Tensor Product Basis Approximations for 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 — This paper studies approximations for a class of nonlinear filters known as Volterra filters. Although the Volterra filter provides a relatively simple and general representation for nonlinear filtering, often it is highly over-parameterized. Due to the large number of parameters, the utility of the Volterra filter is limited. The over-parameterization problem is addressed in this paper using a tensor product basis approximation (TPBA). In many cases a Volterra filter may be well approximated using the TPBA with far fewer parameters. Hence, the TPBA offers considerable advantages over the original Volterra filter in terms of both implementation and estimation complexity. Furthermore, the TPBA provides useful insight into the filter response. This paper studies the crucial issue of choosing the approximation basis. Several methods for designing an appropriate approximation basis and error bounds on the resulting mean-square output approximation error are derived. Certain methods are shown to be nearly optimal.

I. Introduction

Volterra filters have received increasing attention in the recent signal processing literature and have been applied to many signal processing problems such as signal detection [17, 19], estimation [2, 17], adaptive filtering [12], and system identification [6, 8, 10, 11, 14]. The Volterra filter is motivated by Weierstrauss' Theorem, which shows that a Volterra filter provides an arbitrarily accurate approximation to a given continuous function on a compact set. One of the major drawbacks of Volterra filters is the large number of parameters associated with such structures. In this paper, it is shown how the Volterra filter can be approximated to yield parsimonious filter structures that are adequately flexible for large classes of problems. The general nth order Volterra filter is a degree n polynomial mapping from $\mathbb{R}^m \rightarrow \mathbb{R}$. To simplify the presentation, this paper focuses on the homogeneous nth order Volterra filter. The homogeneous nth order Volterra filter is a linear combination of n-fold products of the inputs. Since the general nth order Volterra filter is the sum of linear (1st order) through homogeneous nth order Volterra filters, extensions to the general case are straightforward.

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

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

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

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 set of indices $k_1,\ldots,k_n$ and 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 limit its practical utility to problems involving only modest values of m and n. Therefore, it is desirable to reduce the number of free parameters in the Volterra filter in situations when m and/or n is large. Efforts to reduce Volterra filter complexity are proposed in [1, 4, 6, 9, 10, 11, 14, 20]. Each of these references adopt one of two basic approaches.

In the first approach [6, 9], the Volterra filter is approximated using a cascade structure composed of linear filters in series with memoryless nonlinearities. The output of such cascade models is not linear with respect to the parameters and therefore identifying the globally optimal model parameters is a nonlinear estimation problem. Both [6, 9] suggest algorithms for estimating cascade model parameters, however neither method guarantees globally optimal solutions. This is a drawback of the cascade structure.

The second approach, which is the focus of this paper, is termed the *tensor product basis approximation* TPBA method. The TPBA represents the Volterra filter as a linear combination of tensor products of simple basis vectors. In contrast to the cascade methods, the output of the TPBA is linear in the parameters. Therefore, estimation of the TPBA parameters is a linear estimation problem and hence conditions for global optimality and uniqueness of the estimate are easily established. There are several motivations for the TPBA.

1. Tensor product arises naturally in Volterra filters
2. Provides efficient implementation
3. Reduced parameterization for adaptive filtering and identification problems
4. Provides useful insight into filter behavior

The use of such approximations is not new. Originally, Wiener [20] proposed using a tensor product of the Laguerre functions as a multidimensional basis for representation of the Wiener kernels of a nonlinear system. Implementations and representations of discrete Volterra kernels using the discrete Laguerre basis has been recently examined in [1, 10, 11]. Although the Laguerre basis has many desirable properties, other basis choices are possible. Hence, it is of interest to determine appropriate bases for different nonlinear filtering problems. Choosing a basis for the TPBA is analogous to choosing a filter structure and hence the choice of basis and parameter estimation are separate issues. The focus of this paper is choosing a basis. Methods to determine optimal bases for quadratic filters are given in [4, 11]. In [4] an SV-LU quadratic kernel decomposition is used to implement quadratic filters in an efficient fashion. The notion of the "principal dynamic modes" of a quadratic system is introduced in [11]. The principal dynamic modes are obtained from the eigendecomposition of a matrix composed of the first and second order kernels. Both methods [4, 11] apply only to quadratic Volterra filters.

The basis design methods of this paper are not restricted to quadratic filters and hence extend existing results. They are based on complete or partial characterization of the filter *or input* and are related to two distinct nonlinear optimization problems. The use of input information in the design process appears to be a new contribution. The design methods are based on suboptimal procedures aimed at solving the two optimization problems. Bounds on the approximation error are derived for each method. Two of the design methods are shown to be nearly optimal in the sense that the resulting approximation error is within a factor of the global minimum and conditions that guarantee global optimality are given. The TPBA also provides a practical framework in which to address the trade-off between model complexity and performance. The error performance of the TPBA can be bounded for a specified model complexity (basis dimension) using the approximation error bounds. Alternatively, given a desired error performance, the the required complexity of the TPBA can be deduced.

The paper is organized as follows. The TPBA is introduced in section II and two design criteria for determining an appropriate basis, based on a filter or input error, are proposed. In section III, the filter error criterion is examined. Two basis design methods aimed at minimizing the filter error are developed and the approximation error is bounded for each case. One method is shown to be nearly optimal. The input error criterion is studied in section IV. Two methods are proposed that attempt to minimize the input error and error bounds are derived. One of the input error methods is also shown to be nearly optimal. The implementational complexity of the TPBA is compared to the homogeneous nth order Volterra filter in section V. In section VI, some illustrative examples of the proposed methods are given.

II. Volterra Filter Approximation via Tensor Product Bases

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 [3]. If $\mathrm{A}_i \in \mathbb{R}^{q_i \times p_i}$, $i = 1, \ldots, n$, then $\bigotimes_{i=1}^{n} \mathrm{A}_i = \mathrm{A}_1 \otimes \cdots \otimes \mathrm{A}_n$. Next let $\mathbf{h}$ be an m^n-vector composed of the elements in the 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 corresponding to an $r < m$ dimensional "approximation" subspace $\mathcal{U} \subset \mathbb{R}^m$ and consider approximating $\mathbf{h}$ by $\hat{\mathbf{h}} = \mathrm{P}^{(n)}\mathbf{h}$. This approximation is called a rank r^n TPBA to $\mathbf{h}$. Note that

$$\hat{Y} = \hat{\mathbf{h}}^T \mathbf{X}^{(n)} = \mathbf{h}^T \mathrm{P}^{(n)} \mathbf{X}^{(n)} = \mathbf{h}^T (\mathrm{P}\mathbf{X})^{(n)}. \tag{2}$$

Hence, the output of the approximated Volterra filter is equivalent to the output of the original filter driven by the approximation $\mathrm{P}\mathbf{X}$ of the input. This interpretation of the TPBA is useful in designing the basis using knowledge of the input.

Expressing P as $\mathrm{P} = \mathrm{U}\mathrm{U}^T$, where U is $m \times r$, shows that

$$\hat{Y} = (\mathbf{h}^T \mathrm{U}^{(n)})({\mathrm{U}^{(n)}}^T \mathbf{X}^{(n)}) = \mathbf{h}_{\mathrm{U}}^T \mathbf{X}_{\mathrm{U}}^{(n)}, \tag{3}$$

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 $\mathrm{U}^{(n)}$. Both the vector $\mathbf{h}_{\mathrm{U}}$ and $\mathbf{X}_{\mathrm{U}}^{(n)}$ possess the same types of symmetry as $\mathbf{h}$ and $\mathbf{X}^{(n)}$. Therefore, the Volterra filter $\mathbf{h}_{\mathrm{U}}^T \mathbf{X}_{\mathrm{U}}^{(n)}$ may be implemented in an efficient fashion that accounts for these symmetries. The key point is that $\hat{\mathbf{h}}$ has only $\binom{n+r-1}{n}$ degrees of freedom, far fewer degrees of freedom than $\mathbf{h}$. The degrees of freedom a measure of filter complexity. This complexity affects filter estimation as well as filter implementation. In section V (15), it is shown that, for $m, r \gg n$, the ratio of degrees of freedom in $\hat{\mathbf{h}}$ to degrees of freedom in $\mathbf{h}$

$$\frac{\binom{n+r-1}{n}}{\binom{n+m-1}{n}} \approx \left(\frac{r}{m}\right)^n.$$

Clearly, the reduction in complexity can be dramatic. Several possible applications of the TPBA are outlined next.

Filter Implementation

From an implementation perspective, the cost of computing the transformation $\mathbf{X}_{\mathrm{U}} = \mathrm{U}^T\mathbf{X}$, forming the products in $\mathbf{X}_{\mathrm{U}}^{(n)}$, and computing $\mathbf{h}_{\mathrm{U}}^T\mathbf{X}_{\mathrm{U}}^{(n)}$ is often much less than the cost of forming the products in $\mathbf{X}^{(n)}$ and computing $\mathbf{h}^T\mathbf{X}^{(n)}$. Note that both filters, $\mathbf{h}^T\mathbf{X}^{(n)}$ and $\mathbf{h}_{\mathrm{U}}^T\mathbf{X}_{\mathrm{U}}^{(n)}$, possess the symmetries discussed previously and therefore may be computed in an efficient fashion that accounts for these symmetries. The implementation complexity is examined in section V.

Adaptive Filtering and System Identification

If U is determined from prior knowledge, the TPBA is useful for adaptive filtering and identification problems. In adaptive filtering applications, the TPBA provides an flexible filter structure with far fewer adaptive degrees of freedom than the original Volterra filter. In nonlinear system identification problems, the TPBA has fewer parameters than the original Volterra filter structure and hence more reliable parameter estimates are obtained from finite, noisy data records. Methods for determining an appropriate basis based on incomplete prior knowledge of the filter or input are discussed in sections III and IV respectively. The application of the TPBA to system identification is discussed in the examples of section VI.

Filter Analysis

Note that U also determines a null space of the TPBA filter. That is, any input $\mathbf{X}$ lying in the linear subspace that is orthogonal to the columns of U produces zero output. Hence, given a filter $\mathbf{h}$, a good approximating basis U provides information about the filter response and thus the TPBA is also a useful analysis tool. For example, if the basis U spans a bandpass subspace in the frequency domain, then it may be inferred that $\mathbf{h}$ only responds to the input component in the passband and hence is bandlimited. Another interesting application is demonstrated in Example 1 of section VI of this paper where it is shown that if the basis U consists of a single vector, then $\mathbf{h}$ has a cascade structure.

The main goal of this paper is to suggest several methods for choosing an appropriate basis for the TPBA and to bound the corresponding approximation errors. Several design methods are studied. The methods are based on complete or partial knowledge of either the filter or the input process. Specifically, the design methods for the basis U attempt to minimize the filter error:

$$e_f \triangleq \|\mathbf{h} - \hat{\mathbf{h}}\|_2 = \|(\mathrm{I}^{(n)} - \mathrm{P}^{(n)})\mathbf{h}\|_2, \tag{4}$$

where $\|\cdot\|_2$ denotes the l_2 vector norm, or the input error:

$$e_i \triangleq \|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\| = \mathrm{tr}(\mathrm{E}[(\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)})(\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)})^T])^{1/2}, \tag{5}$$

where E is the expectation operator and tr is the trace operator. The input error arises naturally from the input interpretation (2) of the TPBA. It is easily verified that the mean square output error is bounded by

$$\mathrm{E}[(Y - \hat{Y})^2] \leq e_f^2\, e_i^2. \tag{6}$$

Hence, minimizing either error reduces the bound on the mean-square output error of the filter approximation.

From (6) it is easily seen that if $\mathrm{null}(\mathrm{I}^{(n)} - \mathrm{P}^{(n)})$ denotes the null space of $\mathrm{I}^{(n)} - \mathrm{P}^{(n)}$, then the error is zero if either of the following conditions hold:

A1. $\mathbf{h} \in \mathrm{null}(\mathrm{I}^{(n)} - \mathrm{P}^{(n)})$

A2. $\mathrm{range}(\mathbf{X}^{(n)}) \subset \mathrm{null}(\mathrm{I}^{(n)} - \mathrm{P}^{(n)})$ 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 $\mathbf{h}$, P, and the $2n$th order moments of the input process.

The next two sections consider the following two optimizations problems:

1) Find P to minimize $e_f = \|(\mathrm{I}^{(n)} - \mathrm{P}^{(n)})\mathbf{h}\|_2$ subject to $\mathrm{rank}\mathrm{P} \leq r < m$.

2) Find P to minimize $e_i = \|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|$ subject to $\mathrm{rank}\mathrm{P} \leq r < m$.

One could try to solve both optimization problems and then choose a final basis for the TPBA by combining these results, however this approach is not pursued in the present work.

Can an optimal projection matrix be found in either case? Since the set of rank $\leq r$ orthogonal projection operators on $\mathbb{R}^m$ is compact and because the errors are continuous functions of the projection matrix there is no problem with the existence of a minimizer (see Appendix B, proof of Theorem 1). However, both optimizations are nonlinear and 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. Hence, finding a globally optimal solution may not be feasible. In this paper, several suboptimal approaches are considered. The methods vary in computational complexity and required prior knowledge. Bounds are obtained on the approximation error in each case and two methods are shown to be nearly optimal.

III. Filter Error Designs

In this section, two approaches to designing the tensor product basis based on the filter error are examined. The first approach is in general suboptimal and only requires prior knowledge of the filter's support in the Fourier domain. The second approach requires complete knowledge of the filter and is shown to be nearly optimal in the sense that the resulting filter error $\|(\mathrm{I}^{(n)} - \mathrm{P}^{(n)})\mathbf{h}\|_2$ is within a factor of $\sqrt{n}$ of the global minimum.

A. Method I: Frequency Domain Filter Error Design

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

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

$$\mathrm{W} \triangleq \int_B \mathbf{w}(f)\mathbf{w}^H(f)df, \tag{7}$$

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 not given here. A complete proof of the proposition is found in [15]. Proposition 1 leads to the bound,

$$\int_{B^n} |H(\mathbf{f}) - \hat{H}(\mathbf{f})|^2 d\mathbf{f} \leq \|\mathbf{h}\|_2^2\ \|\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)}\|_2, \tag{8}$$

where the second norm on the right hand side of (8) is the matrix 2-norm. 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 $\leq r$. 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 Appendix B. If $w_1 = 0$, then the eigenvectors of W are the discrete prolate spheroidal sequences [18] and it can be shown that for large m the first $2mw_2$ eigenvalues of W are close to unity and the remainder are approximately zero. Hence, in such cases a rank r^n, $r = 2mw_2$, TPBA is possible with negligible error. In general, the rank of W is proportional to the time-bandwidth product $2m(w_2 - w_1)$. Note that the results easily extend to more general sets than those with the form of B. The following corollary summarizes the results. The proof follows in a straightforward manner using Parseval's Theorem and Theorem 1. The details of the proof are given in [15].

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.

B. Method II: SVD Based Filter Error Design

This design method is based on the singular value decomposition and directly utilizes the filter $\mathbf{h}$. The following theorem suggests a nearly optimal choice of P.

Theorem 2: Let $m, n > 1$ and let h be an nth order symmetric kernel. Define the $m^{n-1} \times m$ matrix

$$\mathrm{H} \triangleq [\mathrm{H}_1^T, \ldots, \mathrm{H}_m^T]^T,$$

where

$$\mathrm{H}_i = \begin{bmatrix} h(i,1,\ldots,1,1) & \cdots & h(i,1,\ldots,1,m) \\ h(i,1,\ldots,2,1) & \cdots & h(i,1,\ldots,2,m) \\ \vdots & & \vdots \\ h(i,m,\ldots,m,1) & \cdots & h(i,m,\ldots,m) \end{bmatrix}, \quad i = 1, \ldots, m.$$

Let $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ denote the singular values of H and let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be the associated right singular vectors. Furthermore, for $r \leq m$, let $\wp_r$ denote the compact set of all $m \times m$ orthogonal projection matrices with rank $\leq r$, and let $\mathrm{P}_{r,\mathrm{H}} \in \wp_r$ be the orthogonal projector onto $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. Then

$$\sum_{i=r+1}^{m} \sigma_i^2 \leq \min_{\mathrm{Q}_1,\ldots,\mathrm{Q}_n \in \wp_r} \| (\bigotimes_{i=1}^{n} \mathrm{Q}_i)\mathbf{h} - \mathbf{h} \|_2^2 \leq \| \mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h} - \mathbf{h} \|_2^2 \leq n \sum_{i=r+1}^{m} \sigma_i^2.$$

Theorem 2 is proved in Appendix C and is an extension of the SV-LU quadratic filter decomposition of [4] to the general Volterra filter case. Note that choosing $\mathrm{P}_{r,\mathrm{H}}$ in this fashion results in an approximation error $\| \mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h} - \mathbf{h} \|_2$ that is within a factor of $\sqrt{n}$ of the global minimum. The following three corollaries summarize some important properties of the approximation $\mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h}$.

Corollary 2.1: There exists a rank r orthogonal projection matrix P such that $\mathrm{P}^{(n)}\mathbf{h} = \mathbf{h}$ if and only if $\mathrm{rank}\mathrm{H} \leq r$. Moreover, if $\mathrm{rank}\mathrm{H} \leq r$, then $\mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h} = \mathbf{h}$.

Proof: If $\mathrm{rank}\mathrm{H} \leq r$, then $\sum_{i=r+1}^{m} \sigma_i^2 = 0$. Hence, by Theorem 2 this implies that $\| \mathbf{h} - \mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h} \|_2^2 = 0$. On the other hand, if $\mathrm{rank}\mathrm{H} > r$, then for every rank r orthogonal projection matrix P, $\| \mathbf{h} - \mathrm{P}^{(n)}\mathbf{h} \|_2^2 = \| \mathrm{H} - \mathrm{P}^{(n-1)}\mathrm{H}\mathrm{P} \|_\mathrm{F}^2 \geq \| \mathrm{H} - \mathrm{H}\mathrm{P} \|_\mathrm{F}^2 > 0$. The identity $\| \mathbf{h} - \mathrm{P}^{(n)}\mathbf{h} \|_2^2 = \| \mathrm{H} - \mathrm{P}^{(n-1)}\mathrm{H}\mathrm{P} \|_\mathrm{F}^2$ follows from Kronecker product identity (P6) in Appendix A and the definition of the Frobenius matrix norm $\| \cdot \|_\mathrm{F}$. $\square$

The next result is immediately obvious from the previous corollary and shows that H can be used to test if $\mathbf{h}$ is factorable.

Corollary 2.2: There exists a $\mathbf{g} \in \mathbb{R}^m$ such that $\mathbf{h} = \mathbf{g}^{(n)}$ if and only if $\mathrm{rank}\mathrm{H} = 1$.

If $\mathrm{rank}\mathrm{H} > r$, then in general the lower bound in Theorem 2 is not achieved by the approximation $\mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h}$ except in the following special cases examined in Corollary 2.3.

Corollary 2.3: Partition H into $m \times m$ symmetric matrices $\mathrm{G}_1, \ldots, \mathrm{G}_{m^{n-2}}$ so that

$$\mathrm{H} = [\mathrm{G}_1, \ldots, \mathrm{G}_{m^{n-2}}]^T.$$

$\| \mathbf{h} - \mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h} \|_2^2 = \sum_{i=r+1}^{m} \sigma_i^2$, the lower bound in Theorem 2, if and only if $\mathrm{P}_{r,\mathrm{H}}$ and G_i commute for every $i = 1, \ldots, m^{n-2}$.

The proof of Corollary 2.3 involves some Kronecker product identities and is given in [15]. Notice that because the quadratic kernel is a symmetric matrix, Corollary 2.3 implies that in the quadratic case $\mathrm{P}_{r,\mathrm{H}}^{(2)}\mathbf{h}$ is always a best approximation. The special case of a quadratic filter was previously treated in [4, 11].

C. Discussion of Methods I and II

Method I (frequency domain design) only requires knowledge of the filter's support in the Fourier domain. In some applications, this prior information may be available without complete knowledge of the filter. Hence, in such cases, this approximation may be used prior to an identification experiment (see Example 1 in section VI). In general, Method I is suboptimal. In contrast, Method II (SVD design) requires complete knowledge of the filter. Method II also has the desirable characterization of near optimality in the sense of Theorem 2. It should be noted that Method II can be also applied in practice to initial kernel estimates obtained using other methods. This may improve the accuracy of the initial estimates by removing basis vectors corresponding to small singular values that may reflect errors in the estimate. Also notice that the use of such initial estimates obviates the need for "exact" knowledge of the filter.

The two filter error methods in this section are easily extended to a non-homogeneous nth order Volterra filter composed of n homogeneous filters (linear through nth order homogeneous). In terms of Method I (frequency domain design), the error bound given in Corollary 1 is extended by computing the error for each homogeneous component separately and using the sum of these bounds as a bound for the error for the complete non-homogeneous Volterra filter. Method 2 (SVD design) has an elegant generalization to the non-homogeneous case. Separately form the H matrix for each homogeneous kernel (*e.g.,* linear is a $1 \times m$ vector, nth order is an $m^{n-1} \times m$ matrix) and stack them to obtain a single $(\sum_{i=1}^{n} m^{i-1}) \times m$ matrix. The dominant right singular vectors of this matrix form a single basis for the complete nth order non-homogeneous Volterra filter.

IV. Input Error Designs

Define the norm of any $q \times 1$ real-valued random vector $\mathbf{Z}$, $q \geq 1$, as $\|\mathbf{Z}\| \triangleq \mathrm{tr}(\mathrm{E}[\mathbf{Z}\mathbf{Z}^T])^{1/2}$. Recall that the input error is defined as

$$e_i = \|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\| = \mathrm{tr}(\mathrm{E}[(\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)})(\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)})^T])^{1/2}. \tag{9}$$

The objective of this section is to find a rank r orthogonal projector P so that $\mathrm{P}\mathbf{X}$ is a good approximator of $\mathbf{X}$ in the sense of (9). Two suboptimal approaches are considered. The first approach utilizes the optimal mean-square rank r approximation of $\mathbf{X}$. That is, the rank r orthogonal projection matrix $\mathrm{P}_{r,\mathrm{R}}$ that minimizes $\|\mathbf{X}-\mathrm{P}\mathbf{X}\|$ over all orthogonal projection matrices P of rank $\leq r$ is computed to obtain the approximation $(\mathrm{P}_{r,\mathrm{R}}\mathbf{X})^{(n)}$ to $\mathbf{X}^{(n)}$. This method is particularly appropriate when $\mathbf{X}$ has a linear correlation structure (*i.e.*, $\mathbf{X}$ is a linear transformation of independent random variables). The second approach is based on the singular value decomposition and is closely related to Method II in the filter error section. The second design is also shown to be nearly optimal in the sense of (9).

A. Method III: Correlation Matrix Based Input Error Design

Theorem 3: Let P be an orthogonal projection matrix on $\mathbb{R}^m$. If $\mathbf{X}$ is an m-dimensional random vector with finite $2n$th order moments, then there exists a constant $0 \leq \alpha_n < \infty$ such that

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

and

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

Theorem 3, which is proved in Appendix D, suggests the choice of P that minimizes $\|\mathbf{X}-\mathrm{P}\mathbf{X}\|^2 = \mathrm{tr}(\mathrm{R}-\mathrm{PR}-\mathrm{RP}+\mathrm{PRP})$, where $\mathrm{R} \triangleq \mathrm{E}[\mathbf{X}\mathbf{X}^T]$ is the autocorrelation matrix of $\mathbf{X}$. Using the eigendecomposition $\mathrm{R}=\mathrm{UDU}^T$ and defining $\mathrm{C}=\mathrm{UD}^{1/2}\mathrm{U}^T$ write

$$\mathrm{tr}(\mathrm{R}-\mathrm{PR}-\mathrm{RP}+\mathrm{PRP}) = \mathrm{tr}((\mathrm{C}-\mathrm{CP})^T(\mathrm{C}-\mathrm{PC})) = \|\mathrm{C}-\mathrm{CP}\|_{\mathrm{F}}^2, \tag{10}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius matrix norm. It is easily established (using Theorem A1 in Appendix A) that a rank r orthogonal projection matrix minimizing (10) 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 3 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 2nd and $2n$th order moment of the each individual random variable in the vectors $\mathbf{X}$ 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 determined independent of $\mathrm{P}_{r,\mathrm{R}}$.

Theorem 4: 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

$$\mathrm{E}[U_j^{2n}] \;\leq\; \alpha_{n,F_j}\,\mathrm{E}[U_j^2]^n,\; j=1,\ldots,q. \tag{11}$$

The proof of Theorem 4 is also in Appendix D. Notice that under the assumptions of Theorem 4, the bounds in Theorem 3 are computed using only the second order moments of $\mathbf{X}$ and the bounds (11) relating the 2nd and $2n$th order moments of the independent $\mathbf{U}$ process. The next corollaries illustrate three important applications.

Corollary 4.1: If $\mathbf{X}$ is jointly Gaussian mean-zero, then

$$\alpha_n = \frac{(2n)!}{n!2^n}. \tag{12}$$

Proof: If $\mathbf{X}$ is jointly Gaussian mean-zero, then there exists a matrix C such that $\mathbf{X} = \mathrm{C}\mathbf{U}$, where $\mathbf{U}$ is a vector of independent zero-mean Gaussian r.v.s. For a zero-mean Gaussian distribution F, irrespective of the variance, it is well known that $\alpha_{n,F} = \frac{(2n)!}{n!2^n}$. □

Corollary 4.2: Let $\{X_k\}_{k\in\mathbb{Z}}$ be a stationary sinusoidal process

$$X_k = \sum_{j=1}^{q} c_j \cos(\omega_j k - \phi_j),$$

where $\{\phi_j\}_{j=1}^{q}$ are i.i.d. uniform on $[-\pi, \pi]$, $c_1, \ldots, c_q \in \mathbb{R}$ and $\omega_1, \ldots, \omega_q \in \mathbb{R}$. If $\mathbf{X} = (X_k, \ldots, X_{k-m+1})^T$, then

$$\alpha_n = \frac{2^n\,(2n-1)!!}{(2n)!!},$$

where $(2n-1)!! \triangleq 1\cdot 3 \cdots 2n-1$ and $(2n)!! \triangleq 2\cdot 4 \cdots 2n$.

The proof is straightforward and is found in [15].

Corollary 4.3: If $U_1, \ldots, U_q$ are independent, symmetric, uniformly distributed random variables, then

$$\alpha_n = \frac{3^n}{2n+1}.$$

Proof: If U_i is uniformly distributed on $[-b_i, b_i]$, where $b_i > 0$, then $\mathrm{E}[U_i^{2n}] = \frac{b_i^{2n}}{2n+1}$. Hence, $\mathrm{E}[U_i^{2n}] = \frac{3^n}{2n+1}\mathrm{E}[U_i^2]^n$. □

B. Method IV: SVD Based Input Error Design

This nearly optimal design method requires complete knowledge of the $2n$th order moments of $\mathbf{X}$ and does not make any assumptions regarding the correlation structure. The following theorem is proved in Appendix E. Recall that the vec operator applied to a matrix stacks the columns of the matrix into a vector.

Theorem 5: Let $\mathrm{R}_n = \mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]$ and let C_n be a matrix square root satisfying $\mathrm{C}_n^2 = \mathrm{R}_n$. Let C be an $m^{(2n-1)} \times m$ matrix of the m^{2n} elements in C_n appropriately ordered so that $\mathrm{vec}(\mathrm{C}) = \mathrm{vec}(\mathrm{C}_n)$. Let $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ denote the singular values of C and let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be the associated right singular vectors. Furthermore, for $r \leq m$, let $\wp_r$ denote the compact set of all $m \times m$ orthogonal projection matrices with rank $\leq r$ and let $\mathrm{P}_{r,\mathrm{C}} \in \wp_r$ be the orthogonal projector onto $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. Then

$$\sum_{i=r+1}^{m} \sigma_i^2 \;\leq\; \min_{\mathrm{Q}_1,\ldots,\mathrm{Q}_n \in \wp_r} \| \mathbf{X}^{(n)} - (\bigotimes_{i=1}^{n} \mathrm{Q}_i)\mathbf{X}^{(n)} \|^2 \;\leq\; \| \mathbf{X}^{(n)} - \mathrm{P}_{r,\mathrm{C}}^{(n)} \mathbf{X}^{(n)} \|^2 \;\leq\; n \sum_{i=r+1}^{m} \sigma_i^2.$$

The following corollary is analogous to Corollary 2.1 and can be proved using Corollary 2.1 and Theorem 5.

Corollary 5.1: There exists a rank r orthogonal projection P such that $\| \mathbf{X}^{(n)} - \mathrm{P}^{(n)}\mathbf{X}^{(n)} \|^2 = 0$ if and only if $\mathrm{rank}\mathrm{C} \leq r$. Moreover, if $\mathrm{rank}\mathrm{C} \leq r$, then $\| \mathbf{X}^{(n)} - \mathrm{P}_{r,\mathrm{C}}^{(n)}\mathbf{X}^{(n)} \|^2 = 0$.

A condition for the global optimality of the projector $\mathrm{P}_{r,\mathrm{C}}$, similar to Corollary 2.3, is also easily established and is not given here.

C. Discussion of Methods III and IV

Method III utilizes knowledge of the second order correlation of $\mathbf{X}$. The design method and error bound only involve second order moments of the $\mathbf{X}$ process, except for the bounding constant α_n. Under the assumption of linearity, α_n is determined using only the 2nd and $2n$th order moments of the underlying independent, symmetric process. In general, Method III is suboptimal.

Method IV requires the $2n$th order moments of $\mathbf{X}$ and does not make any linearity assumptions on $\mathbf{X}$ process. Also, Method IV is nearly optimal in the sense of Theorem 5. The $2n$th order moments are generally more difficult to compute or estimate than the second order correlations. Also, the design method involves computing the square root of an $m^n \times m^n$ matrix, requiring $\mathcal{O}(m^{3n})$ floating point operations, and hence is much more computationally intensive than Method III. However, note that the complexity of Method IV is similar to the complexity of the least squares identification of the original Volterra kernel $\mathbf{h}$.

V. 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 TPBA, the number multiplications required per output using the rank r^n TPBA $\hat{\mathbf{h}}$ 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 TPBA $\hat{\mathbf{h}}$ recall that the output is computed with a $\binom{n+r-1}{n}$ parameter Volterra filter $\mathbf{h}_{\mathrm{U}}$ and the transformed data vector $\mathbf{X}_{\mathrm{U}} = \mathrm{U}^T\mathbf{X}$, where the columns of U span an r-dimensional subspace $\mathcal{U} \subset \mathbb{R}^m$ (3). To form $\mathbf{X}_{\mathrm{U}}$ and all unique products in $\mathbf{X}_{\mathrm{U}}^{(n)}$ 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 TPBA.

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.

$$\eta_p = \frac{\#\mathrm{mults}(\hat{\mathbf{h}})}{\#\mathrm{mults}(\mathbf{h})} = \frac{rm + n\binom{n+r-1}{n}}{n\binom{n+m-1}{n}}, \tag{13}$$

and

$$\eta_s = \frac{\#\mathrm{mults}(\hat{\mathbf{h}})}{\#\mathrm{mults}(\mathbf{h})} = \frac{rm + n\binom{n+r-1}{n}}{\binom{n+m-1}{n} + (n-1)\binom{n+m-2}{n-1}}. \tag{14}$$

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 Stirling's formula $m! \sim \sqrt{2\pi}\, m^{m+1/2}\, e^{-m}$, it follows that

$$\frac{\binom{n+r-1}{n}}{\binom{n+m-1}{n}} \sim \rho^n, \qquad \frac{rm}{\binom{n+m-1}{n}} \sim n!\,\frac{\rho}{m^{n-2}}. \tag{15}$$

Hence,

$$\eta_p \;=\; \frac{\#\mathrm{mults}(\hat{\mathbf{h}})}{\#\mathrm{mults}(\mathbf{h})} \sim \frac{(n-1)!\,\rho}{m^{n-2}} + \rho^n, \tag{16}$$

and

$$\eta_s \;=\; \frac{\#\text{mults}(\hat{\mathbf{h}})}{\#\text{mults}(\mathbf{h})} \sim \frac{n!\,\rho}{m^{n-2}} + n\rho^n \;=\; n\,\eta_p. \tag{17}$$

The above expressions show how the reduction in complexity is related to the ratio of the approximation subspace dimension to m, $\rho \approx \frac{r}{m}$. In the special case of quadratic filters, further simplification is obtained by applying the method proposed in [4].

VI. Numerical Examples

Two examples are studied in this section. The first example demonstrates the filter error design methods applied to a simulated system identification problem. The second example studies the input error design methods for a Laplacian noise input.

A. Example 1 – Filter Error Design

In this example, the performance of the filter error design methods is studied. To accomplish this, the third order nonlinear system given in Fig. 1 is simulated. The system is a cascade of an FIR linear filter L, whose impulse response is depicted as the solid curve in Fig. 2, followed by memoryless, cubic polynomial p, represented by the curve in Fig. 3. The complete system is denoted as F. Cascade systems of this form are often called "Wiener" models [7]. The memory length of L is 40. The input x is i.i.d. uniform on $[-1, 1]$. This input is applied to the system and 2000 input and output samples are collected. The goal is to identify the "unknown" system F from the input and output data.

It is assumed that prior information is available that suggests:

1. The effective memory of the unknown system F is 40.
2. The response of F to sinusoidal inputs with frequency higher than 0.15 times the sampling frequency is negligible.
3. F displays nonlinear behavior up to third order.

Such information may be obtained by impulse and sinusoidal response tests prior to complete identification. In light of this prior information, Theorem 1 suggests that a low-frequency basis may be choosen for a TPBA. The basis is computed by finding the $12 = 40 \times .3$, (memory $\times$ bandwidth), eigenvectors associated with the 12 largest eigenvalues of the 40×40 positive semidefinite matrix

$$\mathrm{W} \triangleq \int_{[-.15,.15]} \mathbf{w}(f)\mathbf{w}^H(f)df, \tag{18}$$

where $\mathbf{w}(f) = (1, e^{i2\pi f}, \ldots, e^{i(39)2\pi f})^H$. Theorem 1 shows that by using this basis the TPBA represents the low-frequency response of F with negligible error. Since the high-frequency response of F is itself negligible, it is reasonable to expect that the TPBA will model F quite well. Using this basis, the third order TPBA (sum of linear, quadratic, and cubic homogeneous TPBA's) has 454 parameters. For comparison, the number of parameters in a third order Volterra filter with memory 40 is $12,340$. From the input and output data records the least squares estimate of the linear, quadratic, and cubic Volterra kernels using the TPBA are obtained. The normalized squared error between the true system kernels, denoted h_1, h_2, and h_3, and the TPBA kernel estimates, $\hat{h}_1$, $\hat{h}_2$, and $\hat{h}_3$, is defined as

$$e_k^2 = \frac{\sum_{i_1,\ldots,i_k=1}^{m} |h_k(i_1,\ldots,i_k) - \hat{h}_k(i_1,\ldots,i_k)|^2}{\sum_{i_1,\ldots,i_k=1}^{m} |h_k(i_1,\ldots,i_k)|^2}, \; k = 1, 2, 3. \tag{19}$$

For this simulation, the errors are $e_1^2 = 3.8751 \times 10^{-2}$, $e_2^2 = 1.6813 \times 10^{-1}$, and $e_3^2 = 1.7587 \times 10^{-1}$. The estimated and true kernels are also visually compared. The dashed curve in Fig. 2 shows the estimated linear kernel. Fig. 4 and 5 depict the true and estimated quadratic kernels respectively. Fig.s 6 and 7 show the 2-dimensional kernel "slices" $\{h_3(i,i,j)\}_{i,j=1}^{40}$ and $\{\hat{h}_3(i,i,j)\}_{i,j=1}^{40}$ of the third order kernels. These kernel slices are representative of the correspondence between the estimated and the true third order kernels.

If $\mathbf{g}$ is the impulse response vector of the linear system L and $\mathbf{x} = (x(k), \ldots, x(k-39))^T$, then the output of F is given by

$$\begin{aligned} z(k) &= 5(\mathbf{g}^T\mathbf{x})^3 - (\mathbf{g}^T\mathbf{x})^2 + \mathbf{g}^T\mathbf{x}, \\ &= 5(\mathbf{g}^{(3)})^T\mathbf{x}^{(3)} - (\mathbf{g}^{(2)})^T\mathbf{x}^{(2)} + \mathbf{g}^T\mathbf{x}. \end{aligned}$$

Written this way, it is easy to see that the vectorized second and third order kernels of F are proportional to $\mathbf{g} \otimes \mathbf{g}$ and $\mathbf{g} \otimes \mathbf{g} \otimes \mathbf{g}$ respectively. Using Theorem 2 and Corollary 2.2, $\mathbf{g}$ may be recovered exactly (up to a constant scale factor) from either the second or third order kernels. For example, in the third order case the $m^2 \times m$ matrix H, formed according to Theorem 2 using the third order kernel, is proportional to $(\mathbf{g} \otimes \mathbf{g})\mathbf{g}^T$. Hence, in this case H is rank 1 and the normalized right singular vector associated with the non-zero singular value is $\mathbf{g}/\|\mathbf{g}\|$. The second order kernel produces the same result. Hence, given only the true system kernels, using Theorem 2 one can deduce the cascade structure of F. The use of Theorem 2 to deduce the cascade structure from a general order Volterra kernel is an extension of the quadratic kernel rank criterion proposed in [7].

If the estimates of the Volterra kernels are sufficiently accurate, then applying Theorem 2 to the estimated kernels should reveal the special structure of the true system F. Using the estimates obtained from the system identification simulation above, an $m \times m$ matrix $\hat{\mathrm{H}}_2$ is formed from the estimate of the second order kernel $\hat{h}_2$, and an $m^2 \times m$ matrix $\hat{\mathrm{H}}_3$ is formed from the estimate of the third order kernel $\hat{h}_3$, both according to Theorem 2. Because the kernels are estimated using a 12 dimensional TPBA, $\hat{\mathrm{H}}_2$ and $\hat{\mathrm{H}}_3$ each have at most 12 non-zero singular values. The a plot of the first 12 singular values of $\hat{\mathrm{H}}_2$ is given in Fig. 8. The first 12 singular values of $\hat{\mathrm{H}}_3$ are plotted in Fig. 9. Note that both $\hat{\mathrm{H}}_2$ and $\hat{\mathrm{H}}_3$ are nearly rank 1 matrices indicating that both the second and third order kernels are well represented as a tensor product of a single basis vector. Furthermore, the right singular vectors corresponding to the single largest singular values of $\hat{\mathrm{H}}_2$ and $\hat{\mathrm{H}}_3$ are nearly the same. These singular vectors also match up well with the normalized estimate of the linear kernel as shown in Fig. 10. On the basis of this comparison, one may infer that the underlying true system is well represented by a cascade of a linear filter with impulse response $\hat{h}_1$ (linear kernel estimate) followed by a memoryless polynomial transformation.

B. Example 2 – Input Error Design

In this example, the input error design methods are examined. Let $\{U_k\}$ be an i.i.d. sequence of Laplace random variables, with density $f_{\mathrm{U}}(u) = e^{-2|u|}$. An MA sequence $\{X_k\}$ is generated by passing $\{U_k\}$ through a 10-tap FIR filter whose impulse response is shown in Fig. 11. Let $\mathbf{X} = \mathbf{X}_k = (X_k, \ldots, X_{k-9})^T$ be the input to a 2nd order homogeneous Volterra filter.

The eigenvalues of $\mathrm{R} = \mathrm{E}[\mathbf{X}\mathbf{X}^T]$, normalized by the largest and arranged in descending order, are depicted in the solid curve of Fig. 12. Note that the last 5 eigenvalues are approximately zero. Since $\mathbf{X}$ is a linear, symmetric process, Theorems 3 and 4 suggest that the first 5 eigenvectors of R (associated with the largest eigenvalues) provide an excellent basis for $\mathbf{X}$. If $\mathrm{P}_{5,\mathrm{R}}$ is the projection matrix corresponding to these five eigenvectors, then Theorems 3 and 4 produce the error bound (for Laplacian distributed random variables the bounding coefficient $\alpha_2 = 6$)

$$e_{\mathrm{R}} = \frac{\|\mathbf{X}^{(2)} - (\mathrm{P}_{5,\mathrm{R}}\mathbf{X})^{(2)}\|}{\|\mathbf{X}^{(2)}\|} \leq 1.1304 \times 10^{-2}. \tag{20}$$

The actual error in this case is $e_{\mathrm{R}} = 9.5386 \times 10^{-4}$. The bound of Theorem 3 overestimates the error by an order of magnitude, but is useful in that it indicates that the worst case error is approximately 1 percent.

A nearly optimal TPBA is obtained by forming the matrix C, as in Theorem 5, from the 4th order moment matrix $\mathrm{E}[\mathbf{X}^{(2)}\mathbf{X}^{(2)^T}]$. The singular values of C are shown (normalized and in decreasing order) in the dashed curve of Fig. 12. Notice that again only 5 singular values are significant. Using the five dominant right singular vectors of C as a basis and forming the corresponding projection matrix $P_{5,C}$, produces the error

$$e_C = \frac{\|\mathbf{X}^{(2)} - (P_{5,C}\mathbf{X})^{(2)}\|}{\|\mathbf{X}^{(2)}\|} = 9.5385 \times 10^{-4}. \tag{21}$$

Hence, in this case both the methods of Theorem 2 and Theorem 5 appear to perform equally well. In fact, the projections are nearly identical and $\|P_{5,R} - P_{5,C}\|_2 = 2.0458 \times 10^{-3}$ (note, for any projection matrix P, $\|P\|_2 = 1$).

Next, the input error design methods are examined for a nonlinear process. For this case, let $\{X_k\}$ be the quadratic process

$$X_k = 0.25\,U_k U_{k-1} \;+\; 0.5\,U_{k-1}U_{k-2} \;+\; 0.5\,U_{k-2}U_{k-3} \;+\; 0.25\,U_{k-3}U_{k-4}. \tag{22}$$

Again, $\mathbf{X} = \mathbf{X}_k = (X_k, \ldots, X_{k-9})^T$ is the input to an 2nd order homogeneous Volterra filter. The singular values of R and C for this case are depicted in the solid and dashed curves of Fig. 13 respectively. The rank 5 approximations of both methods in this case produce the errors

$$e_R = \frac{\|\mathbf{X}^{(2)} - (P_{5,R}\mathbf{X})^{(2)}\|}{\|\mathbf{X}^{(2)}\|} = 3.2090 \times 10^{-2}, \tag{23}$$

$$e_C = \frac{\|\mathbf{X}^{(2)} - (P_{5,C}\mathbf{X})^{(2)}\|}{\|\mathbf{X}^{(2)}\|} = 3.0432 \times 10^{-2}. \tag{24}$$

Notice that the nearly optimal SVD method does produce a slightly lower approximation error than Method III. As a point of interest, in this case the projections are quite different and $\|P_{5,R} - P_{5,C}\|_2 = 3.6018 \times 10^{-1}$.

In the two previous examples, the difference in performance between the two input error methods is slight. However in [16], it is shown that Method IV can perform arbitrarily better than Method III.

VII. Conclusions

The TPBA dramatically reduces the complexity of Volterra filters. Four methods for choosing the approximation basis for the TPBA are studied. The methods vary in computational complexity and required prior knowledge. Two methods are shown to be nearly optimal. In all cases, the approximation error of the TPBA is bounded to quantify the performance of the approximation. It is shown that the TPBA offers a much more efficient implementation than the original Volterra filter. Also, because certain design methods are based on incomplete prior knowledge of the filter (*i.e.,* frequency support) or input (*i.e.,* moments only) such approximations are also useful in reducing the estimation complexity of Volterra filters for identification and modelling problems. Furthermore, the approximation subspace provides useful insight into the response of the Volterra filter. In particular, the approximation subspace may be used to model or detect bandpass behavior and cascade structure as demonstrated in the examples.

Appendix A
Preliminaries

The following classical result regarding low-rank matrix approximations is used in several of the proofs.

Theorem A1 [5, 13]: For every complex-valued matrix $\mathrm{A} \in \mathbb{C}^{q \times m}$, $q \geq m$, there exists a matrix that is a best rank $r < m$ approximation to A, simultaneously with respect to every unitarily invariant norm $\|\cdot\|$ on $\mathbb{C}^{q\times m}$. Moreover, if $\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^T$ is the singular value decomposition of A where $\mathrm{U}^T\mathrm{U} = \mathrm{V}^T\mathrm{V} = \mathrm{I}$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$, $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$, then $\mathrm{A}_r = \mathrm{U}\Sigma_r\mathrm{V}^T$, where $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$, is a best rank $r < m$ approximation.

Corollary A1: If $\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^T$, V_r is an $m \times r$ matrix composed of the r columns of V corresponding to $\sigma_1, \ldots, \sigma_r$, and $\mathrm{P}_r = \mathrm{V}_r\mathrm{V}_r^T$, then $\mathrm{A}_r = \mathrm{A}\mathrm{P}_r$.

The Kronecker product also plays a key role in several of the proofs. The following Kronecker product properties are used extensively. If A is any matrix, then the n-fold Kronecker product of A with itself is denoted

$$\mathrm{A}^{(n)} = \underbrace{\mathrm{A} \otimes \cdots \otimes \mathrm{A}}_{n \text{ times}}.$$

Also, throughout the appendices, let I denote the $m \times m$ identity matrix. See [3] for a review of Kronecker product properties. Dimensions of matrices used in Kronecker product properties:

$\mathrm{A}(p \times q)$	$\mathrm{B}(s \times t)$	$\mathrm{G}(t \times u)$	$\mathrm{H}(p \times q)$
$\mathrm{Q}(q \times q)$	$\mathrm{P}(p \times p)$	$\mathrm{D}(q \times s)$	$\mathrm{R}(s \times t)$

(P1) $(\mathrm{A} \otimes \mathrm{B}) \otimes \mathrm{D} = \mathrm{A} \otimes (\mathrm{B} \otimes \mathrm{D})$

(P2) $(\mathrm{A} + \mathrm{H}) \otimes (\mathrm{B} + \mathrm{R}) = \mathrm{A} \otimes \mathrm{B} + \mathrm{A} \otimes \mathrm{R} + \mathrm{H} \otimes \mathrm{B} + \mathrm{H} \otimes \mathrm{R}$

(P3) $(\mathrm{A} \otimes \mathrm{B})(\mathrm{D} \otimes \mathrm{G}) = \mathrm{AD} \otimes \mathrm{BG}$

(P4) $(\mathrm{A} \otimes \mathrm{B})^T = \mathrm{A}^T \otimes \mathrm{B}^T$

(P5) If $\{\lambda_i\}_{i=1}^p$ are the eigenvalues of P and $\{\mu_j\}_{j=1}^q$ are the eigenvalues of Q, then the pq eigenvalues of $\mathrm{P} \otimes \mathrm{Q}$ are given by products $\{\lambda_i\mu_j\}_{i=1,\ldots,p,\, j=1,\ldots,q}$.

(P6) $\mathrm{vec}(\mathrm{ADB}) = (\mathrm{B}^T \otimes \mathrm{A})\, \mathrm{vec}(\mathrm{D})$

Appendix B

Method I: Frequency Domain Filter Error Design

Proof of Theorem 1: This theorem establishes a minimizer of

$$\min_{\mathrm{P}\in\wp_r} \|\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)}\|_2, \tag{25}$$

where $\wp_r$ is the set of all orthogonal projection matrices on $\mathbb{R}^m$ of rank $\leq r$. To show that $\wp_r$ is compact suppose that $\{\mathrm{Q}_j\}_{j\geq 1}$ is a convergent sequence in $\wp_r$. Then $\mathrm{Q}_j^2 = \mathrm{Q}_j$, $\mathrm{Q}_j^T = \mathrm{Q}_j$, and $\mathrm{rank}\mathrm{Q}_j \leq r$, $\forall j$, and hence $\wp_r$ is closed. Also, since each element in $\wp_r$ is a projection matrix, $\wp_r$ is bounded. Therefore, since $\wp_r$ is finite dimensional, closed, and bounded, it is compact. The error is continuous with respect to P and hence a minimizer exists.

Let $\mathrm{W} = \mathrm{UDU}^T$ be the eigendecomposition of W, where $\mathrm{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. If $\mathrm{C} = \mathrm{UD}^{1/2}\mathrm{U}^T$, then $\mathrm{W} = \mathrm{C}^T\mathrm{C}$. Notice that

$$\mathrm{W} + \mathrm{PWP} - \mathrm{PW} - \mathrm{WP} = (\mathrm{C} - \mathrm{CP})^T(\mathrm{C} - \mathrm{CP}),$$

so that

$$\|\mathrm{W} + \mathrm{PWP} - \mathrm{PW} - \mathrm{WP}\|_2 = \|\mathrm{C} - \mathrm{CP}\|_2^2.$$

Theorem A1 implies that $\mathrm{P}_{r,\mathrm{W}}$, as defined in the statement of Theorem 1, minimizes $\|\mathrm{C} - \mathrm{CP}\|_2$ and hence $\mathrm{P}_{r,\mathrm{W}}$ minimizes $\|\mathrm{W} + \mathrm{PWP} - \mathrm{PW} - \mathrm{WP}\|_2$. It is easily verified that $\|\mathrm{W} + \mathrm{P}_{r,\mathrm{W}}\mathrm{W}\mathrm{P}_{r,\mathrm{W}} - \mathrm{P}_{r,\mathrm{W}}\mathrm{W} - \mathrm{W}\mathrm{P}_{r,\mathrm{W}}\|_2 = \lambda_{r+1}$. Note that $\mathrm{P}_{r,\mathrm{W}}\mathrm{W}\mathrm{P}_{r,\mathrm{W}} = \mathrm{P}_{r,\mathrm{W}}\mathrm{W} = \mathrm{W}\mathrm{P}_{r,\mathrm{W}} = \mathrm{U}\mathrm{D}_r\mathrm{U}^T$, where $\mathrm{D}_r = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$. Kronecker product property (P3) implies that $\mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)} = (\mathrm{P}_{r,\mathrm{W}}\mathrm{W})^{(n)}$. Since $\mathrm{P}_{r,\mathrm{W}}\mathrm{W} = \mathrm{P}_{r,\mathrm{W}}\mathrm{W}\mathrm{P}_{r,\mathrm{W}}$, applying (P3) again shows that $\mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)} = \mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)}\mathrm{P}_{r,\mathrm{W}}^{(n)}$. Therefore,

$$\begin{aligned}
\|\mathrm{W}^{(n)} + \mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)}\mathrm{P}_{r,\mathrm{W}}^{(n)} - \mathrm{P}_{r,\mathrm{W}}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}_{r,\mathrm{W}}^{(n)}\|_2 &= \|\mathrm{W}^{(n)} - (\mathrm{W}\mathrm{P}_{r,\mathrm{W}})^{(n)}\|_2, \\
&= \|(\mathrm{UDU}^T)^{(n)} - (\mathrm{U}\mathrm{D}_r\mathrm{U}^T)^{(n)}\|_2, \\
&= \|\mathrm{U}^{(n)}(\mathrm{D}^{(n)} - \mathrm{D}_r^{(n)})\mathrm{U}^{(n)^T}\|_2.
\end{aligned}$$

The matrix $(\mathrm{D}^{(n)} - \mathrm{D}_r^{(n)})$ is diagonal and positive semidefinite. Furthermore, it is easily verified that the largest element of $(\mathrm{D}^{(n)} - \mathrm{D}_r^{(n)})$ is equal to $\lambda_1^{n-1}\lambda_{r+1}$. Therefore,

$$\|\mathrm{U}^{(n)}(\mathrm{D}^{(n)} - \mathrm{D}_r^{(n)})\mathrm{U}^{(n)^T}\|_2 = \lambda_1^{n-1}\lambda_{r+1}.$$

Hence, to prove the theorem it suffices to show that for every orthogonal projection matrix P with rank $\leq r$ there exists a unit norm vector $\mathbf{e}$ such that

$$\mathbf{e}^T(\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)})\mathbf{e} \ \geq\ \lambda_1^{n-1}\lambda_{r+1}.$$

Let $\mathbf{v}_{\mathrm{P}}$ maximize $\mathbf{v}^T\mathrm{W}\mathbf{v}$ subject to $\mathrm{P}\mathbf{v} = 0$, $\|\mathbf{v}\|_2 = 1$. Then it is easily established that $\mathbf{v}_{\mathrm{P}}^T\mathrm{W}\mathbf{v}_{\mathrm{P}} \geq \lambda_{r+1}$. To see this, note that the problem: maximize $\mathbf{v}^T\mathrm{W}\mathbf{v}$ subject to $\mathrm{P}\mathbf{v} = 0$, $\|\mathbf{v}\|_2 = 1$, is equivalent to: maximize $\mathbf{v}^T\mathrm{P}^\perp\mathrm{W}\mathrm{P}^\perp\mathbf{v}$ subject to $\|\mathbf{v}\|_2 = 1$, where $\mathrm{P}^\perp = \mathrm{I} - \mathrm{P}$. Also note that $\mathbf{v}_{\mathrm{P}}^T\mathrm{P}^\perp\mathrm{W}\mathrm{P}^\perp\mathbf{v}_{\mathrm{P}} = \|\mathrm{W} + \mathrm{PWP} - \mathrm{PW} - \mathrm{WP}\|_2$. Hence, if $\mathbf{v}_{\mathrm{P}}^T\mathrm{P}^\perp\mathrm{W}\mathrm{P}^\perp\mathbf{v}_{\mathrm{P}} < \lambda_{r+1}$, then $\|\mathrm{W} + \mathrm{PWP} - \mathrm{PW} - \mathrm{WP}\|_2 < \lambda_{r+1}$. However, this contradicts the optimality of $\mathrm{P}_{r,\mathrm{W}}$ according to Theorem A1.

Let $\mathbf{u}_1$ denote the unit norm eigenvector of W associated with λ_1 and set $\mathbf{e} = \mathbf{u}_1^{(n-1)} \otimes \mathbf{v}_{\mathrm{P}}$. Using (P3) it follows that $\mathbf{e}^T\mathbf{e} = 1$ and

$$\mathrm{P}^{(n)}\mathbf{e} = (\mathrm{P}\mathbf{u}_1)^{(n-1)} \otimes (\mathrm{P}\mathbf{v}_{\mathrm{P}}) = (\mathrm{P}\mathbf{u}_1)^{(n-1)} \otimes \mathbf{0} = \mathbf{0}.$$

Hence,

$$\begin{aligned}
\mathbf{e}^T(\mathrm{W}^{(n)} + \mathrm{P}^{(n)}\mathrm{W}^{(n)}\mathrm{P}^{(n)} - \mathrm{P}^{(n)}\mathrm{W}^{(n)} - \mathrm{W}^{(n)}\mathrm{P}^{(n)})\mathbf{e} &= \mathbf{e}^T\mathrm{W}^{(n)}\mathbf{e}, \\
&= (\mathbf{u}_1^T\mathrm{W}\mathbf{u}_1)^{n-1}(\mathbf{v}_{\mathrm{P}}^T\mathrm{W}\mathbf{v}_{\mathrm{P}}), \quad \text{by (P3)}, \\
&= \lambda_1^{n-1}(\mathbf{v}_{\mathrm{P}}^T\mathrm{W}\mathbf{v}_{\mathrm{P}}), \\
&\geq \lambda_1^{n-1}\lambda_{r+1}.
\end{aligned}$$

□

Appendix C

Method II: SVD Based Filter Error Design

Proof of Theorem 2: First show that $\| \mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h} - \mathbf{h} \|_2^2 \leq n\sum_{i=r+1}^m \sigma_i^2$. Let

$$e_k = \| \mathbf{h} - (\mathrm{P}_{r,\mathrm{H}}^{(k)} \otimes \mathrm{I}^{(n-k)})\mathbf{h} \|_2^2, \; k = 1, \ldots, n.$$

Note that $(\mathrm{P}_{r,\mathrm{H}}^{(k)} \otimes \mathrm{I}^{(n-k)})$ is an orthogonal projection matrix and use (P3) to establish the identity

$$(\mathrm{P}_{r,\mathrm{H}}^{(k)} \otimes \mathrm{I}^{(n-k)}) = (\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})(\mathrm{I}^{(k-1)} \otimes (\mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})).$$

Using the identity above

$$\begin{aligned}
e_k &= \| \mathbf{h} - (\mathrm{P}_{r,\mathrm{H}}^{(k)} \otimes \mathrm{I}^{(n-k)})\mathbf{h} \|_2^2, \\
&= \| (\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})[\mathbf{h} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h}] \\
&\quad +(\mathrm{I}^{(n)} - \mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})\mathbf{h} \|_2^2.
\end{aligned}$$

Since $(\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})$ and $(\mathrm{I}^{(n)} - \mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})$ are projectors onto orthogonal subspaces

$$e_k = \| (\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})[\mathbf{h} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h}] \|_2^2$$

$$+\,\|\,(\mathrm{I}^{(n)} - \mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})\mathbf{h}\,\|_2^2,$$

$$= \quad \|\,(\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})[\mathbf{h} \,-\, (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h}]\,\|_2^2 \,+\, e_{k-1}. \tag{26}$$

Now, since $(\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})$ is a projection matrix,

$$\|\,(\mathrm{P}_{r,\mathrm{H}}^{(k-1)} \otimes \mathrm{I}^{(n-k+1)})[\mathbf{h} \,-\, (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h}]\,\|_2^2 \leq \|\,\mathbf{h} \,-\, (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h}\,\|_2^2. \tag{27}$$

Furthermore, the symmetry of $\mathbf{h}$ implies that

$$\|\,\mathbf{h} \,-\, (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h}\,\|_2^2 = \|\,\mathbf{h} \,-\, (\mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-1)})\mathbf{h}\,\|_2^2 = e_1. \tag{28}$$

To see this, let $\mathrm{P}_{r,\mathrm{H}}^{\perp} = \mathrm{I} - \mathrm{P}_{r,\mathrm{H}}$. Then using (P2)

$$\mathbf{h} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-k)})\mathbf{h} = (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}}^{\perp} \otimes \mathrm{I}^{(n-k)})\mathbf{h}.$$

Let $\mathbf{u}_i$, $i = 1, \ldots, m$ denote columns of I and let $\mathbf{p}_i$, $i = 1, \ldots, m$ denote the columns of $\mathrm{P}_{r,\mathrm{H}}^{\perp}$. Then each element of $(\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}}^{\perp} \otimes \mathrm{I}^{(n-k)})\mathbf{h}$ has the form $\mathbf{h}^T(\mathbf{u}_{i_1} \otimes \cdots \otimes \mathbf{u}_{i_k} \otimes \mathbf{p}_j \otimes \mathbf{u}_{i_{k+1}} \otimes \cdots \otimes \mathbf{u}_{i_{n-1}})$, for appropriate integers $i_1, \ldots, i_{n-1}, j$. The symmetry of $\mathbf{h}$ implies that for every collection of m-vectors $\{\mathbf{x}_i\}_{i=1}^n$ and every permutation $(\pi(1), \ldots, \pi(n))$ of $(1, \ldots, n)$,

$$\mathbf{h}^T(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_n) = \mathbf{h}^T(\mathbf{x}_{\pi(1)} \otimes \cdots \otimes \mathbf{x}_{\pi(n)}).$$

Hence,

$$\mathbf{h}^T(\mathbf{u}_{i_1} \otimes \cdots \otimes \mathbf{u}_{i_k} \otimes \mathbf{p}_j \otimes \mathbf{u}_{i_{k+1}} \otimes \cdots \otimes \mathbf{u}_{i_{n-1}}) = \mathbf{h}^T(\mathbf{u}_{i_1} \otimes \cdots \otimes \mathbf{u}_{i_{n-1}} \otimes \mathbf{p}_j).$$

For each $i_1, \ldots, i_{n-1}, j$ the term on the right hand side of the equation above is an element of the vector $(\mathrm{P}_{r,\mathrm{H}}^{\perp} \otimes \mathrm{I}^{(n-1)})\mathbf{h}$. Hence, the vectors $(\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{H}}^{\perp} \otimes \mathrm{I}^{(n-k)})\mathbf{h}$ and $(\mathrm{P}_{r,\mathrm{H}}^{\perp} \otimes \mathrm{I}^{(n-1)})\mathbf{h}$ have the same elements and thus both have the same norm.

From (26), (27), and (28) it is easily established that $e_k \leq e_{k-1} + e_1$ and $\|\,\mathbf{h} - \mathrm{P}_{r,\mathrm{H}}^{(n)}\mathbf{h}\,\|_2^2 = e_n \leq n\,e_1$. Notice that $\mathrm{vec}(\mathrm{H}) = \mathbf{h}$, where the vec operator stacks the matrix columns to form a column vector. The Kronecker product vec identity (P6) shows that $\mathrm{vec}(\mathrm{H} - \mathrm{H}\mathrm{P}_{r,\mathrm{H}}) = \mathbf{h} - (\mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-1)})\mathbf{h}$. The Frobenius norm, denoted $\|\cdot\|_{\mathrm{F}}$, is the square root of the sum of the square of every element in the argument. Hence, $\|\,\mathbf{h} - (\mathrm{P}_{r,\mathrm{H}} \otimes \mathrm{I}^{(n-1)})\mathbf{h}\,\|_2^2 = \|\mathrm{H} - \mathrm{H}\mathrm{P}_{r,\mathrm{H}}\|_{\mathrm{F}}^2$. The Frobenius norm is unitarily invariant and therefore Theorem A1 and Corollary A1 imply that $\mathrm{H}\mathrm{P}_{r,\mathrm{H}}$ is a best rank r approximation to H and $e_1 = \|\mathrm{H} - \mathrm{H}\mathrm{P}_{r,\mathrm{H}}\|_{\mathrm{F}}^2 = \sum_{i=r+1}^{m} \sigma_i^2$.

To establish the lower bound consider the following. Suppose that

$$\min_{\mathrm{Q}_1, \ldots, \mathrm{Q}_n \in \wp} \|\,(\bigotimes_{i=1}^{n} \mathrm{Q}_i)\mathbf{h} \,-\, \mathbf{h}\,\|_2^2 = \|(\bigotimes_{i=1}^{n} \bar{\mathrm{Q}}_i)\mathbf{h} \,-\, \mathbf{h}\,\|_2^2,$$

that is, $\{\bar{\mathrm{Q}}_i\}_{i=1}^n$ are minimizers. Note that

$$\begin{aligned}\|(\bigotimes_{i=1}^{n}\bar{\mathrm{Q}}_i)\mathbf{h} - \mathbf{h}\|_2^2 &= \|(\mathrm{I}^{(n)} - \bigotimes_{i=1}^{n}\bar{\mathrm{Q}}_i)\mathbf{h}\|_2^2, \\ &\geq \|(\mathrm{I}^{(n)} - \bar{\mathrm{Q}}_1 \otimes \mathrm{I}^{(n-1)})\mathbf{h}\|_2^2,\end{aligned}$$

since the subspace spanned by the columns of $\bar{\mathrm{Q}}_1 \otimes \mathrm{I}^{(n-1)}$ contains the subspace spanned by the columns of $\bigotimes_{i=1}^n \bar{\mathrm{Q}}_i$. Using (P6), the Frobenius norm, and Corollary A1,

$$\begin{aligned}\|(\mathrm{I}^{(n)} - \bar{\mathrm{Q}}_1 \otimes \mathrm{I}^{(n-1)})\mathbf{h}\|_2^2 &= \|\mathrm{H}\bar{\mathrm{Q}}_1 - \mathrm{H}\|_{\mathrm{F}}^2, \\ &\geq \|\mathrm{H}\mathrm{P}_{r,\mathrm{H}} - \mathrm{H}\|_{\mathrm{F}}^2 = \sum_{i=r+1}^{m} \sigma_i^2.\end{aligned}$$

□

Appendix D
Method III: Correlation Matrix Based Input Error Design

Proof of Theorem 3: Theorem 3 states that if P is an orthogonal projection matrix and $\mathbf{X}$ is an m-dimensional random vector with finite $2n$th order moments, then there exists a constant $0 \leq \alpha_n < \infty$ such that

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

Define $e_k = \mathbf{X}^{(n)} - (\mathrm{I}^{(n-k)} \otimes \mathrm{P}^{(k)})\mathbf{X}^{(n)}$, for $k = 1, \ldots, n$. Then using Kronecker properties (P2) and (P3)

$$\|e_{k+1}\|^2 = \|\mathbf{X}^{(n)} - (\mathrm{I}^{(n-k-1)} \otimes \mathrm{P}^{(k+1)})\mathbf{X}^{(n)}\|^2 = \|\mathrm{Q}_k\bar{e}_{1,k+1} + \mathrm{Q}_k^{\perp}\mathbf{X}^{(n)}\|^2, \tag{29}$$

where $\mathrm{Q}_k = \mathrm{I}^{(n-k)} \otimes \mathrm{P}^{(k)}$, $\mathrm{Q}_k^{\perp} = \mathrm{I}^{(n)} - \mathrm{Q}_k$, and $\bar{e}_{1,k+1} = \mathbf{X}^{(n)} - (\mathrm{I}^{(n-k-1)} \otimes \mathrm{P} \otimes \mathrm{I}^{(k)})\mathbf{X}^{(n)}$. Since Q_k and $\mathrm{Q}_k^{\perp}$ are projectors onto orthogonal subspaces, it follows that

$$\begin{aligned}\|\mathrm{Q}_k\bar{e}_{1,k+1} + \mathrm{Q}_k^{\perp}\mathbf{X}^{(n)}\|^2 &= \mathrm{tr}(\mathrm{E}[(\mathrm{Q}_k\bar{e}_{1,k+1} + \mathrm{Q}_k^{\perp}\mathbf{X}^{(n)})(\mathrm{Q}_k\bar{e}_{1,k+1} + \mathrm{Q}_k^{\perp}\mathbf{X}^{(n)})^T]), \\ &= \mathrm{tr}(\mathrm{Q}_k\mathrm{E}[\bar{e}_{1,k+1}\bar{e}_{1,k+1}^T]\mathrm{Q}_k) + \mathrm{tr}(\mathrm{Q}_k^{\perp}\mathrm{E}[\mathbf{X}^{(n)}(\mathbf{X}^{(n)})^T]\mathrm{Q}_k^{\perp}),\end{aligned}$$

where the facts

$$\mathrm{tr}(\mathrm{Q}_k\mathrm{E}[\bar{e}_{1,k+1}(\mathbf{X}^{(n)})^T]\mathrm{Q}_k^{\perp}) = \mathrm{tr}(\mathrm{E}[\bar{e}_{1,k+1}(\mathbf{X}^{(n)})^T]\mathrm{Q}_k^{\perp}\mathrm{Q}_k) = 0$$

and similarly $\mathrm{tr}(\mathrm{Q}_k^{\perp}\mathrm{E}[\mathbf{X}^{(n)}\bar{e}_{1,k+1}^T]\mathrm{Q}_k) = 0$ are used. Also, since Q_k is a projection matrix, $\mathrm{tr}(\mathrm{Q}_k\mathrm{E}[\bar{e}_{1,k+1}\bar{e}_{1,k+1}^T]\mathrm{Q}_k) \leq \mathrm{tr}(\mathrm{E}[\bar{e}_{1,k+1}\bar{e}_{1,k+1}^T])$. By symmetry $\mathrm{tr}(\mathrm{E}[\bar{e}_{1,k+1}\bar{e}_{1,k+1}^T]) = \|\bar{e}_{1,k+1}\|^2 = \|e_1\|^2$. To see this let $\mathrm{P}^{\perp} = \mathrm{I} - \mathrm{P}$.

Then using Kronecker property (P2) $\bar{e}_{1,k+1} = \mathbf{X}^{(n-k-1)} \otimes \mathrm{P}^{\perp}\mathbf{X} \otimes \mathbf{X}^{(k)}$ and properties (P3) and (P4) show that

$$\|\bar{e}_{1,k+1}\|^2 = \mathrm{tr}(\mathrm{E}[\bar{e}_{1,k+1}\bar{e}_{1,k+1}^T]) = \mathrm{tr}(\mathrm{E}[(\mathbf{X}\mathbf{X}^T)^{(n-k-1)} \otimes \mathrm{P}^{\perp}\mathbf{X}\mathbf{X}^T\mathrm{P}^{\perp} \otimes (\mathbf{X}\mathbf{X}^T)^{(k)}]). \tag{30}$$

The trace is equal to the sum of the eigenvalues and by the Kronecker product eigenvalue property (P5) the ordering of the Kronecker products does not effect the eigenvalues, hence

$$\begin{aligned}
&\mathrm{tr}(\mathrm{E}[(\mathbf{X}\mathbf{X}^T)^{(n-k-1)} \otimes \mathrm{P}^{\perp}\mathbf{X}\mathbf{X}^T\mathrm{P}^{\perp} \otimes (\mathbf{X}\mathbf{X}^T)^{(k)}]) \\
&\qquad = \mathrm{E}[\mathrm{tr}((\mathbf{X}\mathbf{X}^T)^{(n-k-1)} \otimes \mathrm{P}^{\perp}\mathbf{X}\mathbf{X}^T\mathrm{P}^{\perp} \otimes (\mathbf{X}\mathbf{X}^T)^{(k)})], \\
&\qquad = \mathrm{E}[\mathrm{tr}((\mathbf{X}\mathbf{X}^T)^{(n-1)} \otimes \mathrm{P}^{\perp}\mathbf{X}\mathbf{X}^T\mathrm{P}^{\perp})], \\
&\qquad = \|e_1\|^2.
\end{aligned}$$

Finally, note that $\mathrm{Q}_k^{\perp}\mathbf{X}^{(n)} = e_k$ and therefore $\|e_{k+1}\|^2 \leq \|e_1\|^2 + \|e_k\|^2$ and $\|e_n\|^2 \leq n\,\|e_1\|^2$. Hence,

$$\begin{aligned}
\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2 &= \|e_n\|^2 \leq n\,\|e_1\|^2 = n\,\|\mathbf{X}^{(n)} - (\mathrm{I}^{(n-1)} \otimes \mathrm{P})\mathbf{X}^{(n)}\|^2, \\
&= n\,\|\mathbf{X}^{(n-1)} \otimes (\mathbf{X} - \mathrm{P}\mathbf{X})\|^2, \quad \text{by (P2).}
\end{aligned}$$

Now let $\mathbf{X}_1, \ldots, \mathbf{X}_{n-1} = \mathbf{X}$ and let $\mathbf{X}_n = (\mathbf{X} - \mathrm{P}\mathbf{X})$ and let $X_{i,j}$ denote the jth element of the vector $\mathbf{X}_i$. Since $\mathbf{X}$ has finite $2n$th order moments there exists a constant $0 \leq \alpha_n < \infty$ such that $\mathrm{E}[X_{i,j}^{2n}] \leq \alpha_n\, \mathrm{E}[X_{i,j}^2]^n$. Therefore,

$$\begin{aligned}
\|\mathbf{X}^{(n-1)} \otimes (\mathbf{X} - \mathrm{P}\mathbf{X})\|^2 &= \|\mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_n\|^2, \\
&= \sum_{i_1,\ldots,i_n=1}^{m} \mathrm{E}[X_{1,i_1}^2 \cdots X_{1,i_n}^2], \\
&\leq \sum_{i_1,\ldots,i_n=1}^{m} \prod_{j=1}^{n} \mathrm{E}[X_{j,i_j}^{2n}]^{1/n}, \text{ by Hölder's inequality,} \\
&\leq \sum_{i_1,\ldots,i_n=1}^{m} \prod_{j=1}^{n} \alpha_n^{1/n}\mathrm{E}[X_{j,i_j}^2] = \alpha_n \sum_{i_1,\ldots,i_n=1}^{m} \prod_{j=1}^{n} \mathrm{E}[X_{j,i_j}^2], \\
&= \alpha_n \prod_{j=1}^{n} \|\mathbf{X}_j\|^2 = \alpha_n\,\|\mathbf{X}\|^{2(n-1)}\|\mathbf{X} - \mathrm{P}\mathbf{X}\|^2.
\end{aligned}$$

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

The ratio $\frac{\|\mathbf{X}^{(n)} - (\mathrm{P}\mathbf{X})^{(n)}\|^2}{\|\mathbf{X}^{(n)}\|^2}$ quantifies the quality of the approximation $(\mathrm{P}\mathbf{X})^{(n)}$. The numerator is bounded from above as using the previous argument. The denominator $\|\mathbf{X}^{(n)}\|^2$ is bounded from below using Jensen's inequality.

$$\|\mathbf{X}^{(n)}\|^2 = \mathrm{tr}(\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]),$$

$$
\begin{aligned}
&= \operatorname{tr}(\mathrm{E}[(\mathbf{X}\mathbf{X}^T)^{(n)}]), \\
&= \mathrm{E}[\operatorname{tr}(\mathbf{X}\mathbf{X}^T)^n], \\
&= \mathrm{E}[(\mathbf{X}^T\mathbf{X})^n], \\
&\geq (\mathrm{E}[(\mathbf{X}^T\mathbf{X})])^n, \text{ by Jensen's inequality,} \\
&= \|\mathbf{X}\|^{2n}.
\end{aligned}
$$

□

Proof of Theorem 4: Theorem 4 asserts that if $\mathbf{X}$ is a linear transformation of an independent, symmetric process $\mathbf{U}$, then the bounding constant α_n can be determined from the 2nd and $2n$th order moments of $\mathbf{U}$ alone. Consider a random variable

$$
Z = \sum_{j=1}^{q} c_j \, U_j, \tag{31}
$$

where $c_1, \ldots, c_q$ are real numbers. Technically, this proof only requires that the odd order moments of U_j, $j = 1, \ldots, q$ vanish up to order $2n-1$, and the symmetry assumption guarantees this. First consider the second order moment of Z. Due to the independence of the U_j's and the fact that they are zero-mean,

$$
\mathrm{E}[Z^2] \;=\; \sum_{j_1, j_2 = 1}^{q} c_{j_1} c_{j_2} \mathrm{E}[U_{j_1} U_{j_2}] \;=\; \sum_{j=1}^{q} c_j^2 \mathrm{E}[U_j^2]. \tag{32}
$$

Next look at the $2n$th order moment of Z,

$$
\mathrm{E}[Z^{2n}] = \sum_{j_1, \ldots, j_{2n} = 1}^{q} c_{j_1} \cdots c_{j_{2n}} \mathrm{E}[U_{j_1} \cdots U_{j_{2n}}]. \tag{33}
$$

Note that since the distributions are symmetric and by independence, $\mathrm{E}[U_{j_1} \cdots U_{j_{2n}}]$ is zero if the product inside is factorable in the form $U_j^{2p_j+1} \prod_{i \neq j} U_i^{p_i}$, for some j and non-negative integers $p_1, \ldots, p_q$. Hence, any non-zero term in the series above (33) has the form,

$$
c_1^{2p_1} \cdots c_q^{2p_q} \mathrm{E}[U_1^{2p_1} \cdots U_q^{2p_q}], \tag{34}
$$

where $p_1, \ldots, p_q$ is an arbitrary collection of non-negative integers whose sum is n. In fact, the series (33) contains all such terms; that is every product of the form (34) over all $p_1, \ldots, p_q \geq 0$, $\sum_{j=1}^{q} p_j = n$. Also notice that each term of the form (34) may be bounded as

$$
\begin{aligned}
c_1^{2p_1} \cdots c_q^{2p_q} \mathrm{E}[U_1^{2p_1} \cdots U_q^{2p_q}] &\leq c_1^{2p_1} \cdots c_q^{2p_q} \, \mathrm{E}[U_1^{2n}]^{\frac{p_1}{n}} \cdots \mathrm{E}[U_q^{2n}]^{\frac{p_q}{n}}, \text{ by Hölder's inequality,} \\
&\leq c_1^{2p_1} \cdots c_q^{2p_q} \, \alpha_n^{p_1/n} \, \mathrm{E}[U_1^2]^{p_1} \cdots \alpha_n^{p_q/n} \, \mathrm{E}[U_q^2]^{p_q}, \\
&= \alpha_n \, c_1^{2p_1} \cdots c_q^{2p_q} \, \mathrm{E}[U_1^2]^{p_1} \cdots \mathrm{E}[U_q^2]^{p_q},
\end{aligned} \tag{35}
$$

where the fact $\sum_{j=1}^{n} p_j = n$ is used in the last step. Hence,

$$\mathrm{E}[Z^{2n}] \;\leq\; \alpha_n \sum_{p_1,\ldots,p_q \geq 0 : \sum p_j = n} c_1^{2p_1} \cdots c_q^{2p_q} \mathrm{E}[U_1^2]^{p_1} \cdots \mathrm{E}[U_q^2]^{p_q}. \tag{36}$$

Finally note that $\mathrm{E}[Z^2]^n = (\sum_j c_j^2 \mathrm{E}[U_j^2])^n$ is equivalent to the sum above. Hence, $\mathrm{E}[Z^{2n}] \leq \alpha_n \, \mathrm{E}[Z^2]^n$.

If C is an arbitrary $m \times q$ matrix and P is a projection matrix on $\mathbb{R}^m$, then the elements of $\mathbf{X} = \mathrm{C}\mathbf{U}$ and $(\mathrm{I} - \mathrm{P})\mathbf{X}$ are simply linear combinations of the $\mathbf{U}$ process. Hence α_n as defined may be used in Theorem 3 (see proof of Theorem 3). □

Appendix E

Method IV: SVD Based Input Error Design

Proof of Theorem 5: This theorem is proved using an approach similar to the proof of Theorem 2. Three identities are established first. Let $\mathrm{R}_n = \mathrm{E}[\mathbf{X}^{(n)} \mathbf{X}^{(n)^T}]$, C_n be the square root of R_n (*i.e.*, $\mathrm{R}_n = \mathrm{C}_n^2$), and let $\{\mathrm{Q}_i\}_{i=1}^{n}$ be an arbitrary collection of orthogonal projection matrices. Because each Q_i is an orthogonal projection matrix, so is the Kronecker product $(\bigotimes_{i=1}^{n} \mathrm{Q}_i)$. It then follows that the input error may be written as

$$\begin{aligned}
\| \mathbf{X}^{(n)} - (\bigotimes_{i=1}^{n} \mathrm{Q}_i) \mathbf{X}^{(n)} \|^2 &= \mathrm{tr}([\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i)]^T [\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i)]), \\
&= \|\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i)\|_{\mathrm{F}}^2, \qquad (37)
\end{aligned}$$

where $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius norm. It is also easy to show that

$$\mathrm{tr}(\mathrm{R}_n - (\bigotimes_{i=1}^{n} \mathrm{Q}_i) \mathrm{R}_n) = \|\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i)\|_{\mathrm{F}}^2. \tag{38}$$

Using the Kronecker product vec identity (P6) $\mathrm{vec}(\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i)) = \mathbf{c} - ((\bigotimes_{i=1}^{n} \mathrm{Q}_i) \otimes \mathrm{I}^{(n)})\mathbf{c}$, where $\mathbf{c} = \mathrm{vec}(\mathrm{C}_n)$. The Frobenius norm is the square root of the sum of square of every element in the argument. Hence,

$$\begin{aligned}
\|\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i)\|_{\mathrm{F}}^2 &= \|\mathrm{vec}(\mathrm{C}_n - \mathrm{C}_n(\bigotimes_{i=1}^{n} \mathrm{Q}_i))\|_{\mathrm{F}}^2 = \|\mathbf{c} - ((\bigotimes_{i=1}^{n} \mathrm{Q}_i) \otimes \mathrm{I}^{(n)})\mathbf{c}\|_{\mathrm{F}}^2, \\
&= \|\mathbf{c} - ((\bigotimes_{i=1}^{n} \mathrm{Q}_i) \otimes \mathrm{I}^{(n)})\mathbf{c}\|_2^2. \qquad (39)
\end{aligned}$$

The three identities (37),(38), and (39) are used several times in the subsequent analysis. Now to establish the upper bound $\| \mathbf{X}^{(n)} - \mathrm{P}_{r,\mathrm{C}}^{(n)} \mathbf{X}^{(n)} \|^2 \leq n \sum_{i=r+1}^{m} \sigma_i^2$ or equivalently using identities

(37) and (39),

$$\|\mathbf{c} - (\mathrm{P}_{r,\mathrm{C}}^{(n)} \otimes \mathrm{I}^{(n)})\mathbf{c}\|_2^2 \;\leq\; n \sum_{i=r+1}^{m} \sigma_i^2,$$

define $e_k = \|\mathbf{c} - (\mathrm{P}_{r,\mathrm{C}}^{(k)} \otimes \mathrm{I}^{(2n-k)})\mathbf{c}\|_2^2$. Proceeding in the same manner as in the proof of Theorem 2 (see equation (26) in Appendix C) shows that

$$e_k = \|(\mathrm{P}_{r,\mathrm{C}}^{(k-1)} \otimes \mathrm{I}^{(2n-k+1)})[\mathbf{c} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(2n-k)})\mathbf{c}]\|_2^2 + e_{k-1}. \tag{40}$$

Also, since $(\mathrm{P}_{r,\mathrm{C}}^{(k-1)} \otimes \mathrm{I}^{(2n-k+1)})$ is a orthogonal projection matrix

$$\|(\mathrm{P}_{r,\mathrm{C}}^{(k-1)} \otimes \mathrm{I}^{(2n-k+1)})[\mathbf{c} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(2n-k)})\mathbf{c}]\|_2^2 \leq \|\mathbf{c} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(2n-k)})\mathbf{c}\|_2^2. \tag{41}$$

At this point, to establish the upper bound it suffices to show that

$$\|\mathbf{c} - (\mathrm{I}^{(k-1)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(2n-k)})\mathbf{c}]\|_2^2 \;\;=\;\; e_1.$$

To see this, use (P3) to establish

$$\begin{aligned}
\mathrm{tr}(\mathrm{R}_n - (\mathrm{I}^{(n-k)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(k-1)})\mathrm{R}_n) \;\;&=\;\; \mathrm{tr}(\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}] \\
&\qquad -(\mathrm{I}^{(n-k)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(k-1)})\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}]), \\
&=\;\; \mathrm{tr}(\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}] \\
&\qquad -\mathrm{E}[(\mathbf{X}^{(n-k)} \otimes \mathrm{P}_{r,\mathrm{C}}\mathbf{X} \otimes \mathbf{X}^{(k-1)})\mathbf{X}^{(n)^T}]), \\
&=\;\; \mathrm{E}[\mathrm{tr}(\mathbf{X}^{(n)}\mathbf{X}^{(n)^T})] \\
&\qquad -\mathrm{E}[\mathrm{tr}((\mathbf{X}^{(n-k)} \otimes \mathrm{P}_{r,\mathrm{C}}\mathbf{X} \otimes \mathbf{X}^{(k-1)})\mathbf{X}^{(n)^T})].
\end{aligned}$$

Because the trace of a matrix is the sum of the eigenvalues and since (P5) implies that the eigenvalues of $(\mathbf{X}^{(n-k)} \otimes \mathrm{P}_{r,\mathrm{C}}\mathbf{X} \otimes \mathbf{X}^{(k-1)})\mathbf{X}^{(n)^T}$ are the same as the eigenvalues of $(\mathbf{X}^{(n-1)} \otimes \mathrm{P}_{r,\mathrm{C}}\mathbf{X})\mathbf{X}^{(n)^T}$. Therefore,

$$\begin{aligned}
\mathrm{tr}(\mathrm{R}_n - (\mathrm{I}^{(n-k)} \otimes \mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(k-1)})\mathrm{R}_n) \;\;&=\;\; \mathrm{E}[\mathrm{tr}(\mathbf{X}^{(n)}\mathbf{X}^{(n)^T})] \\
&\qquad -\mathrm{E}[\mathrm{tr}((\mathbf{X}^{(n-1)} \otimes \mathrm{P}_{r,\mathrm{C}}\mathbf{X})\mathbf{X}^{(n)^T})], \\
&=\;\; \mathrm{tr}(\mathrm{E}[\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}] \\
&\qquad -\mathrm{E}[(\mathrm{P}_{r,\mathrm{C}}\mathbf{X} \otimes \mathbf{X}^{(n-1)})\mathbf{X}^{(n)^T}]), \\
&=\;\; \mathrm{tr}(\mathrm{R}_n - (\mathrm{P}_{r,\mathrm{C}} \otimes \mathrm{I}^{(n-1)})\mathrm{R}_n).
\end{aligned} \tag{42}$$

Now using the identities (38), (39), and (42),

$$\begin{aligned}
\|\mathbf{c}-(\mathrm{I}^{(k-1)}\otimes \mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(2n-k)})\mathbf{c}\|_2^2 &= \|\mathrm{C}_n-(\mathrm{I}^{(n-k)}\otimes \mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(k-1)})\mathrm{C}_n\|_{\mathrm{F}}^2,\\
&= \mathrm{tr}(\mathrm{R}_n-(\mathrm{I}^{(n-k)}\otimes \mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(k-1)})\mathrm{R}_n),\\
&= \mathrm{tr}(\mathrm{R}_n-(\mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(n-1)})\mathrm{R}_n),\\
&= \|\mathrm{C}_n-(\mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(n-1)})\mathrm{C}_n\|_{\mathrm{F}}^2,\\
&= \|\mathbf{c}-(\mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(2n-1)}))\mathbf{c}\|_2^2 = e_1.
\end{aligned}$$

Hence, combining this with (40) and (41) produces the bound $e_n \le n\, e_1$. Kronecker property (P6) shows that $\mathrm{vec}(\mathrm{CP}_{r,\mathrm{C}}) = (\mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(2n-1)})\mathbf{c}$. Hence, using the Frobenius norm $\|\,\mathbf{c}-(\mathrm{P}_{r,\mathrm{C}}\otimes \mathrm{I}^{(2n-1)})\mathbf{c}\,\|_2^2 = \|\mathrm{C}-\mathrm{CP}_{r,\mathrm{C}}\|_{\mathrm{F}}^2$. The Frobenius norm is unitarily invariant and therefore Theorem A1 and Corollary A1 imply that $\mathrm{CP}_{r,\mathrm{C}}$ is a best rank r approximation to C and $e_1 = \|\mathrm{C}-\mathrm{CP}_{r,\mathrm{C}}\|_{\mathrm{F}}^2 = \sum_{i=r+1}^{m}\sigma_i^2$.

The lower bound is easily established as follows. Suppose that

$$\min_{\mathrm{Q}_1,\ldots,\mathrm{Q}_n\in\wp_r}\|\,\mathbf{X}^{(n)} - (\bigotimes_{i=1}^{n}\mathrm{Q}_i)\mathbf{X}^{(n)}\,\|^2 = \|\,\mathbf{X}^{(n)} - (\bigotimes_{i=1}^{n}\bar{\mathrm{Q}}_i)\mathbf{X}^{(n)}\,\|^2,$$

that is $\{\bar{\mathrm{Q}}_i\}_{i=1}^n$ are minimizers. Note that

$$\|\,\mathbf{X}^{(n)} - (\bigotimes_{i=1}^{n}\bar{\mathrm{Q}}_i)\mathbf{X}^{(n)}\,\|^2 = \|(\mathrm{I}^{(n)}-\bigotimes_{i=1}^{n}\bar{\mathrm{Q}}_i)\mathbf{X}^{(n)}\|^2 \ge \|(\mathrm{I}^{(n)}-\bar{\mathrm{Q}}_1\otimes \mathrm{I}^{(n-1)})\mathbf{X}^{(n)}\|^2,$$

since the subspace spanned by the columns of $\bar{\mathrm{Q}}_1\otimes \mathrm{I}^{(n-1)}$ contains the subspace spanned by the columns of $\bigotimes_{i=1}^{n}\bar{\mathrm{Q}}_i$. Using (38), (39), and Corollary A1,

$$\|(\mathrm{I}^{(n)}-\bar{\mathrm{Q}}_1\otimes \mathrm{I}^{(n-1)})\mathbf{X}^{(n)}\|^2 = \|\mathrm{C}\bar{\mathrm{Q}}_1-\mathrm{C}\|_{\mathrm{F}}^2 \ge \|\mathrm{CP}_{r,\mathrm{H}}-\mathrm{C}\|_{\mathrm{F}}^2 = \sum_{i=r+1}^{m}\sigma_i^2. \tag{43}$$

□

References

[1] S. A. Alshebeili, B. G. Mertzios, and A. N. Venetsanopoulos, "Efficient implementation of nonlinear Volterra systems via Laguerre orthogonal functions," in *Signal Processing VI: Theories and Applications*, J. Vanderwalle, R. Boite, M. Moonen, A. Oosterlinck (eds.), Elsevier Science Publishers, 1992.

[2] P. Bondon, "Polynomial estimation of the amplitude of a signal," *IEEE Trans. Info. Theory*, vol. 40, no. 3, pp. 960-965, May 1994.

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

[4] 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.

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

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

[7] R. Haber, "Structure identification of block-oriented models based on the estimated Volterra kernels," *Int. J. Syst. Sci.*, vol. 20, pp. 1355-1380, 1989.

[8] T. Koh and E. J. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. 33, no. 6. pp. 1445-1455, Dec. 1995.

[9] M. J. Korenberg, "Orthogonal approaches to time-series analysis and system identification,"*IEEE SP Magazine*, vol. 8, no. 3, pp. 29-43, 1991.

[10] V. Z. Marmarelis, "Identification of nonlinear biological systems using Laguerre expansions of kernels,"*Ann. Biomed. Engr.*, vol. 21, pp. 573-589, 1993.

[11] V. Z. Marmarelis, "Nonlinear modelling of physiological systems using principal dynamic modes," in *Advanced Methods of Physiological System Modelling*, Volume 3, pp. 1-27, Plenum Press, New York, 1994.

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

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

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

[15] R. Nowak and B. Van Veen, "Volterra filter approximatations," University of Wisconsin-Madison ECE Technical Report ECE-94-16.

[16] R. Nowak and B. Van Veen, "Volterra filter approximatations," *Proceedings of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, vol. 2, pp. 951-954, Neos Marmaras, Halkidiki, Greece, 1995.

[17] B. Picinbono and P. Duvaut, "Optimal linear-quadratic systems for detection and estima-

tion," *IEEE Trans. Inform. Theory*, vol. 34, pp. 304-311, 1988.

[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] J. D. Taft, "Quadratic-linear filters for signal detection," *IEEE Trans. Signal Proc.*, vol. 39, no. 11, pp. 2557-2559, 1991.

[20] N. Wiener, *Nonlinear Problems in Random Theory*, Wiley and Sons, New York, 1958.