Reduced Paramter Volterra Filters

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

Department of Electrical and Computer Engineering
University of Wisconsin-Madison, WI 53706 USA

Abstract

To reduce the number of parameters in the Volterra filter a tensor product basis approximation is considered. The approximation can be implemented much more efficiently than the original Volterra filter. In addition, because the design methods are based on partial characterization of the Volterra filter, the approximations are also useful in reducing the complexity of identification and modelling problems. Useful bounds are obtained on the approximation error.

1. Introduction

Let $\{X_j\}_{j=1}^m$ be real-valued random variables. The output of an nth order homogeneous Volterra filter applied to $\{X_j\}_{j=1}^m$ is a random variable

$$Y = \sum_{k_1,\ldots,k_n=1}^{m} h(k_1,\ldots,k_n) X_{k_1} \cdots X_{k_n}, \qquad (1)$$

where h, referred to as an nth order Volterra kernel, is deterministic and is real-valued. If $\mathrm{E}[X_j^{2n}] < \infty$, $j = 1,\ldots,m$, then it follows from Hölder's inequality that $\mathrm{E}[Y^2] < \infty$. Throughout this paper, such moment conditions are assumed whenever necessary. Without loss of generality h is assumed to be symmetric. That is, for every permutation $(\pi(1),\ldots,\pi(n))$ of $(1,\ldots,n)$, $h(k_{\pi(1)},\ldots,k_{\pi(n)}) = h(k_1,\ldots,k_n)$, and hence there are $\binom{n+m-1}{n}$ degrees of freedom, or parameters, in h, where $\binom{n+m-1}{n}$ is the binomial coefficient.

The large number of parameters associated with the Volterra filter results in significant computational obstacles except for modest values of m and n. Therefore, it is desirable to reduce the number of free parameters in the Volterra filter. Efforts to reduce Volterra filter complexity are proposed in [2, 4, 5, 7]. In this paper, the number of degrees of freedom, and hence the number of free parameters, in the Volterra filter is reduced using a tensor product basis approximation.

The tensor product basis approximation is a linear combination of tensor products of simple basis vectors. This method was originally introduced in [7] and the present paper represents a refinement of the ideas presented there. The approximation proposed in [2] is related to the tensor product approximations considered here, but only applies to quadratic filters. In [4, 5], the Volterra filter is approximated using a cascade structure composed of linear filters in series with memoryless nonlinearities. This approximation is quite different from those studied here.

There are several motivations for the tensor product basis approximations proposed herein.

1. Tensor product arises naturally in Volterra filters
2. Provides efficient implementation
3. Linear relationship between basis vectors and input
4. Can be designed from incomplete prior knowledge

The fact that there is a linear relationship between the basis vectors and the input makes the analysis of the approximation relatively straightforward and is used to establish useful approximation error bounds. With regard to point 4, it is shown that an appropriate tensor product basis can be designed from incomplete prior knowledge of the Volterra filter characteristics and corresponding approximation error bounds are derived. Such approximations are useful in identification or modelling problems when partial prior information is available.

The tensor product basis approximation and approximation error formulation are given in sections II and III respectively. Two methods for the design of the tensor product basis are presented in sections IV and V. The mean-square approximation error is bounded for each case. In section VI, the implementational complexity of the approximations is compared to the complexity of the original filter and it is shown that a significant savings is obtained.

2. Volterra Filter Approximations

The following convenient notation is employed. If $\mathrm{A} \in \mathbb{R}^{q \times p}$, then define $\mathrm{A}^{(1)} = \mathrm{A}$ and recursively define $\mathrm{A}^{(n)} = \mathrm{A}^{(n-1)} \otimes \mathrm{A}$ for $n > 1$, where $\otimes$ is the Kronecker (tensor) product [1]. Next let $\mathbf{h}$ be an m^n-vector constructed from the elements of the nth order Volterra kernel h and $\mathbf{X} = (X_1,\ldots,X_m)^T$ so that (1) is re-written as $Y = \mathbf{h}^T \mathbf{X}^{(n)}$.

Now let P denote the orthogonal projection matrix

* 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-G-0208, and by the National Institute of Health under Grant NINCDS #NSR 0116436.

0-7803-2431-5/95 $4.00 © 1995 IEEE

corresponding to an $r < m$ dimensional "approximation" subspace $\mathcal{U} \subset \mathbb{R}^m$ and consider approximating $\mathbf{h}$ by $\hat{\mathbf{h}} = \mathrm{P}^{(n)}\mathbf{h}$. Such an approximation is called a rank r^n tensor product basis approximation to $\mathbf{h}$. Note that $\hat{Y} = \hat{\mathbf{h}}^T\mathbf{X}^{(n)} = \mathbf{h}^T\mathrm{P}^{(n)}\mathbf{X}^{(n)} = \mathbf{h}^T(\mathrm{P}\mathbf{X})^{(n)}$. Hence, the output of the approximated Volterra filter is equivalent to the output of the original filter driven by the approximation of the original input. The advantage of using $\hat{\mathbf{h}}$ instead of $\mathbf{h}$ is that $\hat{\mathbf{h}}$ has fewer degrees of freedom. If $\mathrm{P} = \mathrm{U}\mathrm{U}^T$, where U is $m \times r$, then

$$\hat{Y} = (\mathbf{h}^T\mathrm{U}^{(n)})(\mathrm{U}^{(n)^T}\mathbf{X}^{(n)}) = \mathbf{h}_\mathrm{U}^T\mathbf{X}_\mathrm{U}^{(n)}, \qquad (2)$$

where $\mathbf{h}_\mathrm{U} = (\mathrm{U}^{(n)})^T\mathbf{h}$ and $\mathbf{X}_\mathrm{U} = \mathrm{U}^T\mathbf{X}$ is $r \times 1$. Also note that $\hat{\mathbf{h}}$ is constrained to lie in the space spanned by the columns of $\mathrm{U}^{(n)}$. The vector $\mathbf{h}_\mathrm{U}$ possesses same type of symmetry as $\mathbf{h}$ and hence has only $\binom{n+r-1}{n}$ degrees of freedom. Clearly, the reduction in complexity can be dramatic if $r \ll m$.

3. Approximation Error

The mean square output error is expressed and bounded as

$$\begin{aligned} \mathrm{E}[(Y-\hat{Y})^2] &= (\mathbf{Q}\mathbf{h})^T(\mathrm{Q}\,\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]\,\mathrm{Q}^T)(\mathbf{Q}\mathbf{h}), \\ &\le \|\mathbf{Q}\mathbf{h}\|_2^2\ \mathrm{tr}[\mathrm{Q}\,\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]\,\mathrm{Q}^T], \qquad (3) \end{aligned}$$

where $\mathrm{Q} \triangleq \mathrm{I}^{(n)} - \mathrm{P}^{(n)}$ is a projection matrix. From (3) it is easily seen that if null(Q) denotes the null space of Q, then the error is zero if either of the following conditions hold:

A1. $\mathbf{h} \in \mathrm{null}(\mathrm{Q})$
A2. $\mathrm{range}(\mathbf{X}^{(n)}) \subset \mathrm{null}(\mathrm{Q})$ w.p.1.

Of course, in practical situations A1 and A2 may not be exactly satisfied. Deviations in both conditions result in a non-zero output error that is characterized by the $2n$th order moments of the input process.

The bound in (3) leads naturally to the following two optimizations problems:

1) Find P to minimize $\|\mathbf{Q}\mathbf{h}\|_2^2$ subject to $\mathrm{rank}\mathrm{P} \le r < m$.
2) Find P to minimize $\mathrm{tr}(\mathrm{Q}\,\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]\,\mathrm{Q}^T)$ subject to $\mathrm{rank}\mathrm{P} \le r < m$.

Both optimizations are nonlinear and in general a closed form expression for a minimizer is not known to exist. The optimizations may be approached numerically; however, in general the problems are non-convex and finding a globally optimal solution may not be feasible. Hence, two suboptimal approaches are considered. One method attempts to minimize the filter error $\|\mathbf{Q}\mathbf{h}\|_2^2$ and the second method is based on the input error $\mathrm{tr}(\mathrm{Q}\,\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]\,\mathrm{Q}^T)$. In addition, the designs only require partial knowledge of the filter or input respectively. Explicit error bounds are obtained in each case.

4. Filter Error Design

In this section, a method for designing the tensor product basis using the filter error $\|\mathbf{Q}\mathbf{h}\|_2^2$ criteria is examined. This design approach is in general suboptimal and only requires prior knowledge of the filter's support in the Fourier domain.

Let H denote the n-dimensional Fourier transform of the kernel h and $\hat{H}$ denote the Fourier transform of the kernel approximation $\hat{h}$ (corresponding to $\hat{\mathbf{h}} = \mathrm{P}^{(n)}\mathbf{h}$). Let

$$B = [-w_2, -w_1] \cup [w_1, w_2],$$

denote a frequency range of interest, where $0 \le w_1 < w_2 \le 1/2$. Consider approximating H on $B^n \triangleq \underbrace{B \times \cdots \times B}_{n \text{ times}}$.

Define $\mathbf{w}(f) = (1, e^{i2\pi f}, \ldots, e^{i(m-1)2\pi f})^H$,

$$\mathrm{W} \triangleq \int_B \mathbf{w}(f)\mathbf{w}^H(f)df, \qquad (4)$$

and let $\mathbf{f} = (f_1, \ldots, f_n)$.

Proposition 1:

$$\int_{B^n} |H(\mathbf{f}) - \hat{H}(\mathbf{f})|^2 d\mathbf{f} = \mathbf{h}^T[\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)}]\mathbf{h}.$$

The proof of Proposition 1 involves some simple Kronecker product manipulations and is given in [8]. Proposition 1 leads to the bound,

$$\int_{B^n} |H(\mathbf{f}) - \hat{H}(\mathbf{f})|^2 d\mathbf{f} \le \|\mathbf{h}\|_2^2\ \|\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)}\|_2.$$

Thus, for this approximation a logical choice for P is an orthogonal projection matrix that minimizes

$$\|\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)}\|_2.$$

Theorem 1: The orthogonal projection matrix $\mathrm{P}_{r,\mathrm{W}}$ corresponding to the subspace spanned by r eigenvectors associated with the r largest eigenvalues of W minimizes

$$\|\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)}\|_2,$$

over all orthogonal projection matrices of rank $\le r$. Furthermore,

$$\|\mathrm{W}^{(n)} + \mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)}\mathrm{P}_{r,\mathrm{W}}^{(n)} - \mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}_{r,\mathrm{W}}^{(n)}\|_2 = \|\mathrm{W}\|_2^{n-1}\,\|\mathrm{W} - \mathrm{W}\mathrm{P}_{r,\mathrm{W}}\|_2.$$

A proof is given in [8] and is established using Kronecker product properties and the classical results of [3, 6] regarding low-rank matrix approximations. In general, the approximate rank of W is proportional to the time-bandwidth product $2m(w_2 - w_1)$. The results easily extend to more general sets than those with the form of B.

The following corollary, also proven in [8], bounds the error between $\mathbf{h}$ and $\hat{\mathbf{h}}$.

Corollary 1: If $\hat{\mathbf{h}} = \mathrm{P}_{r,\mathrm{W}}^{(n)}\mathbf{h}$ and $|H|^2 \leq \epsilon$ off B^n, then

$$\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2 \leq \|\mathbf{h}\|_2 \lambda_1^{n-1} \lambda_{r+1} + \epsilon,$$

where $\lambda_1 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} \geq \cdots \geq \lambda_m \geq 0$ are the eigenvalues of W.

5. Input Error Design

Define the norm of any $q \times 1$ random vector $\mathbf{Z}$, $q \geq 1$, as $\|\mathbf{Z}\| \triangleq \mathrm{tr}(\mathrm{E}[\mathbf{Z}\mathbf{Z}^T])^{1/2}$. Then

$$\mathrm{tr}(\mathrm{Q}\,\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]\,\mathrm{Q}^T) = \|\mathrm{Q}\mathbf{X}^{(n)}\|^2 = \|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2. \tag{5}$$

Hence, the objective of this section is to find a rank r orthogonal projector P so that PX is a good approximation to X in the sense of (5). Consider the following bound on the input approximation error (5).

Theorem 2: Let P be an orthogonal projection matrix on $\mathbb{R}^m$. If $\mathbf{X}$ is a random m-vector with finite $2n$th order moments, then there exists a constant $0 \leq \alpha_n < \infty$ such that

$$\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2 \leq n\,\alpha_n \|\mathbf{X}\|^{2(n-1)} \|\mathbf{X} - \mathrm{P}\mathbf{X}\|^2.$$

Theorem 2 is established by repeated application of Hölder's inequality and Kronecker product properties. A complete proof is given in [8]. The bound of Theorem 2 suggests the choice of P that minimizes $\|\mathbf{X} - \mathrm{P}\mathbf{X}\|^2 = \mathrm{tr}(\mathrm{R} - \mathrm{P}\mathrm{R} - \mathrm{R}\mathrm{P} + \mathrm{P}\mathrm{R}\mathrm{P})$, where $\mathrm{R} \triangleq \mathrm{E}[\mathbf{X}\mathbf{X}^T]$ is the autocorrelation matrix of X. Using the eigendecomposition $\mathrm{R} = \mathrm{U}\mathrm{D}\mathrm{U}^T$ and defining $\mathrm{C} = \mathrm{U}\mathrm{D}^{1/2}\mathrm{U}^T$ write

$$\begin{aligned} \mathrm{tr}(\mathrm{R} - \mathrm{P}\mathrm{R} - \mathrm{R}\mathrm{P} + \mathrm{P}\mathrm{R}\mathrm{P}) &= \mathrm{tr}((\mathrm{C} - \mathrm{C}\mathrm{P})^T(\mathrm{C} - \mathrm{P}\mathrm{C})), \\ &= \|\mathrm{C} - \mathrm{C}\mathrm{P}\|_{\mathrm{F}}^2, \end{aligned} \tag{6}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius matrix norm.

It is easily established using the result in [3] that a rank r orthogonal projection matrix minimizing (6) is the projection matrix $\mathrm{P}_{r,\mathrm{R}}$ onto the subspace spanned by the eigenvectors associated with the r largest eigenvalues of C or equivalently R. Theorem 2 implies that if $\mathrm{P}_{r,\mathrm{R}}\mathbf{X}$ is a good approximation to $\mathbf{X}$, in the mean-square sense, then $(\mathrm{P}_{r,\mathrm{R}}\mathbf{X})^{(n)}$ may be a good approximation of $\mathbf{X}^{(n)}$ in the same sense. Of course, "how good" depends on α_n and $\|\mathbf{X}\|$. In general, to determine α_n, knowledge of the second and $2n$th order moment of the each individual random variable in the vectors $\mathbf{X}$, $\mathrm{P}_{r,\mathrm{R}}\mathbf{X}$, and $(\mathrm{I} - \mathrm{P}_{r,\mathrm{R}})\mathbf{X}$ is necessary. However, if $\mathbf{X}$ is a linear transformation of independent, symmetric random variables, then α_n is obtained independent of $\mathrm{P}_{r,\mathrm{R}}$.

Theorem 3: If $\mathbf{X}$ is a linear transformation of a vector $\mathbf{U}$ of independent r.v.'s $U_1, \ldots, U_q$ with symmetric distributions $F_1, \ldots, F_q$, then a constant satisfying the inequality in Theorem 3 is given by $\alpha_n \triangleq \max_{j=1,\ldots,q} \alpha_{n,F_j}$, where α_{n,F_j} is a positive number satisfying

$$\mathrm{E}[U_j^{2n}] \leq \alpha_{n,F_j}\,\mathrm{E}[U_j^2]^n, \quad j = 1, \ldots, q. \tag{7}$$

Theorem 3 is proven by showing that, under the stated assumptions, every random variable of the form $Z = \sum_{j=1}^q c_j U_j$ satisfies $\mathrm{E}[Z^{2n}] \leq \alpha_n\,\mathrm{E}[Z^2]^n$. If C is an arbitrary $m \times q$ matrix and P is a projection matrix on $\mathbb{R}^m$, then the elements of $\mathbf{X} = \mathrm{C}\mathbf{U}$, $\mathrm{P}\mathbf{X}$, and $(\mathrm{I} - \mathrm{P})\mathbf{X}$ are simply linear combinations of the U process. Hence, α_n, as defined above, may be used in Theorem 2. The complete proof is also in [8].

Notice that under the assumptions of Theorem 3, the bound in Theorem 2 is computed using only the second order moments of $\mathbf{X}$ and the bounds (7) relating the 2nd and $2n$th order moments of the independent $\mathbf{U}$ process. The bounding constant α_n is easily determined for many common types of distributions. For example, if $\mathbf{X}$ is jointly Gaussian mean-zero, then $\alpha_n = \frac{(2n)!}{n!2^n}$. If $U_1, \ldots, U_q$ are independent, symmetric, uniformly distributed random variables, then $\alpha_n = \frac{3^n}{2n+1}$.

To quantify the quality of this approximation, consider the ratio $\frac{\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2}{\|\mathbf{X}^{(n)}\|^2}$. In the spirit of Theorems 2 and 3, it is of interest to establish bounds on this ratio in terms of the second order moments of $\mathbf{X}$ and the constant α_n. The details of the following two corollaries are found in [8]

Corollary 3.1: If $\mathbf{X}$ satisfies the hypothesis of Theorem 3, then

$$\frac{\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2}{\|\mathbf{X}^{(n)}\|^2} \leq n\,\alpha_n \frac{\|\mathbf{X}\|^{2(n-1)}\|\mathbf{X} - \mathrm{P}\mathbf{X}\|^2}{\sum_{i=1}^m \mathrm{E}^n\{X_i^2\}},$$

where α_n is the constant defined in Theorem 3. If $\mathrm{E}\{X_i^2\} = \sigma^2$, $i = 1, \ldots, m$, then

$$\frac{\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2}{\|\mathbf{X}^{(n)}\|^2} \leq n\,\alpha_n m^{n-1} \frac{\|\mathbf{X} - \mathrm{P}\mathbf{X}\|^2}{\|\mathbf{X}\|^2}.$$

Corollary 3.1 follows by noting that $\|\mathbf{X}^{(n)}\|^2 \geq \sum_{i=1}^m \mathrm{E}[X_i^2]^n$, by Jensen's inequality. The next corollary requires a stronger assumption, but also considerably tightens the bound.

Corollary 3.2: If $\mathbf{X} = \mathrm{C}\mathbf{U}$ where the random variables $\mathbf{U} = (U_1, \ldots, U_q)^T$ are independent and symmetric and $\mathrm{C} = \{c_{i,j}\}$ is $m \times q$ with $c_{i,j} \geq 0$, then

$$\frac{\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2}{\|\mathbf{X}^{(n)}\|^2} \leq n\,\alpha_n \frac{\|\mathbf{X} - \mathrm{P}\mathbf{X}\|^2}{\|\mathbf{X}\|^2},$$

where α_n is defined in Theorem 3.

This corollary is proved by exploiting independence and using Jensen's inequality.

6. Implementational Complexity

The main source of computational burden for the Volterra filter arises in the number of multiplications required per output. To study the relative computational efficiency of the tensor product basis approximations, the number multiplications required per output using the rank r^n tensor product approximation $\hat{\mathbf{h}}$ and original Volterra filter $\mathbf{h}$ is compared.

Two cases are considered. First, the "parallel" implementation of $\mathbf{h}$, in which all products of the input are computed for every output. To form all unique n-fold products of $\mathbf{X}$ requires $(n-1)\binom{n+m-1}{n}$ multiplications and another $\binom{n+m-1}{n}$ multiplications are required to compute the output. Second, consider the "serial" implementation, in which the input is a time-series. In this case, after initialization, only products involving the new input need be computed at each time step. The number of such products is given by the number of ways $n_1 \geq 1, n_2, \ldots, n_m \geq 0$ may be chosen so that $\sum_{i=1}^{m} n_i = n$ or equivalently the number of ways $n_1, n_2, \ldots, n_m \geq 0$ may be chosen so that $\sum_{i=1}^{m} n_i = n-1$ which is $\binom{n-1+m-1}{n-1}$. Hence, the number of multiplications required for a "serial" implementation of $\mathbf{h}$ is $\binom{n+m-1}{n} + (n-1)\binom{n+m-2}{n-1}$.

To study the complexity of the approximation $\hat{\mathbf{h}}$ recall that the output is computed with a $\binom{n+r-1}{n}$ parameter Volterra filter $\mathbf{h}_U$ and the transformed data vector $\mathbf{X}_U = \mathrm{U}^T\mathbf{X}$, where the columns of U span an r-dimensional subspace $\mathcal{U} \subset \mathbb{R}^m$ (2). To form $\mathbf{X}_U$ and all unique products in $\mathbf{X}_U^{(n)}$ requires $rm + (n-1)\binom{n+r-1}{n}$ multiplications (the first term corresponds to the transformation and the second corresponds to formation of the necessary products). With these products in hand, the output is computed with an additional $\binom{n+r-1}{n}$ multiplications. Note that due to the required transformation, no savings is available in the serial implementation using the approximation.

The exact ratios, denoted η_p and η_s, of the number of multiplications using $\hat{\mathbf{h}}$ versus $\mathbf{h}$, for parallel and serial implementations respectively, are given below.

$$\eta_p = \frac{\#\text{mults}(\hat{\mathbf{h}})}{\#\text{mults}(\mathbf{h})} = \frac{rm + n\binom{n+r-1}{n}}{n\binom{n+m-1}{n}}, \tag{8}$$

and

$$\eta_s = \frac{\#\text{mults}(\hat{\mathbf{h}})}{\#\text{mults}(\mathbf{h})} = \frac{rm + n\binom{n+r-1}{n}}{\binom{n+m-1}{n} + (n-1)\binom{n+m-2}{n-1}}. \tag{9}$$

To gain some insight into the behavior of these ratios as a function of subspace dimension, consider the following large m asymptotic analysis. Assume that $n \geq 2$ and let $0 < \rho \leq 1$ be fixed. Let $r = \lceil \rho m \rceil$, the smallest integer greater than or equal to ρm. The number ρ is the ratio of the approximation subspace dimension to m. Using $(1 + \frac{n}{m-1})^{m-1} \sim e^n$, $(1+\frac{n}{m-1})^{n+1/2} \sim 1$, and Stirlings formula $m! \sim \sqrt{2\pi}\, m^{m+1/2}\, e^{-m}$, it follows that

$$\frac{\binom{n+r-1}{n}}{\binom{n+m-1}{n}} \sim \rho^n, \qquad \frac{rm}{\binom{n+m-1}{n}} \sim n!\,\frac{\rho}{m^{n-2}}. \tag{10}$$

Using these approximations

$$\eta_p = \frac{\#\text{mults}(\hat{\mathbf{h}})}{\#\text{mults}(\mathbf{h})} \sim \frac{(n-1)!\,\rho}{m^{n-2}} + \rho^n, \tag{11}$$

and

$$\eta_s = \frac{\#\text{mults}(\hat{\mathbf{h}})}{\#\text{mults}(\mathbf{h})} \sim \frac{n!\,\rho}{m^{n-2}} + n\rho^n = n\,\eta_p. \tag{12}$$

In the special case of quadratic filters, further simplification is obtained by applying the method proposed in [2].

7. Conclusions

The tensor product basis Volterra filter approximations considered in this paper offer considerable savings in terms of implementation complexity. Also, because the design methods are based on incomplete prior knowledge of the filter (*i.e.*, frequency support) or input (*i.e.*, second order moments only) such approximations are also useful in reducing the complexity of Volterra filters for identification and modelling problems. Error bounds are derived that quantify the mean square output error of such approximations.

References

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

[2] H.-H. Chiang, C. L. Nikias, and A. N. Venetsanopoulos, "Efficient implementations of quadratic digital filters," *IEEE Trans. ASSP*, Vol. 34, pp. 1511-1528, 1986.

[3] G. Eckart and G. Young, "The approximation of one matrix by another of lower rank," Psychometrika 1: pp. 211-218, 1936.

[4] W. Frank, "MMD - An efficient approximation to the 2nd order Volterra filter," *Proc. of ICASSP*, vol. 3, pp. 517-520, 1994.

[5] M. J. Korenberg, "Parallel Cascade Identification and kernel estimation for nonlinear systems," Ann. Biomed. Eng., vol. 19, 1991.

[6] L. Mirsky, "Symmetric gauge functions and unitarily invariant norms," *Quart. J. Math. Oxford* (2), 11, pp. 50-59, 1960.

[7] R. Nowak and B. Van Veen, "Volterra filtering with spectral constraints," *Proc. of ICASSP*, vol. 4, pp. 137-140, 1994.

[8] R. Nowak and B. Van Veen, "Volterra filter approximatations," University of Wisconsin-Madison ECE Technical Report ECE-94-16. Also "Volterra filter approximations," submitted for publication to *IEEE Trans. SP*