Gaussian process regression for the estimation of generalized frequency response functions

Gaussian Process Regression for the Estimation of Generalized Frequency Response Functions ⋆

Jeremy Stoddard [a], Georgios Birpoutsoukis [b], Johan Schoukens [c,d], James Welsh [a]

[a] *School of Electrical Engineering and Computing, University of Newcastle, Australia*

[b] *ICTEAM, Université Catholique de Louvain, B1348 Louvain la Neuve, Belgium*

[c] *Department INDI, Faculty of Engineering, Vrije Universiteit Brussel, Belgium*

[d] *Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands*

This is a postprint copy of: J. G. Stoddard, G. Birpoutsoukis, J. Schoukens, & J. S. Welsh (2019). Gaussian process regression for the estimation of generalized frequency response functions. Automatica, 106, 161-167. DOI: 10.1016/j.automatica.2019.05.010

Abstract

Bayesian learning techniques have recently garnered significant attention in the system identification community. Originally introduced for low variance estimation of linear impulse response models, the concept has since been extended to the nonlinear setting for Volterra series estimation in the time domain. In this paper, we approach the estimation of nonlinear systems from a frequency domain perspective, where the Volterra series has a representation comprised of Generalized Frequency Response Functions (GFRFs). Inspired by techniques developed for the linear frequency domain case, the GFRFs are modeled as real/complex Gaussian processes with prior covariances related to the time domain characteristics of the corresponding Volterra series. A Gaussian process regression method is developed for the case of periodic excitations, and numerical examples demonstrate the efficacy of the proposed method, as well as its advantage over time domain methods in the case of band-limited excitations.

Key words: Nonlinear system identification, Gaussian process regression, Generalized frequency response function

1 Introduction

For the identification of nonlinear systems, the Volterra series can be useful as a nonparametric model structure capable of representing any fading-memory system [15]. In the time domain, the truncated Volterra series can be viewed as a generalization of the linear finite impulse response (FIR) model, consisting of a set of multidimensional impulse responses or 'Volterra kernels'. The series can also be expressed in a frequency domain context, where several competing representations can be found in the literature [14]. The most natural representation, and the one considered in this paper, is the generalized frequency response functions (GFRFs), which are defined as the multidimensional Fourier transforms of the corresponding Volterra kernels in the time domain series [8]. This representation leads to a series of complex-valued frequency domain functions of increasing dimension.

The estimation of GFRFs has been addressed previously in the literature. An interpolation method for sec-

⋆ This paper was not presented at any IFAC meeting. Corresponding author J. Stoddard.

Email addresses: Jeremy.Stoddard@uon.edu.au (Jeremy Stoddard), georgios.birpoutsoukis@uclouvain.be (Georgios Birpoutsoukis), Johan.Schoukens@vub.ac.be (Johan Schoukens), james.welsh@newcastle.edu.au (James Welsh).

ond order kernels was developed in [11], which requires multiple input realizations to give an overdetermined problem. Other studies utilized excitations with specially placed harmonics, but such methods become prohibitively complex for high nonlinear orders, and there is a limit to the number of GFRF elements which can be estimated [3],[6]. Alternatively, full GFRF estimates can be obtained from arbitrary excitation by first identifying the time domain Volterra kernels and then transforming the latter into the frequency domain. In the case where a limited frequency band is of interest, however, identification directly in the frequency domain is more beneficial with respect to the number of estimated parameters, as will be highlighted in this paper.

Gaussian process regression (GPR) was originally considered for the identification of linear FIR systems in the time domain [12], where the impulse response is modeled as a zero mean Gaussian process with a prior covariance designed to impose smoothness and exponential decay on the estimated coefficients. For the problem of Volterra series estimation, an extension to the method was proposed in [2] and [1]. The extension relies on the design of new prior covariance structures which are capable of imposing smoothness and decay along the entire (hyper)surface of the Volterra kernels. Earlier related work can also be found in [7].

In the frequency domain, only the linear theory has been fully developed for GPR. Bayesian estimation of transfer functions and their corresponding transients was achieved in [10], by modeling both processes as real/complex Gaussian variables. The prior covariance of the transfer function was constructed by transforming common covariance structures already used in the time domain for impulse response estimation. For the special case of parallel Hammerstein systems, an extension to the linear theory has also been proposed in [16].

The contribution of this paper is to develop a frequency domain Volterra series identification method using GPR, and thus expand the capabilities of nonparametric Bayesian identification. By drawing on concepts introduced in [2] and [10], regularized GFRF estimates can be obtained up to an arbitrary order using a single multisine input excitation, even though the estimation problem is severely rank deficient. For cases where only a limited frequency band is of interest, the proposed method presents a clear advantage over indirect time domain GPR, much like the linear case in [10].

The paper is organized as follows. Section 2 provides the preliminary definitions necessary to formulate the estimation problem. In Section 3, the GPR method is developed. Some numerical examples are presented in Section 4 before the paper is concluded.

2 Preliminaries and Problem Formulation

A brief summary of the relevant concepts from Volterra theory and real/complex Gaussian processes is presented here, followed by a formulation of the identification problem to be considered in this paper.

2.1 Volterra series and GFRFs

We consider the nonlinear systems whose output can be described by the discrete time Volterra series given by,

$$y^0(t) = \sum_{m=1}^{M} \left[\sum_{\tau_1=0}^{n_m-1} \cdots \sum_{\tau_m=0}^{n_m-1} h_m(\tau_1, \ldots, \tau_m) \prod_{\tau=\tau_1}^{\tau_m} u(t-\tau) \right], \tag{1}$$

where y^0 is the Volterra output, u is the applied input, $h_m(\tau_1, \ldots, \tau_m)$ is the m^{th} order Volterra kernel, n_m is the memory length of h_m, and τ_j is the j^{th} lag variable for the kernel. The subscript m gives the dimension of the kernels up to the maximum degree, M.

The representation can be expressed in the frequency domain via Discrete Fourier Transforms (DFTs). For N-sample time domain signals $\{u(t)\}_{t=0}^{N-1}$ and $\{y^0(t)\}_{t=0}^{N-1}$, the corresponding N-point input and output DFT spectra are labelled $U(k)$ and $Y^0(k)$ respectively, and their relationship in steady state is described by [8],

$$\begin{aligned} Y^0(k) &= \sum_{m=1}^{M} Y_m(k), \\ Y_m(k) &= \sum_{k_1+\ldots+k_m=k} H_m(k_1, \ldots, k_m) \prod_{i=1}^{m} U(k_i), \end{aligned} \tag{2}$$

where $H_m(k_1, \ldots, k_m)$ is labelled the m^{th} order GFRF, given by a multidimensional DFT of the m^{th} order time domain kernel, h_m, i.e.

$$\begin{aligned} H_m(k_1, \ldots, k_m) = \sum_{\tau_1=0}^{n_m-1} \cdots \sum_{\tau_m=0}^{n_m-1} h_m(\tau_1, \ldots, \tau_m) \\ \times e^{\frac{-j2\pi k_1 \tau_1}{N}} \cdots e^{\frac{-j2\pi k_m \tau_m}{N}}. \end{aligned} \tag{3}$$

2.2 Real/complex normal distributions

In order to express the distribution of Gaussian random vectors that contain both real and complex entries, the complex normal distribution is not sufficient, since the associated augmented covariance matrix will be singular [10]. Consequently, a hybrid real/complex Gaussian (RCG) distribution framework was developed in [10], where the reader is directed for a thorough treatment of RCG vectors and their properties. The purpose of this section is to define the notation adopted in this paper.

Definition 1 (RCG vector) *A random vector*

$$X = \left[X^{\mathbb{R}^T}\ X^{\mathbb{C}^T}\right]^T, \quad X^{\mathbb{R}} \in \mathbb{R}^{n_r},\ X^{\mathbb{C}} \in \mathbb{C}^{n_c} \tag{4}$$

is said to be a RCG vector if $[X^{\mathbb{R}^T}\ \Re X^{\mathbb{C}^T}\ \Im X^{\mathbb{C}^T}]$ *is Gaussian distributed, where* T *denotes the transpose operator, and* $\Re$ *and* $\Im$ *give the real and imaginary parts respectively.*

Notation 2 (Augmented vector) *For a RCG vector,* X*, the augmented vector is defined as*

$$\widetilde{X} = [X^{\mathbb{R}^T}\ X^{\mathbb{C}^T}\ X^{\mathbb{C}^H}]^T, \tag{5}$$

where H *denotes the Hermitian transpose.*

Notation 3 (Augmented mean and covariance) *A RCG vector* X *has a real/complex normal distribution denoted by*

$$X \sim \mathcal{RCN}(\mu, \Sigma),$$

where $\mu = \boldsymbol{E}\{\widetilde{X}\}$ *is the augmented mean, and* $\Sigma = \boldsymbol{E}\{(\widetilde{X}-\mu)(\widetilde{X}-\mu)^H\}$ *is the augmented covariance. The latter can be decomposed using covariance functions* R*,* Q*,* K*, and relation function* C*, as*

$$\Sigma = \begin{bmatrix} R & Q & \overline{Q} \\ Q^H & K & C \\ \overline{Q^H} & C^H & \overline{K} \end{bmatrix} \tag{6}$$

$$\begin{aligned} \text{where } R &= \boldsymbol{E}\{(X^{\mathbb{R}} - \boldsymbol{E}\{X^{\mathbb{R}}\})(X^{\mathbb{R}} - \boldsymbol{E}\{X^{\mathbb{R}}\})^T\}, \\ Q &= \boldsymbol{E}\{(X^{\mathbb{R}} - \boldsymbol{E}\{X^{\mathbb{R}}\})(X^{\mathbb{C}} - \boldsymbol{E}\{X^{\mathbb{C}}\})^H\}, \\ K &= \boldsymbol{E}\{(X^{\mathbb{C}} - \boldsymbol{E}\{X^{\mathbb{C}}\})(X^{\mathbb{C}} - \boldsymbol{E}\{X^{\mathbb{C}}\})^H\}, \\ C &= \boldsymbol{E}\{(X^{\mathbb{C}} - \boldsymbol{E}\{X^{\mathbb{C}}\})(X^{\mathbb{C}} - \boldsymbol{E}\{X^{\mathbb{C}}\})^T\}. \end{aligned}$$

2.3 Problem formulation

The nonlinear system identification problem is formulated in the frequency domain under steady-state conditions, using the following assumptions and notation:

Assumption 4 *The measured output,* $\{y(t)\}_{t=0}^{N-1}$*, is a Volterra output corrupted by white noise, i.e.*

$$y(t) = y^0(t) + v(t); \quad v(t) \sim \mathcal{N}(0, \sigma_v^2), \tag{7}$$

where $y^0(t)$ *is the noise-free Volterra output from (1).*

Assumption 5 (Steady-state) *The measured input* $\{u(t)\}_{t=0}^{N-1}$ *is one period of an N-periodic sequence which has been applied to the system for sufficiently long such that the system is in steady-state with a corresponding N-periodic output response,* $\{y^0(t)\}_{t=0}^{N-1}$*, whose spectrum can be given by (2).*

Notation 6 *The set of DFT indices,* k*, for which the input spectrum,* $\{U(k) : k \in \mathbb{Z}, 0 \le k \le N/2\}$*, is non-zero is denoted by* $\mathbf{k_u}$*. Likewise, the set of indices for which the output spectrum,* $\{Y(k) : k \in \mathbb{Z}, 0 \le k \le N/2\}$*, is non-zero (due to input excitation) is denoted by* $\mathbf{k_y}$ *(see [9] for the precise definition of this set for a given* $U(k)$ *and* M*).*

Notation 7 *The frequencies of interest for estimation are given by the set* $\mathbf{k_e} = -\mathbf{k_u} \cup \mathbf{k_u}$*.*

Given Assumptions 4-5, the goal is to estimate the GFRF elements, $H_m(k_1, \ldots, k_m)\ \ \forall k_1, \ldots, k_m \in \mathbf{k_e}$ and $m = 1, \ldots, M$.

3 Gaussian Process Regression for GFRFs

Given the assumptions and problem formulation described in Section 2.3, a GPR method is developed here.

3.1 GFRFs as RCG vectors

Definition 8 *The* m*-dimensional array containing* $H_m(k_1, \ldots, k_m),\ \forall k_1, \ldots, k_m \in \mathbf{k_e}$ *has a real/complex vectorized form (shown in (4)) which will be denoted by*

$$H_m^{\mathcal{V}} = \left[H_m^{\mathbb{R}^T}\ H_m^{\mathbb{C}^T}\right]^T, \tag{8}$$

where $H_m^{\mathbb{R}}$ *contains the strictly real components.*

Unlike in [10], where the linear frequency functions are strictly real only at 0 frequency, higher order GFRFs originating from symmetric Volterra kernels present a larger set of real components. Since we are required to express the vectorized GFRFs in the form dictated by (8), it is important to identify the location of all strictly real elements, as addressed in the following theorem.

Theorem 9 *Consider a symmetric* m^{th} *order Volterra kernel,* $h_m(\tau_1, \ldots, \tau_m)$*, as defined in (1), and its corresponding GFRF,* $H_m(k_1, \ldots, k_m)$*, as defined in (3). The strictly real components of* H_m *are defined by the set of DFT indices,*

$$\begin{aligned} \mathbf{k}_m^{\mathbb{R}} = \{k_1, \ldots, k_m \in \mathbb{Z} : \\ \sum_{\mathcal{I} \in S_m(X)} sin\left(\frac{2\pi}{N}(k_1\tau_{i_1} + \ldots + k_m\tau_{i_m})\right) = 0\ \ \forall \tau_i \in \mathbb{N}\}, \end{aligned} \tag{9}$$

where $\mathcal{I} = [i_1, \ldots, i_m]$*, and* $S_m(X)$ *denotes the set of all permutations of the vector* $X = [1, 2, \ldots, m]$*.*

PROOF. Taking the imaginary part of (3),

$$\Im\{H_m(k_1, \ldots, k_m)\} = \sum_{\tau_1=0}^{n_m-1} \cdots \sum_{\tau_m=0}^{n_m-1} h_m(\tau_1, \ldots, \tau_m) \cdot$$

$$sin(\frac{2\pi}{N}(k_1\tau_1 + \ldots + k_m\tau_m)). \quad (10)$$

For a symmetric kernel h_m, the DFT transformation is independent of the ordering of the lags, $\tau_1, \ldots, \tau_m$, so

$$\begin{aligned}
&\Im\{H_m(k_1, \ldots, k_m)\} \\
&= \frac{1}{m!} \sum_{\mathcal{I} \in S_m(X)} \left[\sum_{\tau_1=0}^{n_m-1} \cdots \sum_{\tau_m=0}^{n_m-1} h_m(\tau_{i_1}, \ldots, \tau_{i_m}) \right. \\
&\left. sin\left(\frac{2\pi}{N}(k_1\tau_{i_1} + \ldots + k_m\tau_{i_m})\right)\right] \\
&= \frac{1}{m!} \sum_{\tau_1=0}^{n_m-1} \cdots \sum_{\tau_m=0}^{n_m-1} h_m(\tau_1, \ldots, \tau_m) \\
&\sum_{\mathcal{I} \in S_m(X)} sin\left(\frac{2\pi}{N}(k_1\tau_{i_1} + \ldots + k_m\tau_{i_m})\right).
\end{aligned}$$

Thus, for an arbitrary symmetric h_m, the strictly real elements of H_m will occur when

$$\sum_{\mathcal{I} \in S_m(X)} sin(\frac{2\pi}{N}(k_1\tau_{i_1} + \ldots + k_m\tau_{i_m})) = 0 \ \ \forall \tau_i \in \mathbb{N} \qquad \square$$

To explicitly locate the strictly real elements within the estimation set, $\mathbf{k_e}$, the result in Theorem 9 can be elaborated using the antisymmetry of $sin()$, as shown below for the first two nonlinear orders.

$$\mathbf{k}_1^{\mathbb{R}} = \{k_1 \in \mathbf{k_e} : k_1 = 0\}, \quad (11)$$
$$\mathbf{k}_2^{\mathbb{R}} = \{k_1, k_2 \in \mathbf{k_e} : k_1 = -k_2\}. \quad (12)$$

3.2 The Gaussian assumption

The following assumptions, which are consistent with the assumptions made in [2] and [10], are placed on the vectorized GFRF quantities, $H_m^{\mathcal{V}}$, and will be used in the sequel to derive the output spectrum distribution and define the GPR procedure for GFRF estimation.

Assumption 10 *$H_m^{\mathcal{V}}$ is real/complex normally distributed with zero mean and augmented covariance Σ_m, i.e.*

$$H_m^{\mathcal{V}} \sim \mathcal{RCN}(0, \Sigma_m), \quad (13)$$

where Σ_m is constructed from the covariance and relation functions R_m, Q_m, K_m and C_m as in (6).

Assumption 11 *Two (real/complex vectorized) Gaussian GFRFs of different orders are independent, i.e. $H_i^{\mathcal{V}}$ and $H_j^{\mathcal{V}}$ are independent for $i \neq j$.*

3.3 Prior covariance design

For multidimensional Volterra kernels, covariance functions have already been constructed in the time domain [2], [1], by applying a diagonal/correlated (DC) structure [5] along multiple perpendicular regularizing directions. The resulting covariance matrices are guaranteed to be valid and produce stable kernel realizations. Taking a similar approach to [10], we note that there exists a linear transformation between the vectorized kernel $h_m^{\mathcal{V}}$ and its frequency domain counterpart, $H_m^{\mathcal{V}}$, which will be denoted F_m, i.e.

$$H_m^{\mathcal{V}} = \begin{bmatrix} H_m^{\mathbb{R}} \\ H_m^{\mathbb{C}} \end{bmatrix} = \begin{bmatrix} F_m^{\mathbb{R}} \\ F_m^{\mathbb{C}} \end{bmatrix} h_m^{\mathcal{V}} = F_m h_m^{\mathcal{V}}, \quad (14)$$

where $F_m^{\mathbb{R}}$ produces $H_m^{\mathbb{R}}$ and $F_m^{\mathbb{C}}$ produces $H_m^{\mathbb{C}}$. While it is clear from (3) that F_m should contain appropriate products of exponentials, obtaining the explicit form for F_m and its submatrices requires the following theorem.

Theorem 12 *Consider the m-dimensional array $w_m \in \mathbb{R}^{N \times \ldots \times N}$, and its m-dimensional DFT, $W_m \in \mathbb{C}^{N \times \ldots \times N}$. Consider also their vectorized forms, $w_m^{\mathcal{V}} \in \mathbb{R}^{N^m}$ and $W_m^{\mathcal{V}} \in \mathbb{C}^{N^m}$, where vectorization is performed such that*

$$\begin{aligned}
w_m^{\mathcal{V}} = \big[&w_m(1, 1, \ldots, 1), w_m(2, 1, \ldots, 1), \ldots, \\
&w_m(N, 1, \ldots, 1), w_m(1, 2, \ldots, 1), \ldots, w_m(N, N, \ldots, N)\big],
\end{aligned}$$

and likewise for $W_m^{\mathcal{V}}$. The vectorized arrays are related by,

$$W_m^{\mathcal{V}} = \Psi_m w_m^{\mathcal{V}}, \quad (15)$$

where Ψ_m can be obtained from the recursive definition,

$$\Psi_p = \Psi_{p-1} \otimes \Psi_1, \ \ p = 2, 3, \ldots, m \quad (16)$$

and where Ψ_1 is the $N \times N$ DFT matrix.

PROOF. For $m = 1$, by definition we have:

$$W_1^{\mathcal{V}} = \Psi_1 w_1^{\mathcal{V}} \quad (17)$$

For $m = 2$, the DFT is applied in each dimension:

$$\begin{aligned}
W_2 &= \Psi_1 w_2 \Psi_1^T \quad (18) \\
&= \underbrace{[\Psi_1 \otimes \Psi_1]}_{\Psi_2} w_2^{\mathcal{V}}, \quad (19)
\end{aligned}$$

by the well known property of kronecker products. The proof for $m > 2$ exploits the same property, but requires the use of repeated tensor matricizations on w_m. $\square$

Assuming $h_m^{\mathcal{V}}$ is vectorized as described in Theorem 12, we can form F_m by first computing the matrix Ψ_m, then rearranging and removing rows to correspond to a desired vectorization of $H_m^{\mathcal{V}}$ which satisfies the required form in (8).

The transformation can now be used to convert time domain covariance functions to the frequency domain,

$$\begin{aligned} R_m &= \mathbf{E}\{H_m^{\mathbb{R}} {H_m^{\mathbb{R}}}^T\} = F_m^{\mathbb{R}} \mathbf{E}\{h_m^{\mathcal{V}} {h_m^{\mathcal{V}}}^T\} {F_m^{\mathbb{R}}}^T = F_m^{\mathbb{R}} P_m {F_m^{\mathbb{R}}}^T \\ Q_m &= \mathbf{E}\{H_m^{\mathbb{R}} {H_m^{\mathbb{C}}}^H\} = F_m^{\mathbb{R}} \mathbf{E}\{h_m^{\mathcal{V}} {h_m^{\mathcal{V}}}^T\} {F_m^{\mathbb{C}}}^H = F_m^{\mathbb{R}} P_m {F_m^{\mathbb{C}}}^H \\ K_m &= \mathbf{E}\{H_m^{\mathbb{C}} {H_m^{\mathbb{C}}}^H\} = F_m^{\mathbb{C}} \mathbf{E}\{h_m^{\mathcal{V}} {h_m^{\mathcal{V}}}^T\} {F_m^{\mathbb{C}}}^H = F_m^{\mathbb{C}} P_m {F_m^{\mathbb{C}}}^H \\ C_m &= \mathbf{E}\{H_m^{\mathbb{C}} {H_m^{\mathbb{C}}}^T\} = F_m^{\mathbb{C}} \mathbf{E}\{h_m^{\mathcal{V}} {h_m^{\mathcal{V}}}^T\} {F_m^{\mathbb{C}}}^T = F_m^{\mathbb{C}} P_m {F_m^{\mathbb{C}}}^T \end{aligned} \tag{20}$$

where $P_m = \mathbf{E}\{h_m^{\mathcal{V}} {h_m^{\mathcal{V}}}^T\}$ denotes the tunable DC-based covariance structure designed in [2] and [1].

Remark 13 *In practice, symmetry must be enforced in both the time and frequency domain kernels to guarantee a unique Volterra series representation [15]. Thus, we are required to modify the transformation matrices, F_m, by removing the columns which correspond to redundant time domain parameters, and removing rows that correspond to redundant frequency domain parameters. As an example, the symmetry axes for second order kernels are given in Figure 1, which are used to dictate the location of unique and redundant parameters in those kernels.*

3.4 The output spectrum

The output spectrum of $Y(\mathbf{k_y})$ can be derived in a similar fashion to the linear case. However, we must first restructure the model equation (2) as follows,

$$\begin{aligned} Y(\mathbf{k_y}) &= [\phi_1 \dots \phi_M][{H_1^{\mathcal{V}}}^T \dots {H_M^{\mathcal{V}}}^T]^T + V(\mathbf{k_y}) \\ &= \phi H + V, \end{aligned} \tag{21}$$

where ϕ_m is an appropriate regressor containing the input spectrum products corresponding to $H_m^{\mathcal{V}}$. Note that symmetry should also be enforced in the GFRFs here, which should be reflected in the design of the regressors.

$Y(\mathbf{k_y})$ will also be a real/complex vector in the case where $0 \in \mathbf{k_y}$. Thus, we extend (21) to the augmented output case, resulting in

$$\begin{aligned} \widetilde{Y}(\mathbf{k_y}) &= [\widetilde{\phi_1} \dots \widetilde{\phi_M}][\widetilde{H_1^{\mathcal{V}}}^T \dots \widetilde{H_M^{\mathcal{V}}}^T]^T + \widetilde{V}(\mathbf{k_y}) \\ &= \widetilde{\phi}\widetilde{H} + \widetilde{V}. \end{aligned} \tag{22}$$

Remark 14 *When ϕ is built from a single input/output realization as in (21), the associated linear regression problem will be severely rank deficient for $M > 1$. While it is possible to make the problem overdetermined by 'stacking' a sufficient number of unique input/output realizations, we will focus on the rank deficient case in this paper to highlight the flexibility of GPR.*

From (22), we can now derive the distribution of the output spectrum.

Theorem 15 *For a system whose output is given by (7), with Gaussian $H_m^{\mathcal{V}} \sim \mathcal{RCN}(0, \Sigma_m)$ as described in (13), the output spectrum $Y(\mathbf{k_y})$ is complex normally distributed as follows,*

$$\begin{aligned} Y(\mathbf{k_y}) &\sim \mathcal{RCN}(0, \Sigma_Y), \\ \text{where } \Sigma_Y &= \widetilde{\phi}\Sigma_{tot}\widetilde{\phi}^H + \sigma_v^2 I \\ \text{and } \Sigma_{tot} &= \begin{bmatrix} \Sigma_1 & & 0 \\ & \ddots & \\ 0 & & \Sigma_M \end{bmatrix} \end{aligned} \tag{23}$$

PROOF. Follows from (22), Assumptions 4, 10-11, and the properties of real/complex normal distributions. □

3.5 MAP estimates of the GFRFs

Maximum a posteriori (MAP) estimates for the GFRFs can be obtained from the joint distribution of Y and H, by computing the mean of the conditional distribution $H|Y$. The result is provided in the following theorem.

Theorem 16 *The MAP estimate of $\widetilde{H}$ in (22) is*

$$\hat{\widetilde{H}}_{MAP} = \Sigma_{tot}\widetilde{\phi}^H \Sigma_Y^{-1} \widetilde{Y} \tag{24}$$

PROOF. Follows from the properties of real/complex distributions [10]. □

3.6 Hyperparameter tuning

We must optimize the hyperparameters, η, describing each covariance function $P_m(\eta)$ from [2], which are in turn used to construct Σ_m. As in [2] and [10], this is performed via a marginal likelihood maximization, i.e.

$$\hat{\eta} = \arg\min_{\eta} \widetilde{Y}(\mathbf{k_y})^H \Sigma_Y^{-1}(\eta) \widetilde{Y}(\mathbf{k_y}) + \log\det \Sigma_Y(\eta). \tag{25}$$

Note that the optimization problem is non-convex in general [13], and can be computationally intensive to solve. Numerical techniques have been developed to reduce the computational burden in the case of long data records [4] or high nonlinear orders [17].

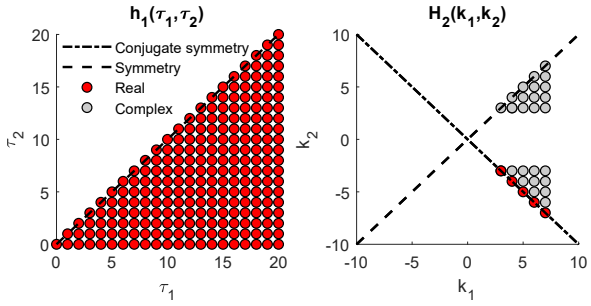

Fig. 1. Parameters requiring estimation in second order band-limited example

3.7 Comparison with time domain approach

There is another viable approach for estimating GFRFs under rank deficient conditions, which is to first compute regularized Volterra kernel estimates in the time domain with memory length $n_m = N$, and subsequently transform them using (3). One advantage of this indirect approach is the freedom to use input/output measurements of arbitrary length which are not necessarily at steady-state, whereas the method proposed here requires such constraints.

If we are only interested in GFRF estimation for a specific frequency band, then the time domain approach has a severe disadvantage when compared to the frequency domain method proposed here. With an indirect time domain approach, we are required to estimate the full set of kernel parameters regardless of frequency band, since each frequency domain parameter depends on the *entire* time domain set (see (3)). In direct frequency domain estimation, the limited frequency band directly reduces the set of parameters to be estimated, and the magnitude of reduction grows rapidly with increasing GFRF order, m, since the band is limited along every dimension of the GFRFs.

An example of this parameter reduction concept is visualized in Figure 1 for a second order example with $N = n_2 = 21$. We excite a discrete set of frequencies, $\mathbf{k_u} = [3, 4, 5, 6, 7]$, and compare the number of parameters requiring estimation in each domain. It is clear that limiting the excited frequency band in two dimensions significantly reduces the parameter set for H_2 to 55 parameters (5 real, 25 complex), while for h_2 the full set of unique Volterra coefficients (231 real) is still required.

4 Simulation Examples

Two simulation studies are presented here in order to demonstrate the performance of the proposed method and compare with the equivalent time domain approach.

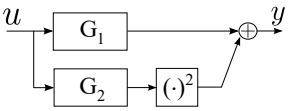

Fig. 2. System structure used for data generation in the rank deficient noise-free case

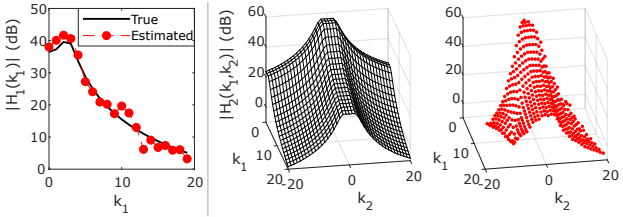

Fig. 3. First order GFRF estimate (left), true 2nd order GFRF (middle) and 2nd order estimate for unique parameters (right)

4.1 Rank deficient noise-free case

A Volterra system output was generated using the block structure in Figure 2, where the linear filters are defined in the z-domain as

$$G_1(z) = \frac{3z}{z^2 - 1.8z + 0.84}, \quad G_2(z) = 0.25G_1(z).$$

The input, $u(t)$, is given by a periodic multisine with period $N = 80$ and $\mathbf{k_u} = \{0, 1, 2, \ldots, 19\}$, which is applied for 10 periods before taking the final period of measurements for estimation purposes, so that all transient effects are negligible.

We consider identification of the system's GFRFs, H_1 and H_2, assuming memory lengths $n_1 = n_2 = N$. The resulting estimation problem is severely rank deficient, with 780 unique parameters requiring estimation, and only 77 corresponding output spectrum measurements. Applying the proposed GPR method, the GFRF estimates are given in Figure 3 alongside the true GFRFs, which can be obtained analytically from the system structure [18]. Despite severe rank deficiency, a reasonable result is achieved at both GFRF orders.

4.2 Limited frequency band with measurement noise

To highlight the advantage of the proposed method over an indirect time domain approach, a Monte Carlo (MC) study was performed. The noise-free outputs were generated using the structure in Figure 4, with linear filters,

$$G_1(z) = \frac{z - 1}{z}, \quad G_2(z) = \frac{1.07z}{z^2 - 1.72z + 0.77}.$$

The input is a multisine with period $N = 55$ and a limited band of excited frequencies, $\mathbf{k_u} = \{4, 5, \ldots, 12\}$. The output, y, is disturbed by white measurement noise.

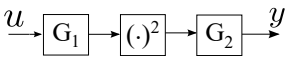

Fig. 4. Wiener-Hammerstein structure used to generate kernel h_2 in the MC study

Note that the system structure can be represented using a single GFRF, H_2 [18], and we will assume $n_2 = N$.

For the MC study three different noise levels, and for each noise level 100 input/output realizations, are considered. For each realization, transient-free estimation data was obtained in the same manner as in Section 4.1, and estimation of the H_2 parameters within the excited band was performed using two separate methods:

(1) **GPR-FD** - GPR in the frequency domain as developed in this paper (i.e. using (24) and (25)).
(2) **GPR-TD** - GPR in the time domain as presented in [2],[1] and using periodicity to define initial conditions, followed by a DFT transformation using (3), and elimination of parameters not relevant to the frequency band of interest.

The noise, $v(t)$, was added in each realization to achieve a Signal-to-Noise Ratio (SNR) of 5dB, 10dB or 20dB.

Estimation errors for each method and realization are quantified using a normalised mean square error (NMSE) metric, given by

$$e_{NMSE} = \frac{\frac{1}{n}\sum_{i=1}^{n} |\hat{H}_2^{\mathcal{V}}(i) - H_2^{\mathcal{V}}(i)|^2}{\frac{1}{n}\sum_{i=1}^{n} |H_2^{\mathcal{V}}(i)|^2}, \qquad (26)$$

where n is the number of GFRF parameters requiring estimation inside the excited frequency band, $\hat{H}_2^{\mathcal{V}}(i)$ is the i^{th} element of the GFRF estimate, and $H_2^{\mathcal{V}}(i)$ contains the true GFRF elements within the frequency band of interest.

The errors resulting from each noise level and method are presented as boxplots in Figure 5. The GPR-FD accuracy is seen to improve with increasing SNR as expected, however it is also clear that for the case of a limited excitation frequency band, the proposed frequency domain regression method performs better than an indirect time domain approach on the same dataset, for the reasons discussed in Section 3.7. More specifically, the number of unique parameters requiring estimation for GPR-TD is 1540 (all real), while for GPR-FD it is only 378 (14 real, 182 complex).

5 Conclusions

In this paper, a GPR method was presented for the estimation of generalized frequency response functions. The method employs a hybrid real/complex Gaussian framework originally developed for the linear case, and translates the properties of previously designed time domain prior covariances to the frequency domain via vectorized DFT transformations. Numerical examples demonstrated the ability of the method to produce reasonable estimates under rank deficient conditions, and the proposed method clearly outperforms the indirect time domain estimation approach in the case of a limited frequency band of interest.

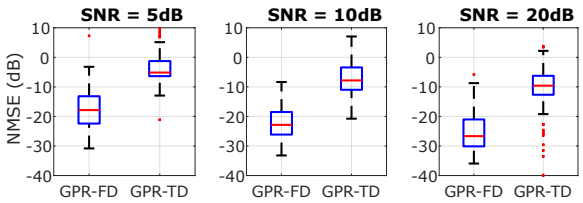

Fig. 5. NMSEs for the limited frequency band MC study

Acknowledgements

This work was supported in part by an Australian government Research Training Program (RTP) scholarship and the Priority Research Centre for Complex Dynamic Systems and Control (CDSC) at the University of Newcastle, Australia, and in part by the Fund for Scientific Research (FWO-Vlaanderen), the Flemish Government (Methusalem), the Belgian Government through the Inter university Poles of Attraction (IAP VII) Program, and the ERC advanced grant SNLSID, under contract 320378.

References

[1] G. Birpoutsoukis, P. Z. Csurcsia, and J. Schoukens. Efficient multidimensional regularization for Volterra series estimation. *Mechanical Systems and Signal Processing*, 104:896–914, 2018.
[2] G. Birpoutsoukis, A. Marconato, J. Lataire, and J. Schoukens. Regularized nonparametric Volterra kernel estimation. *Automatica*, 82:324–327, 2017.
[3] S. Boyd, Y. Tang, and L. Chua. Measuring Volterra kernels. *IEEE Transactions on Circuits and Systems*, 30(8):571–577, 1983.
[4] T. Chen and L. Ljung. Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. *Automatica*, 49:2213–2220, 2013.
[5] T. Chen, H. Ohlsson, G. Goodwin, and L. Ljung. Kernel selection in linear system identification - Part II: a classical perspective. In *Proc. of CDC-ECC*, pages 4326–4331, 2011.
[6] C. Evans, D. Rees, L. Jones, and M. Weiss. Periodic signals for measuring nonlinear Volterra kernels. *IEEE Transactions on Instrumentation and Measurement*, 45(2):362–371, 1996.

[7] M. O. Franz and B. Schölkopf. A unifying view of Wiener and Volterra theory and polynomial kernel regression. *Neural Computation*, 18(12):3097–3118, 2006.

[8] D. George. *Continuous Nonlinear Systems*. MIT RLE Technical Report, 1959.

[9] Z. Q. Lang and S. A. Billings. Output frequencies of nonlinear systems. *International Journal of Control*, 67(5):713–730, 1997.

[10] J. Lataire and T. Chen. Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica*, 72:217–229, 2016.

[11] J. G. Nemeth, I. Kollar, and J. Schoukens. Identification of Volterra kernels using interpolation. *IEEE Transactions on Instrumentation and Measurement*, 51(4):770–775, 2002.

[12] G. Pillonetto and G. De Nicolao. A new kernel-based approach for linear system identification. *Automatica*, 46(1):81–93, 2010.

[13] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[14] D. Rijlaarsdam, P. Nuij, J. Schoukens, and M. Steinbuch. A comparative overview of frequency domain methods for nonlinear systems. *Mechatronics*, 42(Supplement C):11–24, 2017.

[15] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. John Wiley and Sons, 1980.

[16] J. G. Stoddard and J. S. Welsh. Frequency domain estimation of parallel Hammerstein systems using Gaussian process regression. In *Proceedings of the 18th IFAC Symposium on System Identification*, pages 1014–1019, 2018.

[17] J. G. Stoddard, J. S. Welsh, and H. Hjalmarsson. EM-based hyperparameter optimization for regularized Volterra kernel estimation. *IEEE Control Systems Letters*, 1(2):388–393, 2017.

[18] D. T. Westwick and R. E. Kearney. *Identification of Nonlinear Physiological Systems*. John Wiley and Sons, 2003.