Efficient Methods for Identification of Volterra Filter Models

Robert D. Nowak* and Barry D. Van Veen†

Abstract

A major drawback of the truncated Volterra series or "Volterra filter" for system identification is the large number of parameters required by the standard filter structure. The corresponding estimation problem requires the solution of a large system of simultaneous linear equations. Two methods for simplifying the estimation problem are discussed in this paper. First, a Kronecker product structure for the Volterra filter is reviewed. In this approach the inverse of the large correlation matrix is expressed as a Kronecker product of small matrices. Second, a parallel decomposition of the Volterra filter based on uncorrelated, symmetric inputs is introduced. Here the Volterra filter is decomposed into a parallel combination of smaller orthogonal "sub-filters". It is shown that each sub-filter is much smaller than the full Volterra filter and hence the parallel decomposition offers many advantages for estimating the Volterra kernels. Simulations illustrate application of the parallel structure with random and pseudorandom excitations. Input conditions that guarantee the existence of a unique estimate are also reviewed.

Key Words: Volterra filter; nonlinear system identification; persistence of excitation.

---

* Supported by Rockwell International Doctoral Fellowship Program. † Supported in part by the National Science Foundation under Award MIP-895 8559, Army Research Office under Grant DAAH04-93-6-0208, 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.

Unusual Symbols:

$\boldsymbol{\mathcal{X}}$ - boldface calligraphic or script X

$\bar{\boldsymbol{\mathcal{X}}}$ - bar over $\boldsymbol{\mathcal{X}}$

Γ - Greek gamma

$\otimes$ - Kronecker product operator

number of pages: 22

number of tables: 3

number of figures: 3

Running headline: *R.D. Nowak, B. Van Veen \ Identification of Volterra Filters*

Key words: Volterra filter; nonlinear system identification; persistence of excitation.

1. Introduction

The truncated Volterra series or "Volterra filter" is an attractive nonlinear system representation because the parameters of this model are linearly related to the output. Thus, given the system input and output, the Volterra filter parameters are estimated by solving a system of linear equations. The Volterra filter is also a very general nonlinear system representation. It has been shown recently [15] that the truncated (or "doubly finite") Volterra filter provides a uniform approximation of the infinite Volterra series on a ball of bounded inputs for a large class of systems.

One important issue regarding the use of Volterra filter models for system identification is the conditions that the input must satisfy in order to guarantee that the filter parameters are uniquely determined. Such conditions, known as persistence of excitation (*p.e.*) conditions, have been studied for Volterra filters driven by random inputs [2, 13]. It has been shown that a class of deterministic signals known as pseudo-random multilevel sequences (PRMS) are *p.e.* for Volterra filters and are particularly well-suited for identification experiments [13].

A major limitation of the standard Volterra filter model is that it has a large number of parameters for even modest nonlinearity orders and memory lengths. This leads to an estimation problem that requires the solution of a very large system of linear equations, and consequently presents a very large computational burden. Furthermore, this system of linear equations is often ill-conditioned [13]. Simplifications are possible if the input is Gaussian [7], but only for quadratic systems. Such difficulties have prompted several authors [1, 8, 9] to consider orthogonalized Volterra filter structures. Orthogonal structures for the Volterra filter, however, require exact knowledge of the moments of the input distribution or employ sub-optimal, numerically intensive algorithms. Also, orthogonal structures do not give direct estimates of the Volterra kernels. Another approach is based on lattice filter structures [11]. This approach also involves transformations of the input and does not yield direct estimates of the Volterra kernels.

In this paper, two alternative approaches are presented that offer many advantages over existing filter structures. First, a Kronecker product structure for Volterra filters is reviewed [13]. This structure is used to develop an efficient identification algorithm based on uncorrelated or PRMS excitations [14]. Second, a new parallel decomposition Volterra filter (PVF) structure is presented for problems involving uncorrelated, symmetric input signals. Here the standard Volterra filter is decomposed into a parallel structure of smaller "sub-filters" that are obtained directly from the standard Volterra filter structure and require no transformation or orthogonalization. Moreover, it is shown that the sub-filters are orthogonal to each other in the mean-square sense provided the input is uncorrelated and symmetric. The exact size of each sub-filter is derived and is generally much smaller than the total number of parameters in the Volterra kernels. Hence, the PVF structure dramatically reduces the computational burden associated with estimation of the Volterra kernels. It also generally improves the conditioning of the estimation problem.

The PVF structure is appropriate for the class of random inputs that are uncorrelated, symmetric, and satisfy the *p.e.* condition. This class includes Gaussian white noise. The PVF structure can also be used with symmetric PRMS. Unlike their stochastic counterparts, PRMS provide the desired orthogonality over finite data records. This makes PRMS extremely useful for practical identification experiments.

The paper is organized as follows. Section 2 describes the standard Volterra filter structure and contains a brief review of the Kronecker product structure for the Volterra filter. Persistence of excitation conditions for the Volterra filter are given in section 3. The theoretical development of the PVF structure is given in section 4. Section 5 contains a numerical study demonstrating the performance of the PVF structure using random and pseudo-random excitation sequences. A summary is provided in section 6.

2. The Volterra Filter

Consider a single input, single output, discrete time-invariant system with nonlinearities of polynomial order N and memory length M. The system is described by the Volterra filter

$$\begin{aligned} Y_k &= h^0 + \sum_{j_1=1}^{M} h^1_{j_1} X_{k-j_1+1} + \\ &+ \sum_{j_1=1}^{M}\sum_{j_2=1}^{M} h^2_{j_1,j_2} X_{k-j_1+1} X_{k-j_2+1} + \cdots \\ &+ \sum_{j_1=1}^{M} \cdots \sum_{j_N=1}^{M} h^N_{j_1,\cdots,j_N} X_{k-j_1+1} \cdots X_{k-j_N+1}, \end{aligned} \tag{1}$$

where $\{Y_k\}_{k\in\mathbb{Z}}$ is the observed output sequence associated with the input sequence $\{X_k\}_{k\in\mathbb{Z}}$ ($\mathbb{Z}$ is the set of integers). In equation (1), $h^n_{j_1,\,\cdots,j_n}$, $j_1,\ldots,j_n \in \{1,\ldots,M\}$, is referred to as the n-th order Volterra kernel. Note that the Volterra kernel can be assumed symmetric without loss of generality. We will refer to (1) as the standard Volterra filter structure.

Let $E(\cdot)$ denote the expectation operator. For system identification experiments, the input sequence $\{X_k\}$ is often chosen to be a random sequence satisfying the following two conditions:

A1 - 2N-th order stationarity:

$$E(\prod_{i=1}^{M} X_{k-i+1}^{n_i}) \quad \text{is independent of } k$$

for $n_i \in \{0, 1, \ldots, 2N\}$, $i = 1, \ldots, M$, satisfying $\sum_{i=1}^{M} n_i \leq 2N$.

A2 - Uncorrelated up to order 2N:

(*i*) $E(X_k^{2N}) < \infty$, which guarantees all necessary cross moments exist,

(*ii*) $$E(\prod_{i=1}^{M} X_{k-i+1}^{n_i}) \;=\; \prod_{i=1}^{M} E(X_{k-i+1}^{n_i})$$

for $n_i \in \{0, 1, \ldots, 2N\}$, $i = 1, \ldots, M$, satisfying $\sum_{i=1}^{M} n_i \leq 2N$.

Under assumptions $A1$ and $A2$ estimates of the Volterra kernels may be computed in several ways [12].

A Kronecker product structure for the Volterra filter is introduced in [13]. This representation exploits the input conditions $A1$ and $A2$ and results in a simple method for computing the Volterra kernels. Define the vector $\mathbf{X}_k = [1,\ X_k,\ X_k^2,\ \ldots, X_k^N]^T$ and the $p = (N+1)^M$ dimensional vector

$$\mathbf{X}_k^{[M]} \;=\; \mathbf{X}_k \otimes \mathbf{X}_{k-1} \otimes \cdots \otimes \mathbf{X}_{k-M+1}, \tag{2}$$

where $\otimes$ is the Kronecker product operator [5]. All products of the input data values necessary for the evaluation of (1) are included as elements of $\mathbf{X}_k^{[M]}$. The Volterra filter output is a linear combination of these products; hence we rewrite the Volterra filter as

$$Y_k \;=\; \boldsymbol{\theta}^T \mathbf{X}_k^{[M]}, \tag{3}$$

where the elements of the p-dimensional vector $\boldsymbol{\theta}$ correspond to the elements of the Volterra kernels. Note that $\mathbf{X}_k^{[M]}$ contains additional high order "cross-term" products that are not required for the evaluation of (1). Exact equivalence between the N-th order systems described by (3) and (1) is obtained by setting the elements of $\boldsymbol{\theta}$ corresponding to the surplus products equal to zero.

Now if $\{X_k\}$ satisfies $A1$ and all necessary cross moments exist, the optimal estimate of $\boldsymbol{\theta}$ in the mean square sense is found by solving the system of linear equations

$$E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T)\hat{\boldsymbol{\theta}} \;=\; E(\mathbf{X}_k^{[M]} Y_k). \tag{4}$$

An elegant simplification is obtained if the input is sufficiently uncorrelated. First, the uncorrelated assumption $A2$ is slightly strengthened because of the additional higher order terms included in the Kronecker product structure.

$A2'$: For $n_i \in \{0, 1, \ldots, 2N\},\ i = 1, \ldots, M,$

$$E(\prod_{i=1}^{M} X_{k-i+1}^{n_i}) = \prod_{i=1}^{M} E(X_{k-i+1}^{n_i}).$$

Note that, assuming the required cross moments exist, an i.i.d. sequence satisfies $A1$, $A2$, and $A2'$.

The correlation matrix is given by

$$E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T) = E\left((\mathbf{X}_k \otimes \cdots \otimes \mathbf{X}_{k-M+1})\,(\mathbf{X}_k \otimes \cdots \otimes \mathbf{X}_{k-M+1})^T\right). \tag{5}$$

Using the Kronecker transposition property and mixed product rule [5] , (5) is written as

$$E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T) = E\left(\mathbf{X}_k\mathbf{X}_k^T \otimes \cdots \otimes \mathbf{X}_{k-M+1}\mathbf{X}_{k-M+1}^T\right). \tag{6}$$

Now, applying $A2'$, the expectation operator is factored through the Kronecker product producing

$$E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T) = \underbrace{\mathbf{C} \otimes \cdots \otimes \mathbf{C}}_{M \text{ times}}, \tag{7}$$

where the matrix $\mathbf{C} = E(\mathbf{X}_k\mathbf{X}_k^T)$ is an $(N+1) \times (N+1)$ Hankel matrix of the form

$$\mathbf{C} = \begin{bmatrix} 1 & m_1 & m_2 & \cdots & m_N \\ m_1 & m_2 & m_3 & \cdots & m_{N+1} \\ m_2 & m_3 & m_4 & \cdots & m_{N+2} \\ \vdots & \vdots & \vdots & & \vdots \\ m_N & m_{N+1} & m_{N+2} & \cdots & m_{2N} \end{bmatrix} \tag{8}$$

and m_i is the i-th moment of the sequence $\{X_k\}$. Assuming $E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T)$ is invertible, using the Kronecker product inversion property [5] it is readily apparent that

$$(E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T))^{-1} = \mathbf{C}^{-1} \otimes \cdots \otimes \mathbf{C}^{-1}. \tag{9}$$

Hence, solution (4) only requires the inversion of an $(N+1) \times (N+1)$ matrix $\mathbf{C}$ rather than the inversion of the $(N+1)^M \times (N+1)^M$ matrix $E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T)$. Also note that $E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T)$ is invertible if and only if $\mathbf{C}$ is invertible. The input conditions that guarantee the invertibility of $\mathbf{C}$

are studied in the following section.

A similar result is also obtained for PRMS inputs. In [13] it is shown that the sample correlation matrix corresponding to a q-level PRMS ($q \geq N+1$) of degree $m \geq M$ is equivalent to the M-fold Kronecker product

$$\sum_{n=k}^{k+\Gamma} \mathbf{X}_n^{[M]}(\mathbf{X}_n^{[M]})^T \equiv \underbrace{\mathbf{V}\mathbf{V}^T \otimes \cdots \otimes \mathbf{V}\mathbf{V}^T}_{M \text{ times}} \tag{10}$$

where $\Gamma = q^m$ and $\mathbf{V}$ is the $q \times (N+1)$ "extended" Vandermonde matrix generated by the level values

$$\mathbf{V} = \begin{bmatrix} 1 & l_0 & \cdots & l_0^N \\ 1 & l_1 & \cdots & l_1^N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & l_{q-1} & \cdots & l_{q-1}^N \end{bmatrix}. \tag{11}$$

For a more detailed discussion of the Kronecker product structure see [13, 14].

The main drawback of the Kronecker product structure is the inclusion of additional high-order cross-term products. Accurate estimates of the expected value of such products may require an extremely large amount of raw data. The use of PRMS circumvents this problem. However, in practice the Kronecker product structure is only useful for systems with relatively short memory lengths because it contains $(N+1)^M$ parameters.

3. Persistence of Excitation Conditions for Volterra Filters

In this section, we review the persistence of excitation conditions for Volterra filters. Let $\bar{\boldsymbol{\mathcal{X}}}$ be a vector composed of the products of the input $\{X_k\}$ necessary for the evaluation of (1). For example, if $N = 2$ and $M = 2$ then $\bar{\boldsymbol{\mathcal{X}}} \triangleq [1,\ X_k,\ X_{k-1},\ X_k^2,\ X_k X_{k-1},\ X_{k-1}^2]^T$. Also, let $\boldsymbol{\theta}$ be a vector composed of the elements in the Volterra kernels so that $Y_k = \boldsymbol{\theta}^T \bar{\boldsymbol{\mathcal{X}}}$. The ordering of products in $\bar{\boldsymbol{\mathcal{X}}}$ is not important in this section provided it corresponds to the order of kernel elements in $\boldsymbol{\theta}$.

The input sequence is said to be *persistently exciting* (*p.e.*) if the vector $\bar{\boldsymbol{\mathcal{X}}}$ persistently spans the parameter space. Following the definition of *p.e.* inputs for linear filters, a similar definition is applied to the Volterra filter.

Definition 1: If the correlation matrix $E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)$ exists and is non-singular, the sequence $\{X_k\}$ is said to be *p.e.* of degree M and order N or *p.e.*(M, N).

Note that satisfation ot the *p.e.* condition guarantees a unique MSE estimate of the Volterra kernels.

3.1 Random Excitation Sequences

As noted in the previous section, under assumptions $A1$ and $A2'$, the correlation matrix $E(\mathbf{X}_k^{[M]}(\mathbf{X}_k^{[M]})^T)$ corresponding to the Kronecker product structure is invertible if and only if $\mathbf{C}$ (8) is invertible. Since the standard Volterra filter structure (1) is a special case of the Kronecker product structure (3), the invertibility of $\mathbf{C}$ is sufficient to guarantee that $E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)$ is invertible. Also notice that $\mathbf{C}$ is a submatrix of $E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)$. Hence, the invertibility of $\mathbf{C}$ is necessary as well. The following lemma gives a neccesary and sufficient condition for the invertibility of $\mathbf{C}$.

Lemma 3.1: The matrix $\mathbf{C}$ is singular if and only if the random variable X_k takes on at most N distinct values with probability 1. □

Lemma 3.1 and its consequences have been discussed in various forms [2, 6, 13] and originates in the Hamburger moment problem [16]. The proof of the lemma is elementary and follows by noting that the singularity of $\mathbf{C}$ implies that there exists a vector $\mathbf{b} \in \mathbb{R}^{N+1}$ such that $\mathbf{b}^T\mathbf{C}\mathbf{b} = 0$ or $E[(\mathbf{b}^T\mathbf{X}_k)^2] = 0$. Obviously this is equivalent to $\mathbf{b}^T\mathbf{X}_k = 0$ with probability 1. Now $\mathbf{b}^T\mathbf{X}_k$ is a degree N polynomial in the random variable X_k and has at most N distinct zeros with probability 1.

Lemma 3.1 indicates that common inputs for system identification such as Gaussian white noise (GWN) [12] and random multilevel sequences (RMS) [13] are *p.e.* for Volterra filters. RMS are particularly desirable because they are relatively easy to generate. An RMS is defined as follows.

Definition 2: Let $\{X_k\}$ be an i.i.d sequence taking on a finite number of distinct values $\{l_0, l_1, \ldots, l_{q-1}\} \subset \mathbb{R}$ with corresponding positive probabilities $p_0, p_1, \ldots, p_{q-1}, \sum_{i=0}^{q-1} p_i = 1$. The sequence $\{X_k\}$ is called a *q-level RMS*.

Note that an RMS satisfies $A1$, $A2$ and $A2'$ since it is i.i.d. and the level values are finite. Lemma 3.1 indicates that a q-level RMS is $p.e.(N, M)$ for all finite M if and only if $q \geq N + 1$.

3.2 Pseudo-random Excitation Sequences

Pseudo-random binary sequences (PRBS) are widely used for linear system identification because of their many desirable properties and the fact that PRBS can be easily generated using shift registers [17]. Unfortunately, PRBS are not *p.e.* for Volterra filters of polynomial order $N \geq 2$. Pseudo-random multilevel sequences (PRMS) share many of the desirable properties of PRBS and can be used for Volterra filter identification. We define a q-level PRMS as follows.

Definition 3: Let $S_q = \{l_0,\ l_1,\ \ldots,\ l_{q-1}\} \subset \mathbb{R}$ be a set of q distinct level values. Let $\{X_k\}$ be a sequence with elements taken from the set S_q and the m-fold Cartesian product of S_q with itself be denoted as $\bigotimes^m S_q$. If there exists a finite observation interval Γ such that for every k, each ordered m-tuple in $\bigotimes^m S_q$ occurs sequentially in the collection $\{X_n\}_{n=k}^{k+\Gamma}$, then $\{X_k\}$ is called a *q-level PRMS of degree m.*

Note that there are q^m distinct m-tuples in the set $\bigotimes^m S_q$.

Lemma 3.2: A q-level PRMS of degree m is *p.e.*(M, N) if $m \geq M$ and if and only if $q \geq N+1$. $\square$

The proof follows from the Kronecker product form of the correlation matrix (10) by noting that the invertibility of the matrix $\mathbf{V}\mathbf{V}^T$ guarantees the invertibility of the correlation matrix. For details see [14].

A useful class of PRMS are known as maximal length sequences (MLS) [10, 12]. One important property of MLS is known as the *window property.*

Property 1 - The Window Property: If a window of width M is slid along a q-level maximal length sequence of degree M, all but one of the q^M M-tuples is seen exactly once over one period (length $q^M - 1$) of the sequence. The sequence is easily augmented with the remaining M-tuple.

The window property is particularly important for system identification since it implies that an $(N+1)$-level MLS of degree M is *periodically p.e.* for an N-th order filter of memory length M, with period $\Gamma = (N+1)^M$.

A detailed study of random and pseudorandom sequences for Volterra filter identification is given in [14]. The condition of the corresponding system of linear equations is also examined.

4. Parallel Decomposition of the Volterra Filter

In this section an alternative approach is suggested which offers many advantages over existing filter structures. A parallel decomposition Volterra filter (PVF) structure is derived, assuming the input is uncorrelated and symmetric, by decomposing the standard Volterra filter into a parallel structure of smaller "sub-filters". The sub-filters are obtained directly from the standard Volterra filter structure and require no transformation or orthogonalization. Moreover, it is shown that the sub-filters are orthogonal to each other in the mean-square sense provided the input is symmetric and uncorrelated. The size of each sub-filter is much smaller than the total number of parameters in the Volterra kernels. Hence, the PVF structure circumvents the computational burden associated direct kernel estimation methods. It also generally improves the condition of the estimation problem.

4.1 Theoretical Development of Parallel Decomposition

Recall that the output Y_k described by (1) is simply a linear combination of products of the input

$$\prod_{i=1}^{M} X_{k-i+1}^{n_i} , \tag{12}$$

where

$$n_i \in \{0, 1, \ldots, N\},\ i = 1, \ldots, M, \text{ and } \sum_{i=1}^{M} n_i \leq N. \tag{13}$$

The parallel decomposition is obtained by grouping the products (12) into sets that are orthogonal when the input $\{X_k\}_{k\in\mathbb{Z}}$ satisfies $A1$, $A2$, and the symmetry condition:

A3 - 2N-th order symmetric process:

$$E(X_k^{2p+1}) = 0, \quad p = 0, 1, \ldots, N-1.$$

A broad class of signals satisfy $A1$, $A2$, and $A3$ including Gaussian white noise and random multi-level sequences with levels symmetric about the origin.

Since stationarity is assumed ($A1$), we simplify the notation by arbitrarily fixing $k = M$. Hence, the set of random variables (r.v.'s) $\{X_i\}_{i=1}^{M}$ represents the input sequence.

Let $\boldsymbol{\mathcal{X}}$ denote the set of all products of the $\{X_i\}_{i=1}^{M}$ necessary for the evaluation of (1), that is,

$$\boldsymbol{\mathcal{X}} \triangleq \{ \prod_{i=1}^{M} X_i^{n_i} \ : \ \sum_{i=1}^{M} n_i \leq N,\ n_i \in \{0, 1, \ldots, N\} \}. \tag{14}$$

Now $\boldsymbol{\mathcal{X}}$ is decomposed into smaller subsets that are orthogonal to one another in the mean square sense. Assumption $A2$ and the symmetry assumption $A3$ guarantee that the expected value of any product containing at least one r.v. raised to an odd power is identically zero. The desired decomposition is accomplished by appropriately grouping elements of $\boldsymbol{\mathcal{X}}$ according to r.v.'s raised to odd powers.

Let S be the set of all subsets of $\{X_i\}_{i=1}^{M}$ and denote these subsets as $\{I_j\}_{j=1}^{2^M}$ with $I_1 \triangleq \emptyset$. Each $\alpha \in \boldsymbol{\mathcal{X}}$ is a product of r.v.'s raised to even and odd powers where the power 0 is considered an even power. Let α^o be the factor of α consisting of r.v.'s raised to odd powers. For example, if $\alpha = X_1^3 X_2^2 X_3$, then $\alpha^o = X_1^3 X_3$.

We now define the orthogonal subsets of $\boldsymbol{\mathcal{X}}$. For each $j = 1, 2, \ldots, 2^M$ let

$$\boldsymbol{\mathcal{X}}_j \triangleq \{ \alpha \in \boldsymbol{\mathcal{X}} \ : \ \alpha^o = \prod_{X_i \in I_j} X_i^{2p_i+1} \ ,\ p_i \in \{0, 1, \ldots\} \}. \tag{15}$$

In words, $\boldsymbol{\mathcal{X}}_j$ contains products with all r.v.'s in set I_j raised to odd powers and all other r.v.'s raised to even powers. Observe that if $N < M$, which is often the case, some of the sets $\{\boldsymbol{\mathcal{X}}_j\}_{j=1}^{2M}$ are empty. It is easily verified that

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_1 \cup \cdots \cup \boldsymbol{\mathcal{X}}_{2M}. \tag{16}$$

The following proposition shows that these subsets are orthogonal to one another.

Proposition 5.1: If $\alpha \in \boldsymbol{\mathcal{X}}_j$ and $\beta \in \boldsymbol{\mathcal{X}}_k$, then

$$E(\alpha\beta) = \begin{cases} \geq 0 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \tag{17}$$

Proof: Notice that if $j = k$, then $\alpha^o\beta^o$ contains only even-order r.v.'s. Thus, in this case, the product $\alpha\beta$ contains only even-order r.v.'s and $E(\alpha\beta) \geq 0$ almost surely ($a.s.$). If $j \neq k$ then $\alpha\beta$ contains at least one odd-order r.v. Hence, $A2$ and $A3$ imply $E(\alpha\beta) = 0$ $a.s.$ □

Note that the condition $E(\alpha\beta) = 0$, in the case $j = k$, implies a degenerate random process equal to zero $a.s.$ Therefore, in practice, the inequality in (17) is written as a strict inequality. The importance of the Proposition 5.1 is that the Volterra filter (1) can be decomposed into a parallel structure of smaller, orthogonal filter banks.

Of practical interest is the number of products in each "sub-filter", since the number of products corresponds to the number of kernel parameters associated with each sub-filter. For any set B, let card$B$ denote the cardinality or number of elements in the set. Also, let the number of r-combinations of an n-set be denoted as

$$C(n, r) \triangleq \frac{n!}{r!(n-r)!}. \tag{18}$$

The following results are derived in Appendix A. The number of elements in $\boldsymbol{\mathcal{X}}$ is card$\boldsymbol{\mathcal{X}} = C(M+N, M)$, including the zero-order product 1. For each sub-filter, if I_j contains k r.v.'s, then the number of products in $\boldsymbol{\mathcal{X}}_j$ is given by

$$P_k = \begin{cases} C(M + [\frac{N-k}{2}],\ M), & \text{if } 0 \leq k \leq N \\ 0, & \text{if } k > N \end{cases} \tag{19}$$

where $[\cdot]$ is the integer part of the argument. Clearly, the largest sub-filter size is $\mathrm{C}(M + [\frac{N}{2}],\ M)$. Comparing this to card$\boldsymbol{\mathcal{X}} = C(M+N, M)$ it is clear that a significant reduction in size is obtained. For large M and relatively small N, the largest sub-filter size is roughly $\mathcal{O}(\frac{M^{[\frac{N}{2}]}}{[\frac{N}{2}]!})$ while the full Volterra filter is $\mathcal{O}(\frac{M^N}{N!})$.

4.2 Vector Representation and Optimal Volterra Kernels

The number of non-empty sub-filter sets is $s = \sum_{i=0}^{min(M,N)} C(M,i)$. For each $1 \leq j \leq s$, define $\bar{\boldsymbol{\mathcal{X}}}_j$ to be a column vector composed of the products in the set $\boldsymbol{\mathcal{X}}_j$. The order of products in each sub-filter vector $\bar{\boldsymbol{\mathcal{X}}}_j$ is arbitrary. Define the vector

$$\bar{\boldsymbol{\mathcal{X}}} \triangleq [\bar{\boldsymbol{\mathcal{X}}}_1^T,\ \bar{\boldsymbol{\mathcal{X}}}_2^T,\ \ldots, \bar{\boldsymbol{\mathcal{X}}}_s^T]^T. \tag{20}$$

The optimal Volterra kernels in the mean squared error sense are obtained by solving the system of equations

$$E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)\hat{\boldsymbol{\theta}} \ = \ E(\bar{\boldsymbol{\mathcal{X}}}Y), \tag{21}$$

where Y is the output of (1) and $\hat{\boldsymbol{\theta}}$ is a vector of estimated elements in the Volterra kernels. Due to the orthogonality between sub-filters, the matrix $E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)$ is block diagonal, and hence the large system of equations in (21) is reduced to a solving s small systems of equations having the form

$$E(\bar{\boldsymbol{\mathcal{X}}}_j\bar{\boldsymbol{\mathcal{X}}}_j^T)\hat{\boldsymbol{\theta}}_j \ = \ E(\bar{\boldsymbol{\mathcal{X}}}_jY), \ \ 1 \leq j \leq s, \tag{22}$$

where $\hat{\boldsymbol{\theta}}_j$ is a vector of the estimated elements in the Volterra kernels that correspond to the products in $\bar{\boldsymbol{\mathcal{X}}}_j$.

4.3 Example : Third-Order Filter

To illustrate the previous developments, consider a third-order ($N = 3$) Volterra filter. According to the construction developed in section 5.2, the sub-filter vectors have the following forms. The sub-filter vector corresponding to the set $I_1 = \emptyset$ has all even powers and

$$\bar{\boldsymbol{\mathcal{X}}}_1 = [1,\ X_1^2,\ X_2^2,\ \ldots,\ X_M^2]^T. \tag{23}$$

The other sub-filter vectors have three forms. One form has a single r.v. raised to an odd power. For example, if $I_2 = \{X_1\}$ then

$$\bar{\boldsymbol{\mathcal{X}}}_2 = [X_1,\ X_1X_2^2,\ X_1X_3^2,\ \ldots,\ X_1X_M^2,\ X_1^3]^T. \tag{24}$$

Another form has pairs of r.v.'s raised to odd powers. For example, if $I_3 = \{X_1, X_2\}$ then the sub-filter $\bar{\boldsymbol{\mathcal{X}}}_3$ has one element,

$$\bar{\boldsymbol{\mathcal{X}}}_3 = [X_1X_2]. \tag{25}$$

The last form has triples of r.v.'s raised to odd powers. If $I_4 = \{X_1, X_2, X_3\}$ then

$$\bar{\boldsymbol{\mathcal{X}}}_4 = [X_1X_2X_3]. \tag{26}$$

The size of each sub-filter vector is given by (19) and the size of the largest sub-filter vector is $M+1$. Hence, for the third-order case, the complexity of the parameter estimation problem for the largest sub-filter is that of an order $M+1$ FIR filter. The savings in this case is significant since the total number of elements in the Volterra kernels is $\frac{(M+3)!}{M!\,3!} \approx \frac{M^3}{6}$ for large M. For example, if $M = 12$, then the largest sub-filter has only 13 parameters whereas the full Volterra filter has 455 parameters. The computation of the optimal Volterra kernels using the sub-filter structure requires solving several systems of 13 or fewer simultaneous equations (22) while the standard Volterra filter requires solving one system of 455 simultaneous equations (21).

4.4 Estimation from Finite Data Records

The PVF structure is derived assuming the sub-filters are orthogonal to each other. In practice exact orthogonality will not be obtained with random sequences and finite data records because the expectation is replaced by a finite sum. That is, assuming ergodicity, the true correlation matrix is replaced with a sample estimate. The lack of orthogonality introduces an error in the parameter estimates. This error is avoided by using maximum length PRMS. In this case, the exact correlation matrix is obtained over one period of the sequence. If the PRMS level values are chosen symmetrically about the origin, then $A3$ holds in a deterministic sense and the sub-filters are exactly orthogonal to each other over one period.

For example, a third-order filter with memory M requires a four-level MLS of degree M to satisfy the *p.e.* requirement. A symmetric sequence is obtained by choosing the four levels as $+a$, $-a$, $+b$, and $-b$ for $a, b \in \mathbb{R}$. The length of one period of the MLS in this example is $\Gamma = 4^M$.

While exact orthogonality is obtained over one period of an maximum length PRMS, for large memory Volterra filters the required sequence length may be prohibitively long. In this case a shorter or truncated PRMS often yields acceptable estimates. However, as in the random sequence case, errors result from lack of orthogonality between sub-filters.

Another issue is the ill-conditioned nature of the Volterra estimation problem [14]. Ill-conditioning can lead to numerical difficulties as well as erroneous estimates if observation noise is present. The condition number[1] of the correlation matrix $E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)$ grows as the memory and order of non-linearity increase. Unlike a linear filter driven by white noise inputs, when $N > 1$ the condition number depends on the memory length of the filter. In general, the ill-conditioning is even worse

[1] Note that the calculation of the condition number is greatly simplified if the input is uncorrelated and symmetric. If $\{X_k\}$ satisfies $A1$, $A2$, and $A3$ then $\mathrm{cond}_2[E(\bar{\mathcal{X}}\bar{\mathcal{X}}^T)] = \max_{1 \le j \le s} \mathrm{cond}_2[E(\bar{\mathcal{X}}_j\bar{\mathcal{X}}_j^T)]$.

when $E(\bar{\boldsymbol{\mathcal{X}}}\bar{\boldsymbol{\mathcal{X}}}^T)$ is replaced by sample estimates obtained from finite data records. Ill-conditioning problems are often much less severe in the PVF structure because the sub-filter correlation matrices $E(\bar{\boldsymbol{\mathcal{X}}}_j\bar{\boldsymbol{\mathcal{X}}}_j^T)$ are much smaller than the full Volterra filter correlation matrix.

5. Numerical Study

Second and third-order Volterra filters are simulated and identified using Gaussian white noise (GWN), symmetric RMS, and symmetric PRMS excitation sequences to demonstrate application of the PVF structure. These examples also serve to illustrate the errors associated with the use of finite data records. In all simulations the input power is chosen so that the power delivered to each kernel is approximately equal.

First, a second-order filter with memory length 30 is simulated. This second-order filter contains 496 unique kernel parameters. However, the largest sub-filter size in the PVF structure is only 31. An input excitation of length 1×10^4 and average input power 0.66 is applied to the Volterra filter. Table 1 lists the normalized squared error for between the true and estimated kernels for GWN, RMS, and PRMS input sequences. For the n-th kernel, the normalized squared error is defined as

$$e_n^2 = \frac{\sum_{j_1,\ldots,j_n} \mid h^n_{j_1,\ldots,j_n} - \hat{h}^n_{j_1,\ldots,j_n} \mid^2}{\sum_{j_1,\ldots,j_n} \mid h^n_{j_1,\ldots,j_n} \mid^2}, \tag{27}$$

where $\hat{h}^n_{j_1,\ldots,j_n}$ is the estimated kernel. Fig. 1 depicts the estimated first and second-order Volterra kernels.

Next, a third-order filter with memory length 12 is simulated. This third-order filter has 455 unique parameters and the largest sub-filter size is 13. An input excitation of length 2×10^4 and average input power 0.33 is applied to the Volterra filter. The true and estimated kernels obtained using GWN, RMS, and PRMS input sequences are shown in Fig. 2. Table 2 lists the normalized squared errors for each case.

Inspection of the results given in Tables 1 and 2 and Fig. 1 and 2 shows that useful estimates are obtained using the PVF structure with relatively short data records. The errors present in the kernel estimates are due to non-orthogonality between sub-filters that results from using relatively short data records. While these errors are eliminated by using a full period of a PRMS, the required sequence lengths of 3^{30} and 4^{12} in the second and third order examples, respectively, are excessively long. In this simulation, little difference is evident between the estimates obtained by the different excitations. The results for the second-order filter are very good. Better results in the third-order case are expected if a longer data record is used.

Lastly, a third-order filter with memory length 8 is simulated. This third-order filter has 165 unique parameters and the largest sub-filter size is 9. An input excitation of length $4^8 \approx 6.5 \times 10^4$ and average input power 0.33 is applied to the Volterra filter. In this case, one full period of a 4-level maximum length PRMS is used. The true and estimated kernels corresponding to GWN, RMS, and PRMS are shown if Fig. 3. Table 3 lists the normalized squared errors for each case.

Table 3 and Fig. 3 demonstrate that exact results are obtained by the PVF structure with one full period of a PRMS in the absense of observation noise. The errors present in the kernel estimates in the random input cases are once again a result of non-orthogonality between sub-filters due to the finite data record.

6. Conclusions

In this paper, two methods for estimating the Volterra kernels are discussed. In general, the optimal estimate of the Volterra kernels requires the solution of a large system of simultaneous linear equations. If the input sequence is sufficiently uncorrelated, the estimation problem is greatly simplified using the Kronecker product structure filter because the inverse of the large correlation matrix is expressed as a Kronecker product of small matrices.

Assuming the input sequence is sufficiently uncorrelated and symmetric, a parallel decomposition of the Volterra filter is derived that reduces the large estimation problem to a set of parallel sub-problems. Each sub-problem is significantly smaller than the original estimation problem and is easily computed. Simulations demonstrate the utility of this method.

Appendix A.

In the following, the number of r-combinations of an n-set is denoted as $C(n,r)$ as defined in (18). For illustration, first consider the number of unique products in the N-th order Volterra kernel of memory length M. That is, the number of products of the form $\prod_{i=1}^{M} X_i^{n_i}$ with the constraint $\sum_{i=1}^{M} n_i = N$, $n_i \geq 0$, $1 \leq i \leq M$. This number is simply the number of N-selections of an M-set which is well-known to be equal to $C(N+M-1, M-1)$ [4]. The total number of products in all kernels up to the N-th order is determined by requiring $\sum_{i=1}^{M} n_i \leq N$. It is easily verified that this is equivalent to the constraint $\sum_{i=1}^{M+1} n_i = N$, where $n_i \geq 0$ for $1 \leq i \leq M+1$. Hence, the total number of products in an N-th order Volterra series with memory M is $C(N+M, M)$.

Now consider the problem at hand. $\boldsymbol{\mathcal{X}}_j$ is a set of products containing exactly $\text{card} I_j = k$ r.v.'s raised to an odd power. Recall I_j is the set of r.v.'s raised to odd powers. Let $\bar{I}_j$ denote the remaining set of r.v.'s raised to even powers. Each $\alpha \in \boldsymbol{\mathcal{X}}_j$ is written as

$$\alpha = \prod_{X_i \in I_j} X_i^{2n_i+1} \prod_{X_l \in \bar{I}_j} X_l^{2n_l}, \tag{28}$$

with the constraint

$$\sum_{i:X_i \in I_j} (2n_i + 1) \; + \; \sum_{l:X_l \in \bar{I}_j} 2n_l \; \leq \; N, \tag{29}$$

where $n_i, n_l \geq 0$ for all indices. Since there are exactly k r.v.'s in the set I_j, the last constraint is equivalent to

$$\sum_{i:X_i \in I_j} 2n_i \; + \; \sum_{l:X_l \in \bar{I}_j} 2n_l \; \equiv \; \sum_{i=1}^{M} 2n_i \; \leq \; N - k. \tag{30}$$

Dividing both sides by two gives

$$\sum_{i=1}^{M} n_i \; \leq \; [\tfrac{N-k}{2}], \tag{31}$$

where $[\cdot]$ is the integer part of the argument. Hence, the discussion earlier shows that

$$\mathrm{card}\boldsymbol{\mathcal{X}}_j \; = \; C(M + [\tfrac{N-k}{2}], M). \tag{32}$$

References

[1] S. Chen, S. A. Billings and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, Vol. 50, pp. 1873-1896, 1989.

[2] P. Bondon, et al, "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.

[3] P. Bondon, M. Benidir and B. Picinbono, "Polynomial estimation of the amplitude of a signal," *Proc. ICASSP-92*, San Francisco, 1992.

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

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

[6] S. Klein and S. Yasui, "Nonlinear system analysis with non-gaussian white stimuli: general basis functionals and kernels," *IEEE Trans. on Info. Theory*, Vol. IT-25, No. 4, pp. 495-500, July 1979.

[7] T. Koh and E. J. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. on ASSP*, Vol. 33, pp. 1445-1455, Dec. 1985.

[8] M. J. Korenberg, "Orthogonal approaches to time-series analysis and system identification," *IEEE SP Magazine*, Vol. 8, No. 3, pp. 28-43, 1991.

[9] Y. W. Lee and M. Schetzen, "Measurement of the Wiener kernels of a nonlinear system by cross-correlation," *Int. J. Control*, Vol. 2, pp. 237-254, 1965.

[10] F. J. MacWilliams and N. J. A. Sloane, "Pseudo-random sequences and arrays," *Proc. IEEE*, Vol. 64, No. 12, pp. 1715-1728, 1976.

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

[12] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems. The White Noise Approach*, Plenum Press, New York, 1978.

[13] R. D. Nowak and B. Van Veen, "Random and pseudorandom excitation sequences for nonlinear system identification," *IEEE Winter Workshop on Nonlinear DSP*, session 2.1-5, Tampere, Finland, 1993.

[14] R. D. Nowak and B. Van Veen, "Random and pseudorandom excitation sequences for identification of the truncated Volterra series," *to appear in IEEE Trans. Signal Process.*.

[15] I. W. Sandberg, "Uniform approximation with doubly finite Volterra Series", *IEEE Trans. Signal Process.*, Vol. 40, pp. 1438-1442, June 1992.

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

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

TABLE 1

Normalized Squared Error of Estimates for 2nd-Order Filter with Memory 30

kernel	GWN error	RMS error	PRMS error
1	0.020	0.026	0.001
2	0.014	0.020	0.010

TABLE 2

Normalized Squared Error of Estimates for 3rd-Order Filter with Memory 12

kernel	GWN error	RMS error	PRMS error
1	0.073	0.168	0.036
2	0.068	0.014	0.026
3	0.073	0.084	0.102

TABLE 3

Normalized Squared Error of Estimates for 3rd-Order Filter with Memory 8

kernel	GWN error	RMS error	PRMS error
1	0.020	0.058	0.000
2	0.001	0.003	0.000
3	0.014	0.029	0.000

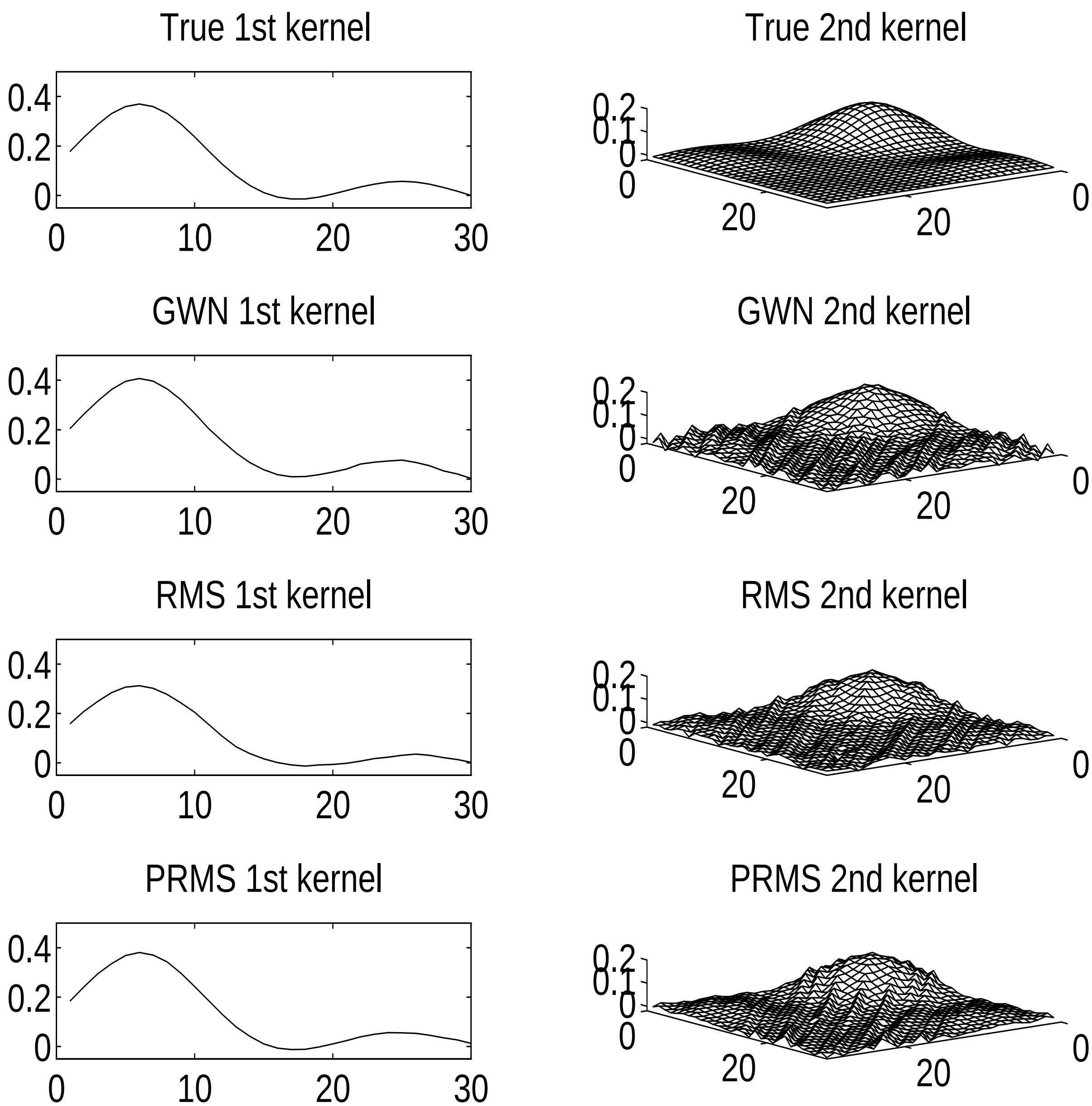

Figure 1. True and estimated kernels for 2nd-order Volterra filter with memory length 30.

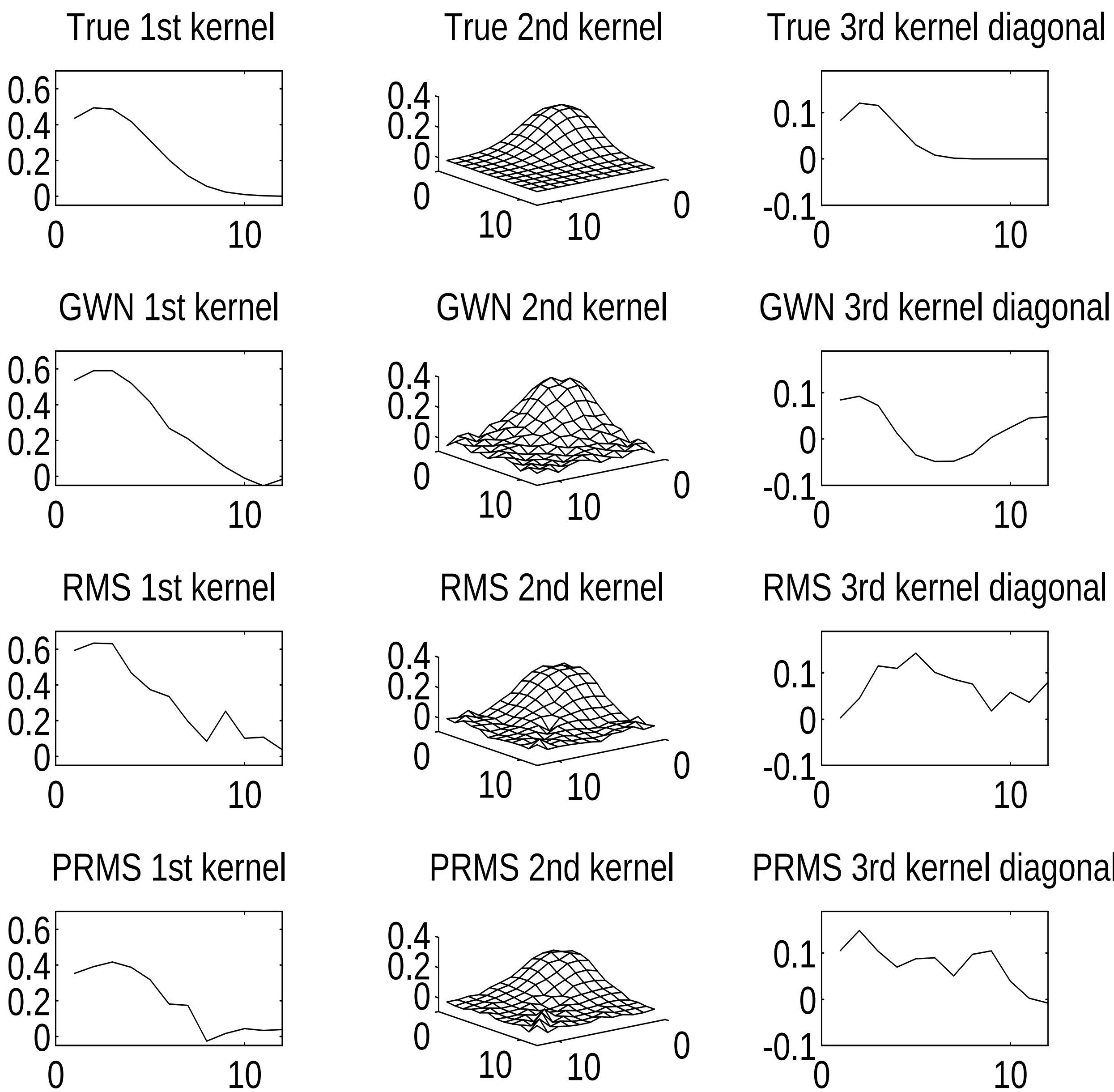

Figure 2. True and estimated kernels for 3rd-order Volterra filter with memory length 12.

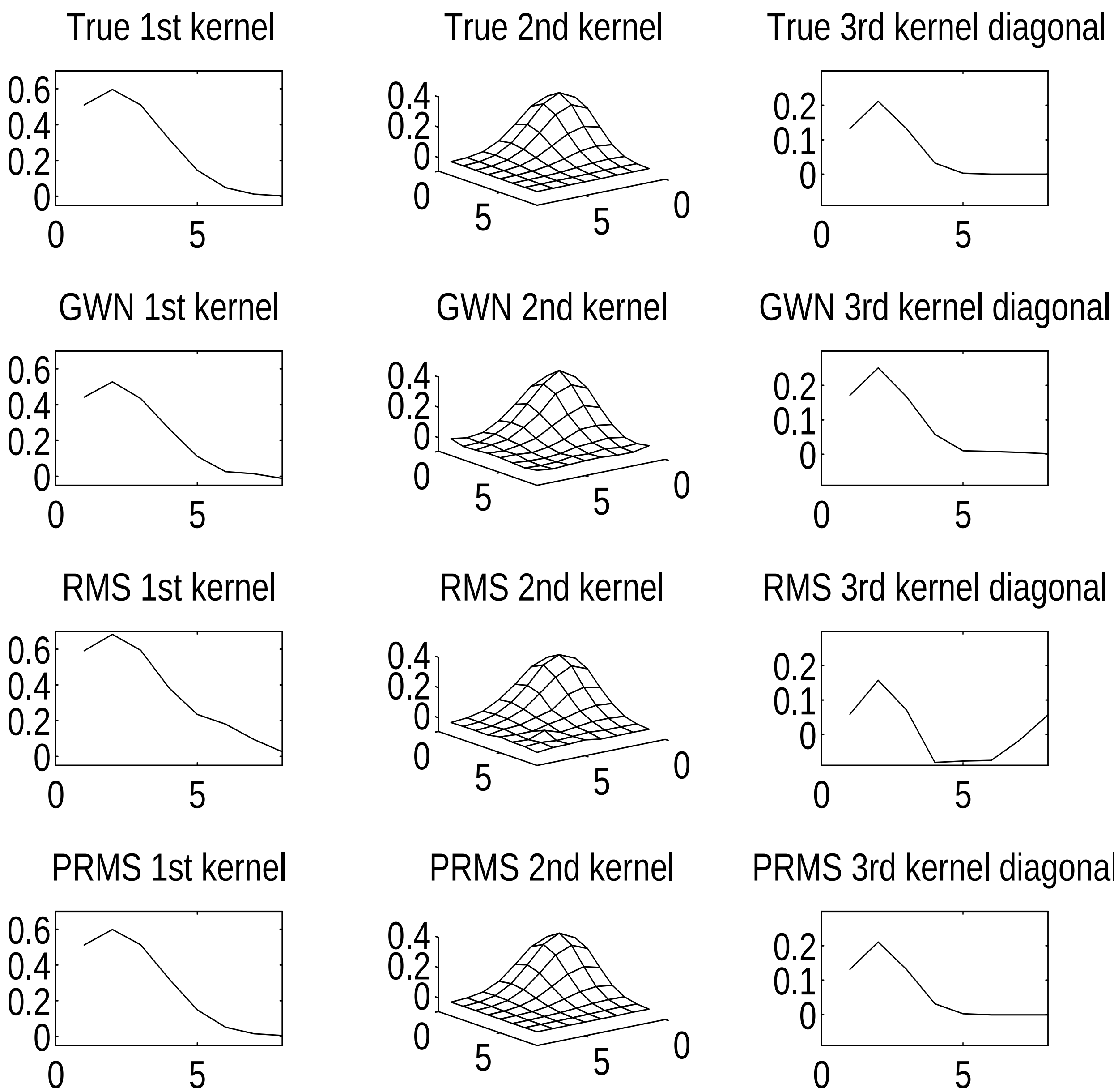

Figure 3. True and estimated kernels for 3rd-order Volterra filter with memory length 8.

Figure Captions:

Fig. 1. True and estimated kernels for 2nd-order Volterra filter with memory length 30.

Fig. 2. True and estimated kernels for 3rd-order Volterra filter with memory length 12.

Fig. 3. True and estimated kernels for 3rd-order Volterra filter with memory length 8.