the numerator and we know $\lambda_l^P \leq \lambda_l$ since $\mathbf{P}$ is a projection. Hence, the numerator of (30) is guaranteed to be less than or equal to the denominator. Therefore the variance of the low rank estimator is less than or equal to the variance of the full rank estimator. The exact reduction in variance depends on the relative noise power in the subspace.

Examination of the special case where $\mathbf{w} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ provides further insight. In this case, $\mathbf{w}^T \mathbf{P} \mathbf{w}$ is a chi-squared random variable with Q degrees of freedom. The n-th order moment of a χ_Q^2 random variable is given by

$$\mu_n = 2^n \frac{\Gamma(\frac{Q}{2} + n)}{\Gamma(\frac{Q}{2})}. \tag{33}$$

Using this fact, the result of Theorem 2 simplifies as follows.

Theorem 3: If $\mathbf{w} \sim N(\mathbf{0}, \mathbf{I})$, at low SNR

$$\begin{aligned} \frac{\mathrm{var}(\mathbf{P}^{(n)} \widehat{\mathbf{m}}_{n,\mathbf{x}})}{\mathrm{var}(\widehat{\mathbf{m}}_{n,\mathbf{x}})} &\approx \frac{\Gamma(\frac{M}{2})\,\Gamma(\frac{Q}{2} + n)}{\Gamma(\frac{Q}{2})\,\Gamma(\frac{M}{2} + n)}, \\ &= \frac{Q(Q+2)\cdots(Q+2(n-1))}{M(M+2)\cdots(M+2(n-1))} \approx (\tfrac{Q}{M})^n. \end{aligned}$$

□

Theorem 3 shows that in the case of white Gaussian noise at low SNR, the variance reduction obtained using the low rank moment estimator is $\mathcal{O}(\frac{Q^n}{M^n})$. The sample cumulant estimator and low rank sample cumulant estimator are formed from the sample moment and low rank sample moment estimators respectively. Hence, similar reductions in variance are achieved by the low rank sample cumulant estimator.

IV. Subspace Design

The goal of subspace design is determining the transformation $\mathbf{T}$ to optimize the tradeoff between bias, variance, and computation. Since all three of these factors depend on the signal processing algorithm of interest, the best results are obtained by developing design procedures for specific algorithms. However, often very good results are obtained by generic design methods such as those discussed in this section. The methods described here attempt to minimize the bias of the subspace moment or cumulant using the smallest possible subspace dimension. This approach is based on the observation that variance and computation tend to increase with dimension.

A. Prior signal information

Prior information about the signal can often be used to design a subspace that results in small bias. The general idea here is to find a $\mathbf{T}$ and corresponding projection matrix $\mathbf{P}$ such that

$$\mathbf{P}\ \mathbf{s} = \mathbf{s} \text{ with probability one} \tag{34}$$

where $\mathbf{s}$ is the signal of interest. If this is satisfied, then

$$\begin{aligned} \mathbf{m}_{n,\mathbf{x}} &= \mathbf{P}^{(n)}\ \mathbf{m}_{n,\mathbf{x}} \\ \mathbf{c}_{n,\mathbf{x}} &= \mathbf{P}^{(n)}\ \mathbf{c}_{n,\mathbf{x}} \end{aligned} \tag{35}$$

and we obtain zero bias. In practice zero bias may not be necessary and we seek an approximate solution to (34).

For example, if the signal of interest is known to be contained within a frequency band $\Omega \subset [-0.5, 0.5]$, then we choose a $\mathbf{T}$ whose columns span this frequency band. This is accomplished by choosing $\mathbf{T}$ to minimize the mean squared error on the band

$$\min_{\mathbf{T}} \int_{\Omega} \| \mathbf{d}(f) - \mathbf{T}\ \mathbf{T}^T\ \mathbf{d}(f) \|_2^2\ df \text{ subject to } \mathbf{T}^T\ \mathbf{T} = \mathbf{I} \tag{36}$$

where $\mathbf{d}(f) = [1\ e^{-j\,2\pi f}\ e^{-j\,4\pi f} \ldots e^{-j\,2(N-1)\pi f}]^H$. This problem can be written

$$\min_{\mathbf{T}} \mathrm{tr}(\mathbf{T}^T\ \mathbf{K}\ \mathbf{T}) \tag{37}$$

where $\mathbf{K} = \int_{\Omega} \mathbf{d}(f)\ \mathbf{d}(f)^H\, df$. The solution is to set $\mathbf{T}$ equal to the eigenvectors of $\mathbf{K}$ associated with significant eigenvalues. The number of significant eigenvalues of $\mathbf{K}$ is proportional to the time bandwidth product [18]. Note that (34) holds approximately if the insignificant eigenvalues are near zero.

If the signal's spectral characteristics are known, then we may weight the mean squared error by the signal power spectrum, $S(f)$, to obtain a more efficient subspace. In this case we solve

$$\min_{\mathbf{T}} \mathrm{tr}(\mathbf{T}^T\ \mathbf{R}\ \mathbf{T}) \tag{38}$$

where $\mathbf{R} = \int_{\Omega}\ \mathrm{S}(f)\mathbf{d}(f)\ \mathbf{d}(f)^H df$. Again $\mathbf{T}$ is set to the eigenvectors of $\mathbf{R}$ corresponding to significant eigenvalues. Signals with strongly colored spectra generally have fewer significant eigenvalues.

B. Prior Cumulant or Moment Information

If the cumulant or moment of the signal is known, we may choose the subspace to minimize the bias for a given dimension. Again, let $\mathbf{G}_{n,\mathbf{x}}$ denote either the cumulant matrix $\mathbf{C}_{n,\mathbf{x}}$ or moment matrix $\mathbf{M}_{n,\mathbf{x}}$ and $\mathbf{g}_{n,\mathbf{x}}$ the corresponding cumulant or moment vector. The squared bias is given by

$$\| \mathbf{g}_{n,\mathbf{x}} - \mathbf{P}^{(n)} \mathbf{g}_{n,\mathbf{x}} \|_2^2, \tag{39}$$

which is equivalently written as

$$\| \mathbf{G}_{n,\mathbf{x}} - \mathbf{P}^{(n-1)} \mathbf{G}_{n,\mathbf{x}} \mathbf{P} \|_F^2, \tag{40}$$

where $\| \cdot \|_F^2$ is the Frobenius matrix norm. Finding $\mathbf{P}$ to directly minimize the bias is difficult because the problem is non-linear and non-convex. The following theorem presents a near optimal solution for minimizing bias.

Theorem 4 : $\mathbf{G}_{n,\mathbf{x}}$ has the singular value decomposition

$$\mathbf{G}_{n,\mathbf{x}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{41}$$

Let $\sigma_1 \geq \ldots \geq \sigma_M \geq 0$, denote the singular values of $\mathbf{G}_{n,\mathbf{x}}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ denote associated orthonormal right singular vectors. Let $\mathbf{T}$ be the matrix whose columns are the Q right singular vectors $\mathbf{v}_1, \ldots, \mathbf{v}_Q$ and let $\mathbf{P}_T$ be the rank Q orthogonal projection matrix $\mathbf{P}_T = \mathbf{T}\mathbf{T}^T$. If $\mathcal{P}_Q$ is the set of all orthogonal projection matrices with rank $\leq Q$, then

$$\sum_{i=Q+1}^{M} \sigma_i^2 \leq \min_{\mathbf{P} \in \mathcal{P}_Q} \| \mathbf{g}_{n,\mathbf{x}} - \mathbf{P}^{(n)} \mathbf{g}_{n,\mathbf{x}} \|_2^2 \leq \| \mathbf{g}_{n,\mathbf{x}} - (\mathbf{P}_T)^{(n)} \mathbf{g}_{n,\mathbf{x}} \|_2^2 \leq n \sum_{i=Q+1}^{M} \sigma_i^2 \tag{42}$$

or, in equivalently in matrix notation,

$$\sum_{i=Q+1}^{M} \sigma_i^2 \leq \min_{\mathbf{P} \in \mathcal{P}_Q} \| \mathbf{G}_{n,\mathbf{x}} - \mathbf{P}^{(n-1)} \mathbf{G}_{n,\mathbf{x}} \mathbf{P} \|_F^2 \leq \| \mathbf{G}_{n,\mathbf{x}} - (\mathbf{P}_T)^{(n-1)} \mathbf{G}_{n,\mathbf{x}} \mathbf{P}_T \|_F^2 \leq n \sum_{i=Q+1}^{M} \sigma_i^2$$

Proof : $\mathbf{G}_{n,\mathbf{x}}$ has the same symmetries as the n^{th} order symmetric Voltera kernel matrix defined in Theorem 2 of [10]. Hence, the proof given in [10] is directly applicable. □

The minimum squared bias associated with a low rank approximation to $\mathbf{G}_{n,\mathbf{x}}$ is bounded below $\sum_{i=Q+1}^{M} \sigma_i^2$. Hence, the low rank estimator $\mathbf{P}_T^{(n)} \mathbf{g}_{n,\mathbf{x}}$ defined in this theorem has a bias that is within a factor $\sqrt{n}$ of the minimum and is therefore near optimal. Unfortunately, the practical utility of this result is limited by the requirement that the cumulant or the moment matrix be known.

V. Application to Linear System Identification

We use the system identification problem studied in [2] by Giannakis and Dandawate to illustrate the results of this paper. Denote the impulse response of the unknown linear time-invariant system by $\mathbf{h}$. If $\{s(k)\}$ is the input to the system, then the output at time k is $v(k) = \mathbf{s}_k^T \mathbf{h}$, where $\mathbf{s}_k = [s(k), \ldots, s(k-m+1)]^T$. We assume that the input is a stationary process and remove the explicit reference to k throughout the remainder of the section. Rather than observing the input $\mathbf{s}$ directly, we assume that we observe a noisy version of $\mathbf{s}$, denoted by $\mathbf{x}$, that is contaminated by an additive, zero-mean, Gaussian noise. We also assume that the observed output is contaminated by an additive, zero-mean, Gaussian noise. If $\mathbf{w}_i$ and w_o represent the Gaussian input and output observation noises, then the observed input and output, respectively, are

$$\mathbf{x} = \mathbf{s} + \mathbf{w}_i, \tag{43}$$

$$y = \mathbf{s}^T \mathbf{h} + w_o. \tag{44}$$

The scenario is depicted in Fig. 1. The goal is to estimate $\mathbf{h}$ given a set of iid observations of $\mathbf{x}$ and y.

The presence of the Gaussian noise introduces a bias into solutions based on second order statistics. An asymptotically unbiased solution may be obtained using the sample third or higher order cumulants [2], since the cumulants are insentitive to Gaussian noise.

Let $\mathbf{h}[i]$ denote the ith element of $\mathbf{h}$. In [2], a third order cumulant based solution is found by solving the matrix equation

$$\mathrm{E}\{y\ x(k+l_1)\ x(k+l_2)\} = \sum_{i=1}^{M} \mathbf{h}[i]\ \mathrm{E}\{x(k-i)\ x(k+l_1)\ x(k+l_2)\}, \tag{45}$$

for various indices l_1 and l_2. These indices are chosen so that a square system of equations is obtained. If the input signal has a symmetric distribution, then the fourth order cumulant must be used since the third order cumulant is zero. In terms of fourth order cumulants and cross cumulants, [2] proposes solving :

$$\mathrm{Cum}\{y, x(k+l_1), x(k+l_2), x(k+l_3)\} = \sum_{i=1}^{M} \mathbf{h}[i]\ \mathrm{Cum}\{x(k-i), x(k+l_1), x(k+l_2), x(k+l_3)\}. \tag{46}$$

for various values l_1, l_2 and l_3.

We reformulate these problems using the matrix and vector representation as follows.

- Third order moment case : multiply both sides of (44) by $\mathbf{x}^{(2)}$ and take expectations to obtain

$$\mathbf{p}_{3,y\mathbf{x}} = \mathbf{C}_{3,\mathbf{x}}\ \mathbf{h}. \tag{47}$$

- Fourth order cumulant case : multiply both sides of (44) by $\mathbf{x}^{(3)} + \mathbf{\Pi}_M \cdot \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{x}$ and take an expectation to obtain

$$\mathbf{p}_{4,y\mathbf{x}} + \mathbf{\Pi}_M \cdot \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{p}_{2,y\mathbf{x}} = \mathbf{C}_{4,\mathbf{x}}\ \mathbf{h} \tag{48}$$

In both cases $\mathbf{p}_{n,y\mathbf{x}} = \mathrm{E}\{y\mathbf{x}^{(n-1)}\}$. The key to expressions (47) and (48) is that in both cases application of the appropriate quantity to the left hand side of (44) and taking expectations produces the desired cumulant. For example, in (47) we compute

$$\begin{aligned} \mathrm{E}\{\mathbf{x}^{(2)}(\mathbf{s}^T\mathbf{h} + w_o)\} &= \mathrm{E}\{\mathbf{x}^{(2)}\mathbf{s}^T\}\mathbf{h} + \mathrm{E}\{\mathbf{x}^{(2)}w_o\}, \\ &= \mathrm{E}\{\mathbf{x}^{(2)}\mathbf{s}^T\}\mathbf{h}, \end{aligned}$$

since $\mathbf{x}$ and w_o are independent and $\mathrm{E}\{w_o\} = 0$. Expanding $\mathrm{E}\{\mathbf{x}^{(2)}\mathbf{s}^T\}$

$$\begin{aligned} \mathrm{E}\{\mathbf{x}^{(2)}\mathbf{s}^T\} &= \mathrm{E}\{(\mathbf{s} + \mathbf{w}_i)^{(2)}\mathbf{s}^T\}, \\ &= \mathrm{E}\{(\mathbf{s}^{(2)} + \mathbf{s} \otimes \mathbf{w}_i + \mathbf{w}_i \otimes \mathbf{s} + \mathbf{w}_i^{(2)})\mathbf{s}^T\}, \\ &= \mathrm{E}\{\mathbf{s}^{(2)}\mathbf{s}^T\} = \mathbf{C}_{3,\mathbf{s}}, \end{aligned}$$

where we have used the fact that $\mathbf{w}_i$ is zero-mean and independent of $\mathbf{s}$. Independence also implies that $\mathbf{C}_{3,\mathbf{x}} = \mathbf{C}_{3,\mathbf{s}} + \mathbf{C}_{3,\mathbf{w}_i}$ and since $\mathbf{w}_i$ is Gaussian $\mathbf{C}_{3,\mathbf{w}_i} = \mathbf{0}$. Hence,

$$\mathrm{E}\{\mathbf{x}^{(2)}\mathbf{s}^T\} = \mathbf{C}_{3,\mathbf{s}} = \mathbf{C}_{3,\mathbf{x}},$$

and (47) is verified. A similar expansion shows that (48) is valid, however it is a bit more tedious because one needs to exploit the structure of the matrix $\mathbf{\Pi}_M$.

If $\mathbf{s}$ lies in a subspace (*i.e.,* $\mathbf{s} = \mathbf{P}\mathbf{s}$, *w.p.*1), then (44) is rewritten

$$\begin{aligned} y &= \mathbf{s}^T\mathbf{P}\mathbf{h} + w_o, \\ &= \mathbf{s}^T\mathbf{T}\mathbf{a} + w_o, \end{aligned} \tag{49}$$

where $\mathbf{a} = \mathbf{T}^T\mathbf{h}$. In this case, we can only identify $\mathbf{a}$, the component of $\mathbf{h}$ that lies in the subspace. Subspace versions of (47) and (48) are obtained by replacing $\mathbf{x}$ with $\mathbf{z} = \mathbf{T}^T\mathbf{x}$ and $\mathbf{h}$ with $\mathbf{a}$. The low rank estimate of $\mathbf{h}$ is given by $\mathbf{Ta}$. Since $\mathbf{s}$ lies in a subspace both $\mathbf{C}_{3,\mathbf{x}}$ and $\mathbf{C}_{4,\mathbf{x}}$ are low rank.

We now compare the performance of the full rank and low rank third order cumulant estimators for the system identification of the 20-tap FIR filter shown in Fig. 2. The input signal $\mathbf{s}$ is generated by passing a zero-mean, non-symmetric, i.i.d. random process through a lowpass Butterworth filter whose frequency response is picture in Fig. 3. The cutoff frequency of this filter is .15 on the normalized frequency band [-.5, .5]. The distribution of the underlying i.i.d. process is $P(x = 1) = 0.1$ and $P(x = -\frac{1}{9}) = 0.9$. The observed input signal $\mathbf{x}$ is obtained by adding Gaussian white noise to $\mathbf{s}$. We consider two scenarios. First, we consider a $N = 500$ data vector estimate with the SNR in $\mathbf{x}$ set at 0dB. Second, the same problem is repeated with $N = 10,000$ data vectors with the SNR = -5dB. In both cases, the measured output y is obtained by adding Gaussian white noise to $\mathbf{h}^T\mathbf{s}$ and the SNR in y is 0dB. Fifty independent trials of both scenarios are employed to evaluate the estimator statistics.

Because the input process is lowpass, a signal subspace is naturally defined. We examine both subspace design methods described in section IV. First, the near optimal signal subspace is computed from the true input cumulant using the method of Section IV-B. Fig. 4 depicts the singular values of $\mathbf{C}_{3,\mathbf{s}}$. Low rank cumulant estimators of rank 2, 4, and 8 are designed using the right singular vectors of $\mathbf{C}_{3,\mathbf{s}}$. These estimators are referred to as *cumulant based* (CB) low rank estimators. The low rank estimators are biased since rank$\mathbf{C}_{3,\mathbf{s}} \geq 8$.

The second method designs the subspace assuming that the signal $\mathbf{s}$ is lowpass using the method of section IV-A with $\Omega = [-0.05, 0.05]$, $\Omega = [-0.10, 0.10]$, and $\Omega = [-0.20, 0.20]$ resulting in three low rank cumulant estimators of rank 2, 4, and 8, respectively. We shall refer to these estimators as *lowpass* (LP) low rank estimators.

Table I and Table II give the estimated bias squared, variance, and MSE for the various cumulant estimators for $n = 500$, SNR= 0dB and $N = 10,000$, SNR= -5dB cases, respectively. All quantities are normalized by the squared norm of the true input cumulant. Theoretically, the bias of the full rank cumulant estimator is zero. The small non-zero values in tables reflect an error associated with using a finite number of trials in the simulation. In all cases, the estimated squared biases are also very close to the corresponding theoretical values. For example, referring to Table I, the true squared bias of the LP rank 2 estimator is 0.7180 and the true squared bias of the CB rank 4 estimator is 0.1532.

The tradeoff between bias and variance is evident in both Tables I and II. As expected, the variance increases and the bias decreases with increasing estimator rank. The rank 8 cumulant estimators have the best performance in the MSE sense and, of the two rank 8 estimators, the CB estimator performs slightly better. The SNR of 0dB is between the high and low SNR cases discussed in Section III. Hence, in that case, we do not expect the degree of improvement in variance predicted at low SNR by our theoretical analysis in Section III. Notice that the reduction in variance, from full rank to low rank estimators, is more dramatic in the -5dB case.

Using each cumulant estimator above, estimates of the system impulse response are derived according

estimator	Bias Squared	Variance	MSE
full rank	0.0081	0.5733	0.5815
LP rank 8	0.0139	0.3081	0.3221
CB rank 8	0.0038	0.3150	0.3188
LP rank 4	0.1758	0.1963	0.3721
CB rank 4	0.1550	0.1961	0.3511
LP rank 2	0.7182	0.0690	0.7872
CB rank 2	0.6285	0.0749	0.7034

Table I. Performance of full rank and low rank estimators of the third order cumulant: $N = 500$, SNR$=$ 0dB.

estimator	Bias Squared	Variance	MSE
full rank	0.0038	0.1902	0.1940
LP rank 8	0.0129	0.0459	0.0589
CB rank 8	0.0018	0.0463	0.0481
LP rank 4	0.1789	0.0195	0.1984
CB rank 4	0.1603	0.0199	0.1802
LP rank 2	0.7128	0.0059	0.7188
CB rank 2	0.6337	0.0066	0.6403

Table II. Performance of full rank and low rank estimators of the third order cumulant: $N = 10,000$, SNR$= -5$dB.

estimator	Bias Squared	Variance	MSE
full rank	0.1640	2.3753	2.5392
LP rank 8	0.2211	1.1457	1.3668
CB rank 8	0.2181	1.1234	1.3414
LP rank 4	0.1895	0.1541	0.3436
CB rank 4	0.1126	0.1467	0.2593
LP rank 2	1.1764	0.0263	1.2028
CB rank 2	1.1260	0.0286	1.1546

Table III. Performance of full rank and low rank estimates of the system impulse response: $N = 500$, SNR= 0dB.

to (47) by replacing the expected values by their respective sample average estimates. Tables III and IV summarize the statistical performance of the system impulse response estimates in the cases $N = 500$, SNR= 0dB and $N = 10,000$, SNR= -5dB, respectively. All quantities are normalized by the squared norm of the impulse response vector $\mathbf{h}$. The bias, variance, and MSE for each system estimator is referenced according to the underlying cumulant estimator. The estimator variance decreases with decreasing rank of the corresponding cumulant. The estimator based on rank 2 cumulant estimates has the smallest variance and the largest bias; again demonstrating the trade-off between bias and variance. The best estimators, in the MSE sense, are those based on the rank 4 low rank cumulant estimates. Of these two, the estimator using the CB low rank cumulant estimate provides the best performance in terms of both bias and variance. Note that best impulse response estimator does not correspond to the best low rank cumulant estimator. This observation supports the point we made previously about obtaining the best performance by designing low rank estimators for the specific problem of interest. To further demonstrate the dramatic improvement obtained using the low rank rather than full rank cumulant estimators for this application consider Fig. 5. Fig. 5 depicts the results of a representative trial of the system identification experiment ($N = 500$, SNR= 0dB). The true impulse response is the solid curve, the full rank cumulant estimator produces the impulse response estimate depicted by the dash-dot curve, and the rank 4 cumulant estimator produces the estimate represented by the dash-dash curve.

estimator	Bias Squared	Variance	MSE
full rank	0.0316	0.5877	0.6193
LP rank 8	0.0597	0.2502	0.3098
CB rank 8	0.0385	0.2118	0.2503
LP rank 4	0.1279	0.0117	0.1396
CB rank 4	0.0353	0.0115	0.0467
LP rank 2	1.0237	0.0001	1.0238
CB rank 2	1.0751	0.0015	1.0766

Table IV. Performance of full rank and low rank estimates of the system impulse response: $N = 10,000$, SNR= -5dB.

VI. Summary

Rank reduction methods offer a general framework for trading bias for variance. Low rank estimators for moments and cumulants are derived in this paper using a tensor product formulation. The algebra of tensor products offers an elegant and mathematically tractable framework for addressing this problem. We develop low rank estimators using linear transformations of the data and address the bias variance tradeoff by comparing the variance of low rank and full rank estimators. The variance reduction attained using low rank estimates can be quite dramatic at low SNR. Furthermore, the computational burden associated with the low rank estimates is much less than the corresponding full rank case. Several strategies for designing the subspace to optimize tradeoffs between bias, variance, and computation are presented. These rely on partial prior knowledge of the signal. The benefits of low rank estimation are illustrated using a system identification problem.

Appendix A
Proof of Proposition 1

Proposition 1:

- The fourth order cumulant matrix of $\mathbf{x}$, $\mathbf{C}_{4,\mathbf{x}}$, is given by :

$$\mathbf{C}_{4,\mathbf{x}} = \mathrm{E}\{\mathbf{x}^{(3)} \otimes \mathbf{x}^T\} + \mathbf{\Gamma}_M \ \mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\} \tag{50}$$

where $\mathbf{\Gamma}_M$ is the sum of three $M^3 \times M^3$ permutation matrices and depends only on M, the size of $\mathbf{x}$.

- The fourth order cumulant vector of $\mathbf{x}$, $\mathbf{c}_{4,\mathbf{x}}$, is given by :

$$\mathbf{c}_{4,\mathbf{x}} = \mathrm{E}\{\mathbf{x}^{(4)}\} + \mathbf{\Pi}_M \ \mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x}^{(2)}\} \tag{51}$$

where $\mathbf{\Pi}_M$ is the sum of three $M^4 \times M^4$ permutation matrices and depends only on M, the size of $\mathbf{x}$.

Proof: We will prove the proposition for the matrix case; the vector case is proved in a similar manner. Before proceeding, we introduce a non-standard indexing convention for the elements of the cumulant matrices. The standard indexing convention for matrices is:

- $\mathbf{C}_{n,\mathbf{x}}[k_1, k_2]$, represents the element in the k_1^{th} row, k_2^{th} column, with $k_1 = 1 \ldots M^{n-1}, k_2 = 1 \ldots M$

The alternate indexing convention for $\mathbf{C}_{n,\mathbf{x}}$ follows naturally from the tensor product formulation:

- $\mathbf{C}_{n,\mathbf{x}}(l_1, l_2, \ldots l_n) = \mathbf{C}_{n,\mathbf{x}}[(l_1 - 1) \ M^{n-2} + \cdots + l_{n-1}, l_n]$

where $l_1, l_2, \ldots, l_n = 1, \ldots, M$. To see why this is a natural indexing convention, consider the nth order moment matrix $\mathbf{M}_{n,\mathbf{x}}(l_1, l_2, \ldots, l_n) = \mathbf{M}_{n,\mathbf{x}}[(l_1 - 1) \ M^{n-2} + \ldots (l_{n-1} - 1), \ l_n] = \mathrm{E}\{\mathbf{x}[l_1]\,\mathbf{x}[l_2] \cdots \mathbf{x}[l_n]\}$. The $(l_1, l_2, \ldots, l_n)$ element is the expected value of the product of the l_1^{th} element of $\mathbf{x}$ with the l_2^{nd} element of $\mathbf{x}, \ldots$, with the l_n^{th} element of $\mathbf{x}$. For simplicity, we refer to the $(l_1 - 1)M^{n-2} + (l_2 - 1)M^{n-3} + \ldots + l_{n-1}$ row of $\mathbf{C}_{n,\mathbf{x}}$ as the $(l_1, l_2, \ldots, l_{n-1})$ row.

The 4th order cumulant is expressed as

$$\begin{aligned} \mathrm{cum}(\mathbf{x}[i], \mathbf{x}[j], \mathbf{x}[k], \mathbf{x}[l]) \ &= \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[j]\ \mathbf{x}[k]\ \mathbf{x}[l]\} \quad - \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[j]\} \cdot \mathrm{E}\{\mathbf{x}[k]\ \mathbf{x}[l]\} \\ & \quad - \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[k]\} \cdot \mathrm{E}\{\mathbf{x}[j]\ \mathbf{x}[l]\} - \mathrm{E}\{\mathbf{x}[j]\ \mathbf{x}[k]\} \cdot \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[l]\} \end{aligned} \tag{52}$$

We will find a matrix $\mathbf{\Gamma}_M$ so that

$$\begin{aligned} \mathrm{cum}(\mathbf{x}[i], \mathbf{x}[j], \mathbf{x}[k], \mathbf{x}[l]) &= \mathbf{C}_{4,\mathbf{x}}(i, j, k, l), \\ &= \mathrm{E}\{\mathbf{x}^{(3)} \otimes \mathbf{x}^T\}(i, j, k, l) + \left(\mathbf{\Gamma}_M \mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\}\right)(i, j, k, l). \end{aligned} \tag{53}$$

Note that first term in (52) is

$$\mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[j]\ \mathbf{x}[k]\ \mathbf{x}[l]\} = \mathbf{M}_{4,\mathbf{x}}(i, j, k, l).$$

This shows that the 4th order moments are properly ordered in $\mathbf{C}_{4,\mathbf{x}}$. The last three terms of equation (52) are all found in the matrix $\mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\}$ at different positions. The matrix $\mathbf{\Gamma}_M$ produces the pairing of second order moments needed to satisfy (53).

Let $\mathbf{\Gamma}_M$ be composed of three permutation matrices $\mathbf{\Gamma}_{1,M}$, $\mathbf{\Gamma}_{2,M}$, and $\mathbf{\Gamma}_{3,M}$ defined as follows.

- $\mathbf{\Gamma}_{1,M}$ is equal to the $(M^3 \times M^3)$ identity matrix
- $\mathbf{\Gamma}_{2,M}$ places the $(i,k,j)^{\text{th}}$ row of $\mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\}$ in the $(i,j,k)^{\text{th}}$ row.
- $\mathbf{\Gamma}_{3,M}$ places the $(j,k,i)^{\text{th}}$ row of $\mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\}$ in the $(i,j,k)^{\text{th}}$ row.

We obtain a simple expression for $\mathbf{\Gamma}_{1,M}$, $\mathbf{\Gamma}_{2,M}$, and $\mathbf{\Gamma}_{3,M}$ by using the notation introduced in [1]. The q-vector which has a "1" in the k^{th} position and zero elsewhere is called the *unit vector* and is denoted $\mathbf{e}_k^{q\times 1}$. The *elementary* $(p \times q)$ *matrix* $\mathrm{E}_{ij}^{p\times q}$ is defined as $\mathrm{E}_{ij}^{p\times q} \triangleq \mathbf{e}_i^{p\times 1}\ \mathbf{e}_j^{q\times 1^T}$ and has a "1" in the $[i,j]$ position, i^{th} row j^{th} column, and zero everywhere else. Define the permutation matrices :

$$
\begin{aligned}
\mathbf{\Gamma}_{1,M} &= \sum_{i,j,k=1}^{M} \mathrm{E}_{ii}^{M\times M} \otimes \mathrm{E}_{jj}^{M\times M} \otimes \mathrm{E}_{kk}^{M\times M}, \\
\mathbf{\Gamma}_{2,M} &= \sum_{i,j,k=1}^{M} \mathrm{E}_{ii}^{M\times M} \otimes \mathrm{E}_{jk}^{M\times M} \otimes \mathrm{E}_{kj}^{M\times M}, \\
\mathbf{\Gamma}_{3,M} &= \sum_{i,j,k=1}^{M} \mathrm{E}_{ij}^{M\times M} \otimes \mathrm{E}_{jk}^{M\times M} \otimes \mathrm{E}_{ki}^{M\times M}.
\end{aligned} \tag{54}
$$

These matrices are square $M^3 \times M^3$ and have precisely a single "1" in each row and column. Now, let

$$
\mathbf{\Gamma}_M \triangleq -(\mathbf{\Gamma}_{1,M} + \mathbf{\Gamma}_{2,M} + \mathbf{\Gamma}_{3,M}). \tag{55}
$$

The matrix representation of the fourth order cumulant is then given by

$$
\mathbf{C}_{4,\mathbf{x}} = \mathrm{E}\{\mathbf{x}^{(3)} \otimes \mathbf{x}^T\} + \mathbf{\Gamma}_M \cdot \mathrm{E}\{\mathbf{x}^{(2)}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\} \tag{56}
$$

To verify (56) consider:

$$
\begin{aligned}
\mathbf{C}_{4,\mathbf{x}}(i,j,k,l) &= \mathrm{E}\{\mathbf{x}^{(3)} \otimes \mathbf{x}^T\}(i,j,k,l) + (\mathbf{\Gamma}_M \cdot \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\})(i,j,k,l), \\
&= \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[j]\ \mathbf{x}[k]\ \mathbf{x}[l]\} - (\mathrm{E}\{\mathbf{x} \otimes \mathbf{x}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\})(i,j,k,l) \\
&\quad -(\sum_{a,b,c=1}^{M} \mathrm{E}_{aa}^{M\times M} \otimes \mathrm{E}_{bc}^{M\times M} \otimes \mathrm{E}_{cb}^{M\times M} \cdot \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\})(i,j,k,l) \\
&\quad -(\sum_{a,b,c=1}^{M} \mathrm{E}_{ab}^{M\times M} \otimes \mathrm{E}_{bc}^{M\times M} \otimes \mathrm{E}_{ca}^{M\times M} \cdot \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\})(i,j,k,l), \\
&= \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[j]\ \mathbf{x}[k]\ \mathbf{x}[l]\} - \mathrm{E}\{\mathbf{x}[i]\ \mathbf{x}[j]\} \cdot \mathrm{E}\{\mathbf{x}[k]\ \mathbf{x}[l]\} \\
&\quad -(\mathrm{E}_{ii}^{M\times M} \otimes \mathrm{E}_{jk}^{M\times M} \otimes \mathrm{E}_{kj}^{M\times M} \cdot \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\})(i,j,k,l) \\
&\quad -(\mathrm{E}_{ij}^{M\times M} \otimes \mathrm{E}_{jk}^{M\times M} \otimes \mathrm{E}_{ki}^{M\times M} \cdot \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}\} \otimes \mathrm{E}\{\mathbf{x} \otimes \mathbf{x}^T\})(i,j,k,l),
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{E}\{\mathbf{x}[i]\,\mathbf{x}[j]\,\mathbf{x}[k]\,\mathbf{x}[l]\} - \mathrm{E}\{\mathbf{x}[i]\,\mathbf{x}[j]\}\cdot \mathrm{E}\{\mathbf{x}[k]\,\mathbf{x}[l]\} \\
&\quad -(\mathrm{E}\{\mathrm{E}_{ii}^{M\times M}\mathbf{x}\otimes \mathrm{E}_{jk}^{M\times M}\mathbf{x}\}\otimes \mathrm{E}\{\mathrm{E}_{kj}^{M\times M}\mathbf{x}\otimes \mathbf{x}^T\})(i,j,k,l) \\
&\quad -(\mathrm{E}\{\mathrm{E}_{ij}^{M\times M}\mathbf{x}\otimes \mathrm{E}_{jk}^{M\times M}\mathbf{x}\}\otimes \mathrm{E}\{\mathrm{E}_{ki}^{M\times M}\mathbf{x}\otimes \mathbf{x}^T\})(i,j,k,l).
\end{aligned}
$$

Now use the fact that the l^{th} element of $\mathrm{E}_{lm}^{M\times M}\mathbf{x}$ is $\mathbf{x}[m]$, so

$$
\begin{aligned}
\mathbf{C}_{4,\mathbf{x}}(i,j,k,l) &= \mathrm{E}\{\mathbf{x}[i]\,\mathbf{x}[j]\,\mathbf{x}[k]\,\mathbf{x}[l]\} - \mathrm{E}\{\mathbf{x}[i]\,\mathbf{x}[j]\}\cdot \mathrm{E}\{\mathbf{x}[k]\,\mathbf{x}[l]\} \\
&\quad -\mathrm{E}\{\mathbf{x}[i]\,\mathbf{x}[k]\}\cdot \mathrm{E}\{\mathbf{x}[j]\,\mathbf{x}[l]\} \\
&\quad -\mathrm{E}\{\mathbf{x}[j]\,\mathbf{x}[k]\}\cdot \mathrm{E}\{\mathbf{x}[i]\,\mathbf{x}[l]\}, \\
&= \mathrm{cum}(\mathbf{x}[i], \mathbf{x}[j], \mathbf{x}[k], \mathbf{x}[l]).
\end{aligned}
$$

The vector form for the 4th order cumulant is established by a similar argument and we only give the result here for the sake of brevity.

$$
\mathbf{c}_{4,\mathbf{x}} = \mathrm{E}\{\mathbf{x}^{(4)}\} + \mathbf{\Pi}_M \cdot \mathrm{E}\{\mathbf{x}^{(2)}\}\otimes \mathrm{E}\{\mathbf{x}^{(2)}\} \tag{57}
$$

where:

$$
\begin{aligned}
\mathbf{\Pi}_M &\triangleq -(\mathbf{\Pi}_{1,M} + \mathbf{\Pi}_{2,M} + \mathbf{\Pi}_{3,M}), \\
\mathbf{\Pi}_{1,M} &= \sum_{i,j,k,l=1}^{M} \mathrm{E}_{ii}^{M\times M}\otimes \mathrm{E}_{jj}^{M\times M}\otimes \mathrm{E}_{kk}^{M\times M}\otimes \mathrm{E}_{ll}^{M\times M}, && (58) \\
\mathbf{\Pi}_{2,M} &= \sum_{i,j,k,l=1}^{M} \mathrm{E}_{ii}^{M\times M}\otimes \mathrm{E}_{jk}^{M\times M}\otimes \mathrm{E}_{kj}^{M\times M}\otimes \mathrm{E}_{ll}^{M\times M}, && (59) \\
\mathbf{\Pi}_{3,M} &= \sum_{i,j,k,l=1}^{M} \mathrm{E}_{ij}^{M\times M}\otimes \mathrm{E}_{jk}^{M\times M}\otimes \mathrm{E}_{ki}^{M\times M}\otimes \mathrm{E}_{ll}^{M\times M}. && (60)
\end{aligned}
$$

□

Appendix B
Proof of Proposition 2

Proposition 2: Let $\mathbf{A}$ be a $Q\times M$ $(Q < M)$ deterministic matrix. Define the zero-mean real-valued random variable $\mathbf{z}$ as the product of $\mathbf{A}$ and $\mathbf{x}$:

$$
\mathbf{z} = \mathbf{A}\ \mathbf{x}. \tag{61}
$$

Then the following holds :

$$
\mathbf{m}_{n,\mathbf{z}} = \mathbf{A}^{(n)}\ \mathbf{m}_{n,\mathbf{x}},
$$

$$
\begin{aligned}
\mathbf{M}_{n,\mathbf{z}} &= \mathbf{A}^{(n-1)}\ \mathbf{m}_{n,\mathbf{x}}\ \mathbf{A}^T, \\
\mathbf{c}_{n,\mathbf{z}} &= \mathbf{A}^{(n)}\ \mathbf{c}_{n,\mathbf{x}}, \\
\mathbf{C}_{n,\mathbf{z}} &= \mathbf{A}^{(n-1)}\ \mathbf{C}_{n,\mathbf{x}}\ \mathbf{A}^T.
\end{aligned} \tag{62}
$$

Proof: From [1], we know that if $\mathbf{A}$ is a $Q \times M$ matrix and $\mathbf{x}$ an M-vector, then

$$
(\mathbf{A}\ \mathbf{x})^{(n-1)} \otimes (\mathbf{A}\ \mathbf{x})^T = \mathbf{A}^{(n-1)}\ (\mathbf{x}^{(n-1)} \otimes \mathbf{x}^T)\ \mathbf{A}^T. \tag{63}
$$

This implies

$$
\mathbf{m}_{n,\mathbf{z}} = \mathrm{E}\{\mathbf{z}^{(n)}\} = \mathrm{E}\{(\mathbf{A}\ \mathbf{x})^{(n)}\} = \mathbf{A}^{(n)}\ \mathrm{E}\{\mathbf{x}^{(n)}\} = \mathbf{A}^{(n)}\ \mathbf{m}_{n,\mathbf{x}} \tag{64}
$$

and

$$
\mathbf{M}_{n,\mathbf{z}} = \mathrm{E}\{\mathbf{z}^{(n-1)} \otimes \mathbf{z}^T\} = \mathbf{A}^{(n-1)}\ \mathbf{M}_{n,\mathbf{x}}\ \mathbf{A}^T. \tag{65}
$$

For the fourth order cumulant we need to do a little more work. Only the vector form is derived here; the matrix form follows in a similar fashion. Note that

$$
\begin{aligned}
\mathbf{c}_{4,\mathbf{z}} &= \mathbf{m}_{4,\mathbf{z}} + \mathbf{\Pi}_Q \cdot \mathbf{m}_{2,\mathbf{z}} \otimes \mathbf{m}_{2,\mathbf{z}}, \\
&= \mathbf{A}^{(4)}\ \mathbf{m}_{4,\mathbf{x}} + \mathbf{\Pi}_Q \cdot \mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)} \cdot \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{m}_{2,\mathbf{x}}, \\
&= \mathbf{A}^{(4)}\ \mathbf{m}_{4,\mathbf{x}} + \mathbf{\Pi}_Q \cdot \mathbf{A}^{(4)} \cdot \mathbf{m}_{2,\mathbf{x}} \otimes \mathbf{m}_{2,\mathbf{x}}.
\end{aligned} \tag{66}
$$

In order to show $\mathbf{c}_{4,\mathbf{z}} = \mathbf{A}^{(4)}\mathbf{c}_{4,\mathbf{x}}$, we must establish that $\mathbf{\Pi}_Q \cdot \mathbf{A}^{(4)} = \mathbf{A}^{(4)} \cdot \mathbf{\Pi}_M$. Recall that $\mathbf{\Pi}_M$ and $\mathbf{\Pi}_Q$ depend only on the size of $\mathbf{x}$ and $\mathbf{z}$, respectively, and that $\mathbf{\Pi}_J = -(\mathbf{\Pi}_{1,J} + \mathbf{\Pi}_{2,J} + \mathbf{\Pi}_{3,J})$ where $\mathbf{\Pi}_{1,J}, \mathbf{\Pi}_{2,J},$ and $\mathbf{\Pi}_{3,J}$ are given in (58), (59), and (60). We claim that for $i = 1, 2, 3$:

$$
\mathbf{\Pi}_{i,Q} \cdot \mathbf{A}^{(4)} = \mathbf{A}^{(4)} \cdot \mathbf{\Pi}_{i,M}. \tag{67}
$$

We verify (67) for the case $i = 3$. The other cases are established in a similar manner.

$$
\begin{aligned}
\mathbf{\Pi}_{i,Q} \cdot \mathbf{A}^{(4)} &= \sum_{i,j,k,l=1}^{Q} (\mathrm{E}_{ij}^{Q\times Q} \otimes \mathrm{E}_{jk}^{Q\times Q} \otimes \mathrm{E}_{ki}^{Q\times Q} \otimes \mathrm{E}_{ll}^{Q\times Q}) \cdot \mathbf{A}^{(4)}, \\
&= \sum_{i,j,k,l=1}^{Q} (\mathrm{E}_{ij}^{Q\times Q} \cdot \mathbf{A}) \otimes (\mathrm{E}_{jk}^{Q\times Q} \cdot \mathbf{A}) \otimes (\mathrm{E}_{ki}^{Q\times Q} \cdot \mathbf{A}) \otimes (\mathrm{E}_{ll}^{Q\times Q} \cdot \mathbf{A}), \qquad (68)
\end{aligned} \tag{69}
$$

We now show that (68) may be rewritten as

$$
\begin{aligned}
\mathbf{\Pi}_{i,Q} \cdot \mathbf{A}^{(4)} &= \sum_{t,u,v,w=1}^{M} (\mathbf{A} \cdot \mathrm{E}_{tu}^{M\times M}) \otimes (\mathbf{A} \cdot \mathrm{E}_{uv}^{M\times M}) \otimes (\mathbf{A} \cdot \mathrm{E}_{vt}^{M\times M}) \otimes (\mathbf{A} \cdot \mathrm{E}_{ww}^{M\times M}), \qquad (70) \\
&= \mathbf{A}^{(4)} \cdot \mathbf{\Pi}_{i,M}.
\end{aligned}
$$

In order to do this, first write (see [1]) :

$$
\begin{aligned}
\mathbf{A} &= \sum_{t}^{Q}\sum_{t'}^{M}\mathbf{A}[t,t']\cdot \mathrm{E}_{tt'}^{Q\times M},\\
\mathbf{A}^T &= \sum_{t}^{Q}\sum_{t'}^{M}\mathbf{A}[t,t']\cdot \mathrm{E}_{t't}^{M\times Q}.
\end{aligned}
$$

Define $\delta_{ab}=1$ for $a=b$ and $\delta_{ab}=0$ for $a\neq b$. Use the properties in Table I and Table II of [1] to write

$$
\begin{aligned}
& \sum_{i,j,k,l=1}^{Q}(\mathrm{E}_{ij}^{Q\times Q}\cdot\mathbf{A})\otimes(\mathrm{E}_{jk}^{Q\times Q}\cdot\mathbf{A})\otimes(\mathrm{E}_{ki}^{Q\times Q}\cdot\mathbf{A})\otimes(\mathrm{E}_{ll}^{Q\times Q}\cdot\mathbf{A})\\
= & \sum_{i,j,k,l=1}^{Q}(\sum_{u'=1}^{Q}\sum_{u=1}^{M}\mathrm{E}_{ij}^{Q\times Q}\cdot\mathbf{A}[u',u]\cdot\mathrm{E}_{u'u}^{Q\times M})\otimes(\sum_{v'=1}^{Q}\sum_{v=1}^{M}\mathrm{E}_{jk}^{Q\times Q}\cdot\mathbf{A}[v',v]\cdot\mathrm{E}_{v'v}^{Q\times M})\\
& \otimes(\sum_{t'=1}^{Q}\sum_{t=1}^{M}\mathrm{E}_{ki}^{Q\times Q}\cdot\mathbf{A}[t',t]\cdot\mathrm{E}_{t't}^{Q\times M})\otimes(\sum_{w'=1}^{Q}\sum_{w=1}^{M}\mathrm{E}_{ll}^{Q\times Q}\cdot\mathbf{A}[w',w]\cdot\mathrm{E}_{w'w}^{Q\times M}),\\
= & \sum_{i,j,k,l=1}^{Q}(\sum_{u'=1}^{Q}\sum_{u=1}^{M}\mathrm{E}_{iu}^{Q\times M}\cdot\mathbf{A}[u',u]\cdot\delta_{ju'})\otimes(\sum_{v'=1}^{Q}\sum_{v=1}^{Q}\mathrm{E}_{jv}^{Q\times M}\cdot\mathbf{A}[v',v]\cdot\delta_{kv'})\\
& \otimes(\sum_{t'=1}^{Q}\sum_{t=1}^{M}\mathrm{E}_{kt}^{Q\times M}\cdot\mathbf{A}[t',t]\cdot\delta_{it'})\otimes(\sum_{w'=1}^{Q}\sum_{w=1}^{M}\mathrm{E}_{lw}^{Q\times M}\cdot\mathbf{A}[w',w]\cdot\delta_{lw'}),\\
= & \sum_{i,j,k,l=1}^{Q}(\sum_{u=1}^{M}\mathrm{E}_{iu}^{Q\times M}\cdot\mathbf{A}[j,u])\otimes(\sum_{v=1}^{M}\mathrm{E}_{jv}^{Q\times M}\cdot\mathbf{A}[k.v])\\
& \otimes(\sum_{t=1}^{M}\mathrm{E}_{kt}^{Q\times M}\cdot\mathbf{A}[i,t])\otimes(\sum_{w=1}^{M}\mathrm{E}_{lw}^{Q\times M}\cdot\mathbf{A}[l,w]),\\
= & \sum_{i,j,k,l=1}^{Q}\sum_{u,v,w,t=1}^{M}\mathbf{A}[j,u]\cdot\mathbf{A}[k,v]\cdot\mathbf{A}[i,t]\cdot\mathbf{A}[l,w]\cdot(\mathrm{E}_{iu}^{Q\times M}\otimes\mathrm{E}_{jv}^{Q\times M}\otimes\mathrm{E}_{kt}^{Q\times M}\otimes\mathrm{E}_{lw}^{Q\times M}),\\
= & \sum_{t,u,v,w=1}^{M}\sum_{i,j,k,l=1}^{Q}(\mathrm{E}_{iu}^{Q\times M}\cdot\mathbf{A}[i,t])\otimes(\mathrm{E}_{jv}^{Q\times M}\cdot\mathbf{A}[j,u])\otimes(\mathrm{E}_{kt}^{Q\times M}\cdot\mathbf{A}[k,v])\otimes(\mathrm{E}_{lw}^{Q\times M}\cdot\mathbf{A}[l,w]),\\
= & \sum_{t,u,v,w=1}^{M}(\sum_{i=1}^{Q}\sum_{i'=1}^{M}\mathrm{E}_{ii'}^{Q\times M}\cdot\mathbf{A}[i,i']\cdot\mathrm{E}_{tu}^{M\times M})\otimes(\sum_{j=1}^{Q}\sum_{j'=1}^{M}\mathrm{E}_{jj'}^{Q\times M}\cdot\mathbf{A}[j,j']\cdot\mathrm{E}_{uv}^{M\times M})\\
& (\sum_{k=1}^{Q}\sum_{k'=1}^{M}\mathrm{E}_{kk'}^{Q\times M}\cdot\mathbf{A}[k,k']\cdot\mathrm{E}_{vt}^{M\times M})\otimes(\sum_{l=1}^{Q}\sum_{l'=1}^{M}\mathrm{E}_{ll'}^{Q\times M}\cdot\mathbf{A}[l,l']\cdot\mathrm{E}_{ww}^{M\times M}),\\
= & \sum_{t,u,v,w=1}^{M}(\mathbf{A}\cdot\mathrm{E}_{tu}^{M\times M})\otimes(\mathbf{A}\cdot\mathrm{E}_{uv}^{M\times M})\otimes(\mathbf{A}\cdot\mathrm{E}_{vt}^{M\times M})\otimes(\mathbf{A}\cdot\mathrm{E}_{ww}^{M\times M}).
\end{aligned}
$$

□

References

[1] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory," *IEEE Trans. Circuits and Systems*, Vol. 25, pp. 772-781, 1978

[2] G. B. Giannakis, A. V. Dandawate, "Higher-Order Statistics-Based Input/Output Systen Identification and Application to Noise Cancellation," *Circuits Systems Signal Process*, Vol. 10, No. 4, pp. 485-511, 1991

[3] A. Depopoulos, G. B. Giannakis, "Consistent Identification of Stochastic Linear Systems with Noisy Input-Output Data", *Automatica*, vol.30, No. 8, pp. 1271-1294, 1994

[4] L. L. Scharf, D. W. Tufts, "Rank Reduction for Modeling Stationary Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-35, No. 3, pp. 350-355, March 1987

[5] C. L. Nikias, M. R. Raghuveer, "Bipectrum Estimation: A Digital Signal Processing Framework," *Proceedings of the IEEE*, Vol. 75, No. 7, pp. 869-891, July 1987

[6] A.L. Swindlehurst, T. Kailath, "Detection and Estimation Using the Third Moment Matrix," *ICASSP 1989*, E6.7, pp. 2325-2328, 1989

[7] P. A. Delaney, "Signal Detection Using Third-Order Moments," *Circuits Systems Signal Processing*, Vol. 13, No. 4, pp. 481-496,1994

[8] M. J. Hinich, "Higher Order Cumulants and Cumulant Spectra," *Circuits Systems Signal Process.*, Vol. 13, No. 4, pp.391-402, 1994

[9] I. I. Jouny, R. L. Moses, "The Bispectrum of Complex Signals: Definitions and Properties," *IEEE Transactions on Signal Processing*, Vol. 40, No. 11, November 1992

[10] R. D. Nowak, B. D. Van Veen, "Tensor Product Basis Approximations for Volterra Filters," *to appear in IEEE Transactions on Signal Processing*, February 1996

[11] R. Pan, C. L. Nikias, "Harmonic Decomposition Methods in Cumulant Domains," *Proc. ICASSP' 88* (New York), Apr. 1988, pp. 2356-2359

[12] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables*, Marcel Dekker, Inc., New York, 1992.

[13] D. E. Cartwright, "A Unified Analysis of Tides and Surges around North and East Britain," *Philos. Trans. Royal Society of London Ser A 263*, pp. 1-55, 1968

[14] L. Ljung, "System Identification: Theory for the User," Prenctice-Hall, Englewood Cliffs, NJ, 1987

[15] A. P. Petropulu, C. L. Nikias "Signal Reconstruction from the Phase of the Bispectrum," *IEEE Transaction on Signal Processing*, Vol. 40, No. 3, March 1992

[16] A. E. Cetin, "An Iterative Algorithm for Signal Reconstruction from Bispectrum," *IEEE Transactions on Signal Processing*, Vol. 39, No. 12, December 1991

[17] M. M. Chang, A. M. Tekalp, A. T. Erdem, "Blur Identification Using the Bispectrum," *IEEE Transactions on Signal Processing*, Vol. 39, No. 10, October 1991

[18] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty - V : The Discrete Case," *Bell Syst. Tech. J.*, Vol. 40, pp. 1371-1429, 1978

[19] P. Forster, C. L. Nikias "Bearing Estimation in the Bispectrum Domain," *IEEE Transactions on Signal Processing*, Vol. 39, No. 9, September 1991

[20] A. M. Richardson, W. S. Hodgkiss "Bispectral analysis of underwater acoustic data," *J. Acoust. Soc. Am.*, Vol. 96, No. 2, Pt. 1, August 1994, pp. 828-837