

IDENTIFICATION OF INFINITE DEGREE VOLTERRA SERIES IN THE TIME AND FREQUENCY DOMAINS

Yufeng Wan, Tony J. Dodd, Robert F. Harrison

*Department of Automatic Control and Systems
Engineering, The University of Sheffield, UK*

Abstract: Volterra series (VS) are widely used in non-linear dynamical system identification. Much physical information about a system can be extracted from the corresponding VS model. Most non-linear frequency domain representations have been based on VS models through the application of the Fourier transform. But the fact that the number of the parameters to be identified in a VS model increases exponentially with the size of the VS model restricts its application. The involvement of kernel methods has been shown significantly to reduce the burden of the computation, with the potential to increase the practical usability of VS and the methods that are based on it. This paper presents the identification of infinite degree, finite memory length, time-invariant, discrete VS from the general reproducing kernel Hilbert space point of view and introduces its extension to the estimation of generalized frequency response functions. *Copyright © 2005 IFAC*

Keywords: Volterra Series, nonlinear systems, dynamics, intelligent machines, Hilbert spaces, frequency responses

1. INTRODUCTION

Volterra series (VS) models can be regarded as a generalization of Taylor series. They are widely used in solving non-linear dynamical system identification problems. They can represent a wide range of non-linear dynamical systems in terms of their individual polynomial components (Boyd and Chua, 1985). Without the involvement of autoregressive terms, such models can provide prediction with consistent precision over the domain of interest for “fading memory” type systems (Boyd and Chua, 1985), provided that measured inputs are accurate. They are also the basis of many non-linear frequency domain representations (Peyton-Jones and Billings, 1989). Because of these and some other practical advantages, such as the linear-in-parameters property (Schetzen, 1980), much has been done in the

development and application of VS models. At the same time, the fact that the number of parameters to be determined increases exponentially with the size of the VS models restricts their practical application.

Kernel methods have been shown to be of help in solving this problem (Harrison, 1999; Drezet, 2001; Dodd and Harrison, 2002a; Dodd and Harrison, 2002b; Wan *et al.*, 2003; Franz and Schölkopf, 2004). Previously, a specially constructed exponential kernel has been involved to estimate a general infinite degree, finite memory length¹ VS in generalized Fock space (Zyla and De Figueiredo, 1983; Wan *et al.*, 2003). Since our focus is on the

¹ The use of a finite memory length method implies that the dynamics under consideration must be of the class of “fading memory” systems (Boyd and Chua, 1985).

discrete time situation, the Fock space framework is overly general. This paper solves the problem from a simpler mathematical point of view, while presenting a detailed exploration of its properties. The extension of the idea to the estimation of generalized frequency response functions (GFRFs) directly, in the frequency domain, is also provided.

The construction of the reproducing kernel Hilbert space (RKHS) containing infinite degree, finite memory length VS, and the corresponding exponential kernel are discussed in the next section. In Section 3 the method of computing the least squares (LS) approximation of the target VS using the exponential kernel, constructed in Section 2, is discussed. The recovery of the estimated VS from the kernel model is introduced in Section 4. A new method for estimating the corresponding GFRFs is discussed in Section 5, and, in Section 6 a synthetic system identification example is analyzed. Finally, the characteristics of the method are summarized.

2. ESTIMATION IN RKHS

Consider a time-invariant, discrete-time, finite memory length, infinite degree VS

$$y(\underline{u}) = h_0 + \sum_{n=1}^{\infty} \left\{ \sum_{m_1=0}^{M-1} \cdots \sum_{m_n=0}^{M-1} h_n(m_1, \dots, m_n) \prod_{j=1}^n u_{m_j} \right\} \quad (1)$$

where $\underline{u} \triangleq [u_0, \dots, u_{M-1}]^T$ is the vector of lagged input samples, and M is the memory length.

A sufficient, but not necessary, condition that guarantees the convergence of Eq.(1) is that (Schetzen, 1980):

$$\sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} \cdots \sum_{m_n=0}^{M-1} |h_n(m_1, m_2, \dots, m_n)| < \infty \quad (2)$$

This series can be written in a compact form

$$y(\underline{u}) = \underline{h}^T \underline{\phi}(\underline{u}) \quad (3)$$

in which \underline{h} is a coefficient vector consisting of all the Volterra coefficients in Eq.(1) and $\underline{\phi}(\underline{u})$ is a vector of all possible polynomial combinations of the inputs, $u_i, i = 0, \dots, M-1$ with the first element 1. Furthermore, we can express Eq.(3) in vector form for a finite sample of N input-output pairs as

$$\underline{y} = \underline{h}^T \Phi \quad (4)$$

where $\underline{y} = [y(\underline{u}_1), y(\underline{u}_2), \dots, y(\underline{u}_N)]$ and $\Phi = [\underline{\phi}(\underline{u}_1), \underline{\phi}(\underline{u}_2), \dots, \underline{\phi}(\underline{u}_N)]$.

It can be shown that the space \mathcal{H} , spanned by $\underline{\phi}(\cdot)$, is a RKHS, equipped with a positive definite reproducing kernel k .

Firstly, given two functions f, g in \mathcal{H} , we can define the corresponding inner product as

$$\begin{aligned} \langle f, g \rangle_{\mathcal{H}} &= \langle (\underline{h}^f)^T \underline{\phi}^f, (\underline{h}^g)^T \underline{\phi}^g \rangle_{\mathcal{H}} \\ &= \sum_{l=1}^{\infty} \frac{h_{l(n, m_1, \dots, m_n)}^f h_{l(n, m_1, \dots, m_n)}^g}{\lambda_l} \end{aligned} \quad (5)$$

where $\underline{h}^f, \underline{\phi}^f$ are the coefficient and polynomial vectors of the function f , respectively, and $\underline{h}^g, \underline{\phi}^g$ are defined in the same way. Additionally, the role of λ_l will be revealed shortly but for now can be considered as a sequence of positive numbers.

Then the kernel k is defined as

$$k(\underline{u}_i, \underline{u}_j) = \sum_{l=1}^{\infty} \lambda_l \phi_l(\underline{u}_i) \phi_l(\underline{u}_j) = \langle \underline{\varphi}(\underline{u}_i), \underline{\varphi}(\underline{u}_j) \rangle_{l_2} \quad (6)$$

where $\underline{\varphi}(\cdot)$ is a vector with l th element $\varphi_l(\cdot) = \sqrt{\lambda_l} \phi_l(\cdot)$.

We can now prove k is a reproducing kernel in \mathcal{H} in two steps:

- (1) For a fixed input sequence \underline{u} , given Eq.(6), a new vector \underline{h}^k can be defined with l th element $h_l^k = \lambda_l \phi_l(\underline{u}), l = 1, 2, \dots, \infty$, then

$$\begin{aligned} k(\underline{u}, \cdot) &= \sum_{l=1}^{\infty} \lambda_l \phi_l(\underline{u}) \phi_l(\cdot) = \sum_{l=1}^{\infty} h_l^k \phi_l(\cdot) \\ &= (\underline{h}^k)^T \underline{\phi}(\cdot) \end{aligned} \quad (7)$$

which implies $k(\underline{u}, \cdot) \in \mathcal{H}$.

- (2) According to the inner product definition (5), for every f in \mathcal{H} ,

$$\begin{aligned} \langle k(\underline{u}, \cdot), f(\cdot) \rangle_{\mathcal{H}} &= \langle (\underline{h}^k)^T \underline{\phi}(\cdot), (\underline{h}^f)^T \underline{\phi}(\cdot) \rangle_{\mathcal{H}} \\ &= \sum_{l=1}^{\infty} \frac{h_l^k h_l^f}{\lambda_l} = \sum_{l=1}^{\infty} \frac{\lambda_l \phi_l(\underline{u}) h_l^f}{\lambda_l} \\ &= \langle \phi(\underline{u}), \underline{h}^f \rangle_{l_2} = (\underline{h}^f)^T \underline{\phi}(\underline{u}) \\ &= f(\underline{u}) \end{aligned} \quad (8)$$

Eq.(8) demonstrates the reproducing property of the kernel, $\langle k(\underline{u}, \cdot), f(\cdot) \rangle_{\mathcal{H}} = f(\underline{u})$.

A function k that has such characteristics is called the ‘‘reproducing kernel’’ (Aronszajn, 1950) of the space \mathcal{H} . The space \mathcal{H} , equipped with this k , is known as a RKHS. The properties of k and \mathcal{H} are summarized in the Moore-Aronszajn theorem (Wahba, 1990).

Given the kernel $k(\underline{u}, \cdot)$, every function f in \mathcal{H} , which corresponds to a particular VS with the form of Eq.(1), can be expressed as:

$$f(\cdot) = \sum_i \alpha_i k(\underline{u}_i, \cdot) \quad (9)$$

in which $\alpha_i \in \mathbb{R}$.

So far, the representation of infinite degree, finite memory length VS in terms of k has been presented. However, since the evaluation of k involves an infinite number of ϕ_l , it is computationally

infeasible. By using a specially constructed kernel, the computation of ϕ_l can be mapped to a finite dimensional space, making computation feasible.

Consider the exponential kernel

$$k_e(\underline{u}, \underline{v}) = \exp\left(\frac{\langle \underline{u}, \underline{v} \rangle_{l_2}}{p}\right)$$

(Zyla and De Figueiredo, 1983; Wan *et al.*, 2003), where $p \in \mathbb{R}^+$ is a constant adjusting the coefficients in the following expansion sequence

$$\begin{aligned} k_e(\underline{u}, \underline{v}) &= \exp\left(\frac{\langle \underline{u}, \underline{v} \rangle_{l_2}}{p}\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\left(\sum_{i=0}^{M-1} u_i v_i\right)^n}{p^n} \\ &= 1 + \frac{1}{p} \sum_{i=0}^{M-1} u_i v_i + \frac{1}{2!p^2} \left(\sum_{i=0}^{M-1} u_i v_i\right)^2 + \dots \\ &\quad + \frac{1}{h!p^h} \left(\sum_{i=0}^{M-1} u_i v_i\right)^h + \dots \end{aligned} \quad (10)$$

which corresponds to an infinite dimensional space spanned by $\left(\sum_{i=0}^{M-1} u_i v_i\right)^h$ with $h = 0, 1, \dots, \infty$, consisting of all possible polynomial terms. Therefore, we have

$$\exp\left(\frac{\langle \underline{u}, \underline{v} \rangle_{l_2}}{p}\right) = \sum_{l=1}^{\infty} \lambda_l \phi_l(\underline{u}) \phi_l(\underline{v}) \quad (11)$$

Hence, it can be shown that Eq.(9) is equivalent to

$$f(\cdot) = \sum_i \alpha_i k_e(\underline{u}_i, \cdot) \quad (12)$$

3. LS APPROXIMATION

Given a sample of size N , the representation of Eq.(12) becomes approximate but finite. The LS approximation \hat{y} of a VS f with the form (1) in a RKHS \mathcal{H}_N , spanned by $k_e(\underline{u}_1, \cdot), \dots, k_e(\underline{u}_N, \cdot)$, is given by the projection of y in \mathcal{H}_N ,

$$\hat{y}(\underline{u}) = \sum_{i=1}^N \alpha_i k_e(\underline{u}_i, \underline{u}), \quad (13)$$

Express Eq.(13) in vector form,

$$\hat{\underline{y}} = K \underline{\alpha} \quad (14)$$

where

$$\begin{aligned} \hat{\underline{y}} &= [\hat{y}(\underline{u}_1), \dots, \hat{y}(\underline{u}_N)]^T \\ \underline{\alpha} &= [\alpha_1, \dots, \alpha_N]^T \end{aligned}$$

and K is the exponential kernel Gram matrix,

$$K_{ij} = k_e(\underline{u}_i, \underline{u}_j) = \exp\left(\frac{\langle \underline{u}_i, \underline{u}_j \rangle_{l_2}}{p}\right) \quad (15)$$

Given that $\underline{u}_1, \dots, \underline{u}_N$ are distinct, the kernel Gram matrix, K , is nonsingular (Zyla and De Figueiredo, 1983), providing a unique solution.

$$\underline{\alpha} = K^{-1} \underline{y} \quad (16)$$

3.1 Numerical consideration

Although the above statement guarantees a unique solution to Eq.(16) in principle, the fundamental condition that $\underline{u}_i, i = 1, 2, \dots, N$, are distinct elements in \mathbb{R}^M is not easily achieved numerically in practice, even in the noise-free situation, especially when N becomes large. In a fixed hypervolume of input space, as N increases, the sample data will tend to be less well separated, and eventually some will be too close to be numerically distinguishable. Thus, at least two rows/columns of K will be approximately co-linear, leading to an ill-conditioned K . However, increasing the memory length M makes K less likely to be ill-conditioned. This will be discussed in more detail in the following section.

In numerical computation, Tikhonov (ridge or weight decay) regularization (with parameter, ρ) may be introduced to solve Eq.(16), therefore, $\tilde{\underline{\alpha}} = (K + \rho I)^{-1} \underline{y}$. Since our ultimate purpose is system identification in a noisy environment, we adopt regularization and accept the biased solution $\tilde{\underline{\alpha}}$.

4. RECOVERY OF ORIGINAL VS

VS models are widely used in analyzing non-linear systems because they contain important characteristics of the physical systems and are qualitatively well-behaved, like linear, finite impulse response models. An infinite degree VS can, in principle, represent the corresponding fading memory, non-linear system to arbitrary accuracy, so the recovery of the VS model may be of importance in many cases.

The equation to calculate the original Volterra kernels (coefficients) $\hat{h}_n(m_1, \dots, m_n)$, corresponding to the LS approximation $\hat{y}(\underline{u})$ in Eq.(13), has been given in (Wan *et al.*, 2003),

$$\begin{aligned} \hat{h}_n(m_1, \dots, m_n) &= \frac{1}{n!p^n} \sum_{i=1}^N \alpha_i \prod_{j=1}^n u_{im_j} \\ &\quad (n = 0, 1, \dots, \infty) \end{aligned}$$

which can be rewritten as

$$\hat{h}_{l(n, m_1, \dots, m_n)} = \frac{1}{n!p^n} \sum_{i=1}^N \alpha_i \phi_{l(n, m_1, \dots, m_n)} \quad (17)$$

Returning to Eq.(12) and (13): if \hat{y} represents the VS accurately, the coefficients, corresponding to \hat{y} , are an accurate representation of the coefficients of the original series because the VS representation of a non-linear system is unique, if it exists.

A general infinite degree VS, f , which could be represented in the form of Eq.(12), has an infinite

number of independent coefficients. But a VS, \hat{y} , represented in the form of Eq.(13), has only N independent Volterra coefficients because every $k_e(\cdot, \cdot)$ has a fixed set of coefficients, corresponding to the polynomial terms that it contains, i.e., the polynomial expansion of every $\alpha_i k_e(\cdot, \cdot)$ has only one independent coefficient.

A simple example follows to make the idea clearer. Assume two data samples, u_1 and u_2 , are available for estimation. Without loss of generality, assume the system is static (no memory). Then we have

$$\hat{y}(u) = \sum_{i=1}^N \alpha_i k_e(u_i, u) = \alpha_1 k_e(u_1, u) + \alpha_2 k_e(u_2, u) \quad (18)$$

Substituting Eq.(10) into Eq.(18) gives

$$\begin{aligned} \hat{y}(u) &= \alpha_1 \exp\left(\frac{u_1 u}{p}\right) + \alpha_2 \exp\left(\frac{u_2 u}{p}\right) \\ &= \alpha_1 + \alpha_1 \left(\frac{u_1}{p}\right) u + \alpha_1 \left(\frac{u_1^2}{2!p^2}\right) u^2 + \dots \\ &\quad + \alpha_2 + \alpha_2 \left(\frac{u_2}{p}\right) u + \alpha_2 \left(\frac{u_2^2}{2!p^2}\right) u^2 + \dots \end{aligned}$$

Since u_1 and u_2 are fixed, the number of independent coefficients in terms of $u^i, i = 0, 1, \dots, \infty$, in the series, is two, but the infinite degree VS should be of the form,

$$f(u) = \beta_0 + \beta_1 u + \beta_2 u^2 + \dots$$

where $\beta_0, \beta_1, \beta_2, \dots$ are, in general, independent. If we call the RKHS containing f , \mathcal{H}_∞ , the one containing \hat{y} , \mathcal{H}_2 in this case, then \hat{y} is the projection of f in \mathcal{H}_2 . Correspondingly, the Volterra coefficients, extracted from \hat{y} , are LS approximations of the original coefficients in \mathcal{H}_2 . As the number of distinct samples, N , increases, the accuracy with which \hat{y} represents f increases, and, in the limit, \hat{y} becomes f . Obviously, this is not computationally feasible, so a trade-off must be made between the accuracy of the approximation and the efficiency of the computation.

Another important parameter which affects the accuracy of the estimation is the memory length M of the VS. First, let's consider the effect of M on the numerical condition of K . As M increases, more elements are involved in \underline{u} , increasing the likelihood that the \underline{u}_i s are distinct (assuming the input is sufficiently exciting).

Second, in practice, increasing M reduces the accuracy of the whole algorithm. Consider the way we construct K as Eq.(15). Because no practical measurements are totally free of errors, the larger M is, the more elements are involved in a single input vector \underline{u}_i . Thus, more errors, coming from the measurements of the sample data, are involved in the computation of K_{ij} . From this step, what we use is a noisy kernel matrix K , leading to noisy

results in all the following steps. Consequently, performance deteriorates with M .

5. GENERALIZED FREQUENCY RESPONSE FUNCTIONS

The n th order GFRF is the n -fold Fourier transform of the n th order Volterra coefficients:

$$\begin{aligned} H_n(\omega_1, \dots, \omega_n) &= \sum_{m_1=0}^{M-1} \dots \sum_{m_n=0}^{M-1} h_n(m_1, \dots, m_n) \\ &\quad \times \exp\left(-j \sum_{a=1}^n \omega_a m_a\right) \end{aligned} \quad (19)$$

where j is the imaginary unit.

Given $e^{jx} = \cos x + j \sin x$, Eq.(19) becomes

$$\begin{aligned} H_n(\omega_1, \dots, \omega_n) &= \sum_{m_1=0}^{M-1} \dots \sum_{m_n=0}^{M-1} h_n(m_1, \dots, m_n) \cos\left(\sum_{a=1}^n \omega_a m_a\right) \\ &\quad - j \sum_{m_1=0}^{M-1} \dots \sum_{m_n=0}^{M-1} h_n(m_1, \dots, m_n) \sin\left(\sum_{a=1}^n \omega_a m_a\right) \end{aligned}$$

As discussed in the previous sections, by using the exponential kernel method, any Volterra coefficient h_n can be estimated, in principle. Then, through Eq.(19), H_n can be calculated.

Although H_n can be derived through h_n , all the errors in the estimated h_n are transferred to H_n , making the result relatively less reliable. So a method to compute H_n directly through a kernel method similar to that used in the time domain, is introduced in the next subsection.

5.1 Direct GFRF Computation

In the probing method (Bedrosian and Rice, 1971), the input signal

$$u(t) = \sum_{i=1}^R \exp(j\omega_i t) \quad (20)$$

is invoked to probe the system to reveal information in the frequency domain.

The output of a VS, given by Eq.(1), given the

input $u(t) = \sum_{i=1}^R \exp(j\omega_i t)$, is

$$\begin{aligned} y(t; \underline{H}, \omega) &= \sum_{n=1}^{\infty} \sum \frac{n!}{n_1! \dots n_c!} H_n(j\omega_1, \dots, j\omega_n) \\ &\quad \times \exp(j(\omega_1 + \dots + \omega_n)t) \end{aligned} \quad (21)$$

where the second summation is over all combinations of R frequencies taken n at a time, \underline{H} is the

column vector consisting of individual frequency responses², $\underline{\omega}$ is the column vector consisting of every frequency component in the input signal, c is the number of distinct frequencies in the combination, n_i is the number of repetitions of the i th distinct frequency, and $n_1 + n_2 + \dots + n_c = n$ (Peyton-Jones and Billings, 1989).

Eq.(21) can be rewritten as

$$y(t; \underline{H}, \underline{\omega}) = \underline{H}^T(\underline{\omega})\underline{\phi}(t; \underline{\omega}) \quad (22)$$

where $\underline{\phi}(t; \underline{\omega})$ is the column vector of all possible polynomial terms of the elements in $\underline{x}(t)$, $\underline{x}(t) = [x_1(t), \dots, x_N(t)]^T$, $x_a(t) = \exp(j\omega_a t)$, $a = 1, \dots, N$.

Given Eq.(20), $\underline{\omega}$ is a constant vector. Considering $\underline{x}(t)$ as an independent variable, we have

$$y(\underline{x}) = \underline{H}^T \underline{\phi}(\underline{x}) \quad (23)$$

Noting that the form of Eq.(23) is the same as that of Eq.(3), it is natural to apply the same approach to Eq.(23) to find the LS approximation $\underline{\hat{H}}$ of \underline{H} , in a RKHS \mathcal{H}_f , as we did for \hat{h} in Eq.(17).

Replacing \underline{u} and \underline{h} with \underline{x} and \underline{H} , respectively, we have the following, using the same derivations as in Sections 2, 3 and 4,

$$\hat{y}(\underline{x}) = \sum_{i=1}^{N_f} \tilde{\beta}_i k_e(\underline{x}_i, \underline{x}) \quad (24)$$

$$\tilde{\underline{\beta}} = (K + \rho_f I)^{-1} \underline{\hat{y}} \quad (25)$$

$$\hat{H}_{l(n, \omega_1, \dots, \omega_n)} = \frac{1}{n! p_f^n} \sum_{i=1}^{N_f} \tilde{\beta}_i \phi_{l(n, \omega_1, \dots, \omega_n)} \quad (26)$$

6. EXAMPLE

A synthetic Wiener-type model (Fig.1) is considered in this section. This is a finite memory length,

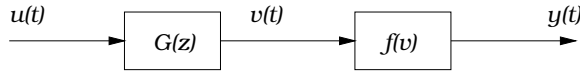


Fig. 1. Synthetic Wiener model.

infinite degree nonlinear dynamical system:

$$v(t) = 1.5u(t) + 2.5u(t-1) + 2u(t-2) \quad (27)$$

$$y(t) = \frac{1}{1 + e^{-v(t)}} \quad (28)$$

To reveal the even order Volterra terms, Eq.(28) is expanded on the point 1.2226, the mean value of $v(t)$, which makes us choose input signal $u(t)$ to be uniformly distributed on $[-0.8, 1.2]$. We have

$$y(t) = 0.7725 + 0.1757(v(t) - 1.2226) - 0.0479 \times (v(t) - 1.2226)^2 + \mathcal{O}(v(t)^3) \quad (29)$$

Substituting Eq.(27) into (29), it follows that

$$y(t) = 0.7725 + 0.1757 \left[1.5 \left(u(t) - \frac{1.2226}{6} \right) + 2.5 \left(u(t-1) - \frac{1.2226}{6} \right) + 2 \left(u(t-2) - \frac{1.2226}{6} \right) \right] - 0.0479 \left[1.5 \left(u(t) - \frac{1.2226}{6} \right) + 2.5 \left(u(t-1) - \frac{1.2226}{6} \right) + 2 \left(u(t-2) - \frac{1.2226}{6} \right) \right]^2 + \mathcal{O}(v(t)^3)$$

To excite the even order terms we bias $u(t)$ by $-\frac{1.2226}{6}$ giving the first and second order Volterra coefficients:

$$h_1 = [0.2636 \ 0.4392 \ 0.3514]$$

$$h_2 = - \begin{bmatrix} 0.1078 & 0.1796 & 0.1437 \\ 0.1796 & 0.2994 & 0.2395 \\ 0.1437 & 0.2395 & 0.1916 \end{bmatrix}$$

Assuming $M(= 3)$ is known, a kernel model is constructed to identify the system, in the noise free case, with the kernel constant $p = 0.1$ and the regularization $\rho = 1 \times 10^{-12}$. We generate 1998 + 2M input/output pairs and divide them for training and validation, respectively³.

The l_2 -norms of the first and second order Volterra coefficients estimation errors are 1.2583×10^{-4} and 7.6693×10^{-5} , respectively. Identification performance in the frequency domain is shown in Figs.2 and 3.

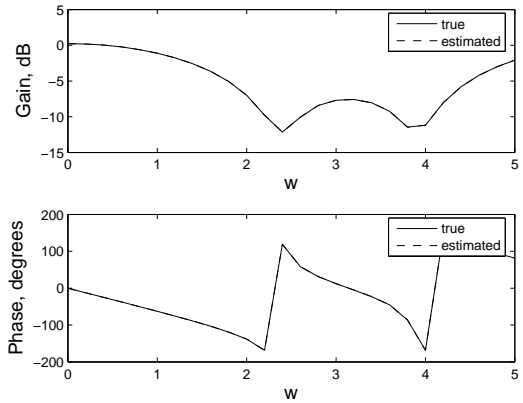


Fig. 2. Gain and phase of the first order GFRF. The true and estimated values are too close to distinguish.

The difference between the true and estimated GFRFs is too trivial to be identified by eyes. The maximum absolute errors in the gain and phase are 0.07 dB and 0.7 degrees, respectively.

² Assume the $H_n(j\omega_1, \dots, j\omega_n)$ is symmetric.

³ Half of the data are used for training and the remainder for testing, ensuring 1000 samples regardless of M .

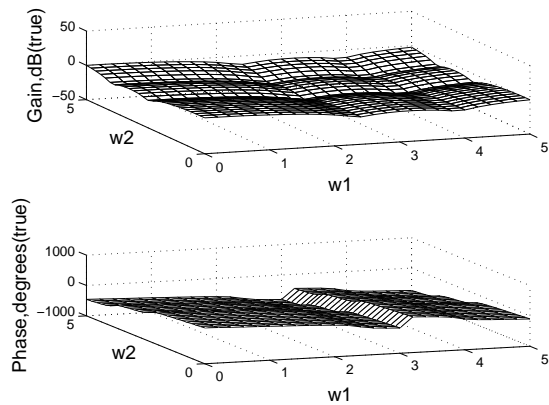


Fig. 3. True gain and phase of the second order GFRF.

7. CONCLUSION

By using the exponential kernel method we can, in principle, estimate any system that can be represented by time-invariant, infinite degree, finite memory length VS to arbitrary accuracy. The accuracy of the model output depends on sample size, memory length and noise level. The Volterra coefficients of the estimated system can also be recovered from the kernel model. By extending the method into the frequency domain, GFRFs can be estimated directly via a kernel method.

Although the number of sample data available for estimation restricts the precision of the approximation, the kernel method significantly reduces the computational burden in estimating “large” VS models, hence, expanding the application of VS in both time and frequency domains.

The application of our estimation method is underway in radio frequency power amplifier distortion cancelation and friction process identification. Early results have shown its promise in these non-linear dynamical system identification problems, which are difficult to address using conventional formulations.

ACKNOWLEDGEMENT

YW would like to thank University of Sheffield for its financial support.

REFERENCES

- Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society* **68**, 337–404.
- Bedrosian, E. and S. O. Rice (1971). The output properties of volterra systems (nonlinear systems with memory) driven by harmonic and gaussian inputs. *Proceedings of the IEEE* **59**, 1688–1707.
- Boyd, S. and L. O. Chua (1985). Fading memory and the problem of approximating non-linear operators with volterra series. *IEEE Transactions on Circuits and Systems* **Cas-32**, 1150–1161.
- Dodd, T. J. and R. F. Harrison (2002a). A new solution to Volterra series estimation. In: *CD-ROM Proceedings of the 2002 IFAC World Congress*. Barcelona.
- Dodd, T. J. and R. F. Harrison (2002b). Estimating Volterra filters in Hilbert space. In: *Proceedings of IFAC International Conference on Intelligent Control Systems and Signal Processing*. University of Algarve, Faro. pp. 538–543.
- Drezet, P. M. L. (2001). Kernel Methods and Their Application to Systems Identification and Signal Processing. Phd thesis. The University of Sheffield. Sheffield.
- Franz, Matthias O. and Bernhard Schölkopf (2004). Implicit estimation of Wiener series. In: *Machine Learning for Signal Processing XIV*. Sao Luis.
- Harrison, R. F. (1999). Computable Volterra filters of arbitrary degree. MAE Technical Report 3060. Princeton University. Princeton.
- Peyton-Jones, J. C. and S. A. Billings (1989). Recursive algorithm for computing the frequency response of a class of non-linear difference equation models. *International Journal of Control* **50**, 1925–1940.
- Schetzen, M. (1980). *The Volterra and Wiener Theories of Non-linear Systems*. New York: Wiley.
- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics. Philadelphia.
- Wan, Y., T. J. Dodd and R. F. Harrison (2003). Infinite degree Volterra series estimation. In: *CD-ROM Proceedings of the Second International Conference on Computational Intelligence, Robotics and Autonomous Systems*. Singapore.
- Zyla, L. V. and R. J. P. De Figueiredo (1983). Non-linear system identification based on a Fock space framework. *SIAM Journal on Control and Optimization* **21**, 931–939.