Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Nonparametric Identification of Linear Time-Varying Systems using Gaussian Process Regression ⋆

N. Hallemans * J. Lataire * R. Pintelon *

* *Vrije Universiteit Brussel, Brussels, Belgium (e-mail: noel.hallemans@vub.be, john.lataire@vub.be, rik.pintelon@vub.be).*

Abstract: Linear time-varying systems are a class of systems, the dynamics of which evolve in time. This results in a time-varying frequency response function where each frequency has a time-varying gain. In classical identification techniques, basis functions are employed to fit these time-varying gains. In this paper a new method based on Gaussian process regression is presented. The advantage of the proposed method is a more convenient model structure and model order selection.

Keywords: Identification, Linear time-varying systems, Gaussian processes, Machine learning

1. INTRODUCTION

This paper describes methods for identifying linear time-varying (LTV) systems. The need for modelling LTV systems appears in diverse engineering disciplines. For instance, in aeronautics, the resonance frequency of the wings of an airplane depend on the flight speed and height. Different approaches exist in order to describe LTV systems, one of them is to use recursive identification methods as in Ljung and Söderström (1983), another way is to model the linear ordinary differential equation with time-varying coefficients as described in Lataire and Pintelon (2011). In this paper, however, a frequency domain approach is used, where the time-varying frequency response function (TV-FRF) is identified nonparametrically over frequency and parametrically over time. For this parametric identification over time, classical methods (Lataire et al., 2012) use basis function regression. However, this method has some disadvantages regarding model structure and model order selection. This makes the method hard to use for non-experts in system identification. In order to bypass this problem, a new method is proposed where Gaussian process regression, as described in Williams and Rasmussen (2006), is used instead of a basis function approach. The main advantage is that a convenient model complexity selection procedure is available. Hence, the modelling is simplified for non-experts in system identification.

This paper is structured as follows. First, Section 2 introduces the TV-FRF. Next, Section 3 explains the current methods for the identification of linear time-varying systems in a nutshell. Further, Section 4 introduces Gaussian process regression and applies it to a very simple regression problem. Next, Section 5 employs Gaussian process regression for identifying the TV-FRF at a particular frequency and Section 6 explains briefly how to obtain the TV-FRF with a multisine excitation, together with results on measured data. Finally, Section 7 concludes the Gaussian process approach by balancing the advantages and disadvantages.

2. AN ELEMENTAL KNOWLEDGE ON LINEAR TIME-VARYING SYSTEMS

Linear time-varying systems can be described using their two-dimensional impulse response function $g(t,\tau)$, which is the response of the system at time t when an impulse has been applied at time τ. In contrast with linear time-invariant systems, the impulse response is explicitly dependent on the time at which the impulse has been applied. The input u - output y relation is then given by Kwakernaak and Sivan (1991)

$$y(t) = \int_{-\infty}^{+\infty} g(t,\tau)u(\tau)d\tau. \quad (1)$$

By transforming to the frequency domain, the latter equation yields (Zadeh, 1950)

$$y(t) = \mathcal{L}^{-1}\{G(s,t)U(s)\}, \quad (2)$$

where $U(s)$ is the Laplace transform of $u(t)$ and $G(s,t)$ is called the time-varying transfer function and is given by the Laplace transform of the impulse response, i.e.

$$G(s,t) = \int_{0}^{+\infty} g(t,t-\tau)e^{-s\tau}d\tau, \quad (3)$$

where s is the Laplace variable. From equation (2), the identification of the LTV system boils down to obtaining the time-varying frequency response function (TV-FRF) $G(j\omega,t)$, which is a two-dimensional complex function. In this paper the system will be identified nonparametrically. This means that instead of considering both $j\omega$ and t as continuous variables, ω will be discretised and one needs to find a continuous complex time function, which is called a time-varying gain, for each frequency $\omega_e \in \mathbb{W}$, where $\mathbb{W}$ is the discrete set of frequencies. An illustration of a TV-FRF is shown in Fig. 1, with the time-varying gain at 1 Hz indicated in red.

⋆ This research was supported in part by the Fund for Scientific Research (FWO Vlaanderen), and in part by the Flemish Government (Methusalem Grant METH1).

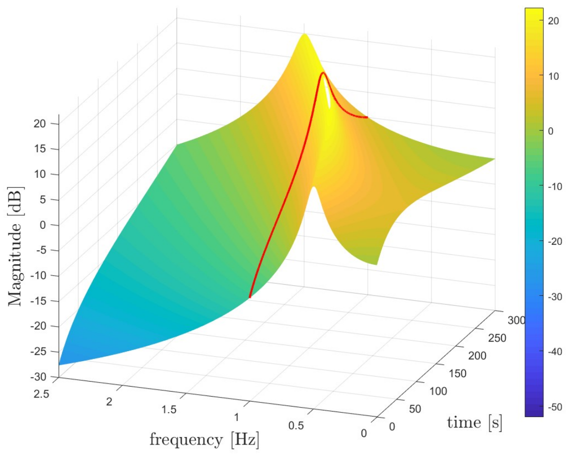

Fig. 1. Illustration of a TV-FRF $G(j\omega, t)$ where the time-varying gain $G(j\omega_e, t)$ for $\omega_e = 2\pi$ rad/s is indicated in red.

3. STATE OF THE ART: IDENTIFICATION OF LINEAR TIME-VARYING SYSTEMS

In the literature, Lataire et al. (2012) and Pintelon et al. (2015), the identification problem is tackled by projecting the time-varying gain at a particular frequency ω_e on a set of basis functions $b_p(t)$, i.e. a series expansion is used,

$$G(j\omega_e, t) = \sum_{p=0}^{N_p} G_p(j\omega_e) b_p(t). \quad (4)$$

The identification then boils down to obtaining the $N_p + 1$ complex scalars $G_p(j\omega_e)$. This is done for all $\omega_e \in \mathbb{W}$. This is solved in Lataire et al. (2012), where Legendre polynomials are used as basis functions. When using multiple periods of a multisine as an excitation for a linear time-varying system, the effect of the time variation can be distinguished in the output spectrum. Using this frequency domain information, the coefficients $G_p(j\omega_e)$ where $p = 0, ..., N_p$ and $\omega_e \in \mathbb{W}$ can be estimated using simple regression. Withal, this method has a disadvantage: the user of the identification tools has to choose the number of basis functions N_p needed to describe the system optimally. This is a non-trivial problem, especially for non-experts in system identification. In the remainder of this paper this problem will be by-passed by using a Gaussian process approach. To begin with, Gaussian processes are defined and applied to a simple time domain regression problem.

4. GAUSSIAN PROCESS REGRESSION

4.1 What are Gaussian processes?

A Gaussian process is an extension of a multivariate Gaussian distribution, where instead of drawing vectors from the distribution, one draws continuous functions. This is now further formalised. The notational conventions of Williams and Rasmussen (2006) are used.

Definition 1. Consider the stochastic process $f(t) \in \mathbb{R}$ which depends on the continuous time variable $t \in [0, T]$. We call $f(t)$ a Gaussian process if any vector $\mathbf{f}$, given by

$$\mathbf{f} = \begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_N) \end{bmatrix} \qquad t_1, t_2, ..., t_N \in [0, T], N \in \mathbb{N} \quad (5)$$

follows a multivariate Gaussian distribution, i.e. $\mathbf{f} \sim \mathcal{N}(\mu, K)$. In other words,

$$p(\mathbf{f}) = \frac{1}{\sqrt{(2\pi)^N |K|}} \exp\big(-\frac{1}{2}(\mathbf{f}-\mu)^T K^{-1}(\mathbf{f}-\mu)\big), \quad (6)$$

where $|K| = \det(K)$.

As this multivariate normal distribution is completely determined by the mean vector μ and covariance matrix K, a Gaussian process is completely determined by a mean function, which is set to zero for simplicity,

$$\mu(t) = \mathbb{E}\{f(t)\} = 0 \quad (7)$$

and covariance function

$$k(t, t') = \text{cov}\{f(t), f(t')\} = \mathbb{E}\{f(t)f(t')\}. \quad (8)$$

One denotes the Gaussian process

$$f(t) \sim \mathcal{GP}(0, k(t, t')), \quad (9)$$

which is a distribution of continuous functions. These Gaussian processes showcase very useful results in the following regression problem in the presence of noise.

4.2 Usefulness of Gaussian process regression

Problem 1. (Time domain regression). Given the prior knowledge about the signal and the noise, $f(t) \sim \mathcal{GP}(0, k(t, t'))$ and $v(t) \sim \mathcal{GP}(0, \sigma_v^2 \delta(t - t'))$, given data $y(t_i) = f(t_i) + v(t_i)$, find the distribution of the model conditioned on the data $y = [y(t_1), ..., y(t_N)]^T$.

This is a simple regression problem, where the data y is for instance a vector with temperatures measured in Brussels on a sunny day in May, shown as dots in Fig. 2. Elegant expressions exist to solve this problem. It is shown in Williams and Rasmussen (2006) Appendix A.2 that for normally distributed real zero mean vectors $a \in \mathbb{R}^{N\times 1}$ and $b \in \mathbb{R}^{N\times 1}$ with joint covariance

$$\text{cov}\Big\{\begin{bmatrix} a \\ b \end{bmatrix}\Big\} = \begin{bmatrix} \mathbb{E}\{aa^T\} & \mathbb{E}\{ab^T\} \\ \mathbb{E}\{ba^T\} & \mathbb{E}\{bb^T\} \end{bmatrix} = \begin{bmatrix} A & C^T \\ C & B \end{bmatrix}, \quad (10)$$

the posterior distribution of model a conditioned on data b yields

$$a \mid b \sim \mathcal{N}(C^T B^{-1} b, A - C^T B^{-1} C). \quad (11)$$

One applies this very convenient formula to the regression problem, and finds

$$\text{cov}\Big\{\begin{bmatrix} \mathbf{f} \\ y \end{bmatrix}\Big\} = \begin{bmatrix} K & K \\ K & K + \sigma_v^2 I \end{bmatrix}, \quad (12)$$

where

$$[K]_{m,n} = k(t_m, t_n) \quad m, n = 1, ..., N. \quad (13)$$

And hence using (11) the posterior distribution is given by

$$\mathbf{f} \mid y \sim \mathcal{N}(\hat{\mathbf{f}}, K_{\hat{\mathbf{f}}}) \quad (14)$$

with mean vector

$$\hat{\mathbf{f}} = K\big(K + \sigma_v^2 I\big)^{-1} y \quad (15)$$

and covariance matrix

$$K_{\hat{\mathbf{f}}} = K - K\big(K + \sigma_v^2 I\big)^{-1} K. \quad (16)$$

Equations (15) and (16) give the posterior distribution at the time instants of the drawn data. Furthermore, Gaussian processes also allow interpolation such that the posterior distribution is available for a continuous time variable t,

$$f(t) \mid y \sim \mathcal{GP}(\hat{f}(t), k_{\hat{f}}(t,t')), \quad (17)$$

with mean function

$$\hat{f}(t) = k(t,\mathbf{t})\big(K + \sigma_v^2 I\big)^{-1} y \quad (18)$$

and covariance function

$$k_{\hat{f}}(t,t') = k(t,t') - k(t,\mathbf{t})\big(K + \sigma_v^2 I\big)^{-1} k(\mathbf{t},t'), \quad (19)$$

where $\mathbf{t} = [t_1 \; t_2 \; ... \; t_N]^T$,

$$k(t,\mathbf{t}) = [k(t,t_1) \; k(t,t_2) \; ... \; k(t,t_N)] \quad (20)$$

and

$$k(\mathbf{t},t') = \big[k(t_1,t') \; k(t_2,t') \; ... \; k(t_N,t')\big]^T. \quad (21)$$

This is a very useful result because the posterior distribution is still a Gaussian process with an analytical mean function and covariance function. Note that also uncertainty bounds can be computed

$$\mathrm{p}\big(|f(t) - \hat{f}(t)| < \sqrt{2}\, \mathrm{erf}^{-1}(\alpha)\sigma_{\hat{f}}(t)\big) = \alpha, \quad (22)$$

where p is the probability, erf is the error function and $\sigma_{\hat{f}}^2(t) = k_{\hat{f}}(t,t)$.

In the remainder of this paper these formulas are applied to the fitting of the TV-FRF (3). Yet, the covariance function $k(t,t')$ was not fixed to a particular shape, this is done in the next paragraph.

4.3 Choice of covariance function

In this paper the squared exponential kernel is used as a covariance function,

$$k(t,t') = a \exp\Big(-\frac{(t-t')^2}{l^2}\Big). \quad (23)$$

Note that many other types of kernels exist. However, the squared exponential kernel is commonly used in the case of smooth functions. This covariance function depends on two hyper parameters a and l, which set, respectively, the variance and the correlation length. The higher l the more correlation between two different time instants of the model $f(t)$. Now remains the question, how do the hyper parameters a and l have to be chosen? This is solved in the next paragraph, employing the marginal likelihood.

4.4 Hyper parameter selection via Marginal likelihood

The previous paragraphs assume that the prior knowledge about the distribution of the model and noise are known. Evidently, this is not the case in practice, and hence the hyper parameters a, l and σ_v have to be estimated. Denote $\Theta = \{a, l, \sigma_v\}$ which is the set of hyper parameters. The distribution of the data y yields

$$y \sim \mathcal{N}(0, K_y) \quad (24)$$

where $K_y = K + \sigma_v^2 I$. Note that this covariance matrix depends on the three hyper parameters. One now defines the marginal likelihood as the probability of the data y conditioned on the hyper parameters Θ,

$$p(y \mid \Theta) = \frac{1}{\sqrt{(2\pi)^N |K_y|}} \exp\Big(-\frac{1}{2} y^T K_y^{-1} y\Big). \quad (25)$$

Obviously, one chooses the hyper parameters Θ such that the marginal likelihood is maximised. Moreover, it is numerically advantageous to maximise the logarithm of the marginal likelihood

$$\log p(y \mid \Theta) = -\frac{1}{2} y^T K_y^{-1} y - \frac{1}{2}\log|K_y| - \frac{N}{2}\log 2\pi \quad (26)$$

and hence

$$\hat{\Theta} = \arg\max_{\Theta} \log p(y \mid \Theta). \quad (27)$$

In this paper the MATLAB routine *global search* is used for solving (27), this procedure minimises a cost function under constraints. The constraints are the bounds on the hyper parameters, $l \in [l_{\min}, l_{\max}]$ etc. These bounds still have to be chosen by the user. Also, initialisation values should be chosen for the hyper parameters.

An illustration of the mean function $\hat{f}(t)$ with 95% uncertainty bounds for temperature data in Brussels is shown in Fig. 2. The estimated hyper parameters are $l = 7.97$ h, $a = 372.09°\mathrm{C}^2$ and $\sigma_v^2 = 0.52°\mathrm{C}^2$.

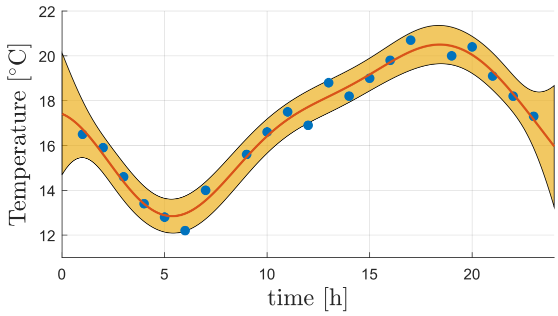

Fig. 2. Regression of temperature data in Brussels. Dots: data y. Full line: $\hat{f}(t) = \mathbb{E}\{f(t) \mid y\}$. Area: 95 % uncertainty bounds.

5. IDENTIFICATION OF THE TIME-VARYING GAIN USING GAUSSIAN PROCESS REGRESSION

5.1 Posterior distribution conditioned on windowed frequency domain data

In this Section we want to identify the TV-FRF (3) at a particular frequency ω_e, which is called a time-varying gain. We first formulate the problem, which is similar to *Problem 1*, but more involved as the function to estimate is complex and will be conditioned on windowed frequency domain data.

Problem 2. (Regression of the time-varying gain). Given the prior knowledge about respectively the time-varying gain and noise,

$$G(j\omega_e, t) \sim \mathcal{GP}_c(0, k(t,t')) \quad (28)$$

and

$$v(t) \sim \mathcal{GP}(0, \sigma_v^2 \delta(t-t')), \quad (29)$$

given time domain measurements

$$y = [y(t_1), ..., y(t_N)]^T, \quad (30)$$

where $y(t) = G\{u(t)\} + v(t)$, $u(t) = \cos(\omega_e t)$, $t_i = (i-1)/f_s$ and f_s is the sampling frequency of the measurement, find the a posteriori distribution of the model conditioned on a windowed set of data points Y_w, i.e.

$$G(j\omega_e, t) \mid Y_w. \quad (31)$$

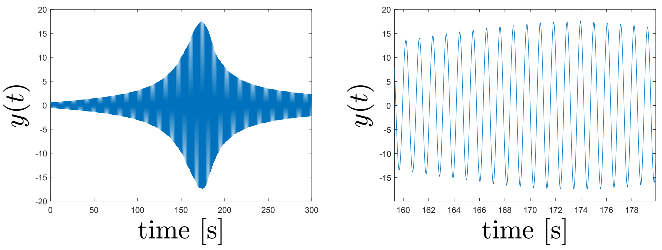

Fig. 3. Output signal of the LTV system in Fig. 1 when excited with a sine at $f_e = 0.94$ Hz. Left: entire response for measurement length $T = 300$ s, right: zoomed version.

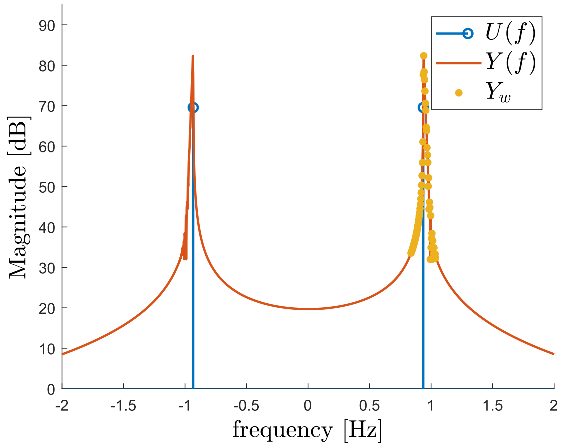

Fig. 4. Input and output frequency spectrum of the LTV system of Fig. 1 when excited by a sine at $\omega_e = 1.88\pi$ rad/s. The windowed data Y_w is indicated as yellow dots, $U(f) = \mathcal{F}_t\{u(t)\}(f)$ and $Y(f) = \mathcal{F}_t\{y(t)\}(f)$.

Here, Y_w is a vector containing $N_w = 2\Delta + 1$ frequency domain output data points centered around ω_e, i.e. in the band

$$\mathbb{B} = \{\omega^w_{-\Delta}, \omega^w_{-\Delta+1}, ..., \omega^w_{\Delta-1}, \omega^w_{\Delta}\}, \tag{32}$$

where $\omega^w_m = \omega_e + \frac{2\pi m}{T} > 0$ and $T = N/f_s$. Hence,

$$Y_w = F_w y \tag{33}$$

where F_w consists of N_w rows of the DFT matrix, viz.

$$[F_w]_{m,n} = \frac{1}{\sqrt{N}} \exp(-j\omega^w_{m-(\Delta+1)} t_n) \tag{34}$$

for $n = 1, ..., N$ and $m = 1, ..., N_w$. As an illustration, Fig. 3 shows the output signal $y(t)$ of the LTV system in Fig. 1 when excited by a sine at frequency $\omega_e = 1.88\pi$ rad/s and Fig. 4 shows Y_w where $N_w = 61$.

Important to note is that the function to be estimated, $G(j\omega_e, t)$, is complex. The following assumption ensures that the prior is a complex Gaussian process with a real covariance function.

Assumption 1. (Prior knowledge). The real and imaginary parts of the complex function are mutually uncorrelated real zero mean Gaussian processes, i.e.

$$\begin{bmatrix} \text{re}\{G(j\omega_e, t)\} \\ \text{im}\{G(j\omega_e, t)\} \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k(t,t')/2 & 0 \\ 0 & k(t,t')/2 \end{bmatrix} \right), \tag{35}$$

where $k(t, t') \in \mathbb{R}$ is the squared exponential kernel, which is justifiable when the time-varying gain is a smooth function in time. One then computes that

$$G(j\omega_e, t) \sim \mathcal{GP}_c(0, k(t, t')). \tag{36}$$

The initial step is to obtain an expression for Y_w based on $G(j\omega_e, t)$. We start from the continuous response of an LTV system to a sinusoidal excitation. From (2), this yields

$$y(t) = |G(j\omega_e, t)| \cos\left(\omega_e t + \angle G(j\omega_e, t)\right) + v(t). \tag{37}$$

First, one notices that $y(t)$ is a sinusoidal signal modulated with the time-varying gain $G(j\omega_e, t)$ and hence all information about $G(j\omega_e, t)$ is confined in $y(t)$. This is clearly visible in Fig. 3.

In order to obtain $G(j\omega_e, t)$ as a complex signal, we consider the noiseless analytical signal of (37) divided by two, which yields

$$y_a(t) = \frac{1}{2}\big(y(t) + jH(y)(t)\big) \tag{38}$$

$$= \frac{1}{2} G(j\omega_e, t) \exp(j\omega_e t) \in \mathbb{C}, \tag{39}$$

where H denotes the Hilbert transform. Transforming this analytical signal to the frequency domain yields

$$Y_a(j\omega) = \mathcal{F}_t\{y_a(t)\}(j\omega) \tag{40}$$

$$= \frac{1}{2} \int_{-\infty}^{+\infty} G(j\omega_e, t) e^{-j(\omega-\omega_e)t} dt \tag{41}$$

$$= \frac{1}{2} G(j\omega_e, j(\omega - \omega_e)), \tag{42}$$

where $G(j\omega_e, j\omega) = \mathcal{F}_t\{G(j\omega_e, t)\}(j\omega)$. Hence, the noiseless analytical output spectrum of the system is equal to the spectrum of the time-varying gain, but shifted over ω_e. Based on (39) one can write Y_w (33) as

$$Y_w = F_w D G_{\omega_e} + V, \tag{43}$$

where $V \sim \mathcal{N}_c(0, \sigma_V^2 I)$ represents the DFT of the noise,

$$[D]_{m,n} = \exp(j\omega_e t_m)\delta_{mn} \quad m, n = 1, 2, ..., N \tag{44}$$

and

$$G_{\omega_e} = [G(j\omega_e, t_1)\ G(j\omega_e, t_2)\ ...\ G(j\omega_e, t_N)]^T \tag{45}$$

is the time-varying gain vector we want to obtain. Note that Y_w is a windowed and sampled version of $Y_a(j\omega)$ defined in (42). Now, the posterior distribution $G_{\omega_e} \mid Y_w$ can be computed. We start by writing down the joint covariance of G_{ω_e} and Y_w,

$$\text{cov}\left\{ \begin{bmatrix} G_{\omega_e} \\ Y_w \end{bmatrix} \right\} = \begin{bmatrix} K & KD^H F_w^H \\ F_w DK & F_w DKD^H F_w^H + \sigma_V^2 I \end{bmatrix}, \tag{46}$$

where

$$K = \text{cov}\{G_{\omega_e}\} \text{ and } [K]_{m,n} = k(t_m, t_n) \quad m, n = 1, ..., N. \tag{47}$$

From (11), the posterior distribution is a multivariate complex normal distribution

$$G_{\omega_e} \mid Y_w \sim \mathcal{N}_c(\hat{G}_{\omega_e}, K_{\hat{G}_{\omega_e}}), \tag{48}$$

where

$$\hat{G}_{\omega_e} = KD^H F_w^H (F_w DKD^H F_w^H + \sigma_V^2 I)^{-1} Y_w \tag{49}$$

and

$$K_{\hat{G}_{\omega_e}} = K - KD^H F_w^H (F_w DKD^H F_w^H + \sigma_V^2 I)^{-1} F_w^H DK. \tag{50}$$

Continuous time estimates can be found as

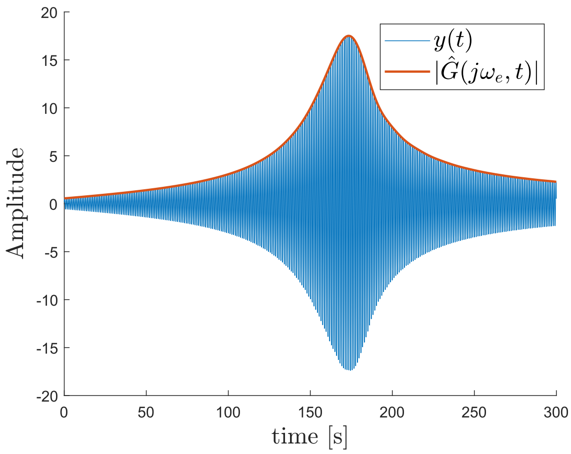

Fig. 5. Time-varying gain $\hat{G}(j\omega_e, t)$ given by (49) conditioned on frequency domain data Y_w.

$$G(j\omega_e, t) \mid Y_w \sim \mathcal{GP}_c\big(\hat{G}(j\omega_e, t), k_{\hat{G}_{\omega_e}}(t, t')\big), \quad (51)$$

where

$$\hat{G}(j\omega_e, t) = k(t, \mathbf{t})D^H F_w^H (F_w D K D^H F_w^H + \sigma_V^2 I)^{-1} Y_w \quad (52)$$

and

$$k_{\hat{G}_{\omega_e}}(t, t') = k(t, t') - \quad (53)$$
$$k(t, \mathbf{t})D^H F_w^H (F_w D K D^H F_w^H + \sigma_V^2 I)^{-1} F_w^H D k(\mathbf{t}, t'). \quad (54)$$

Results are shown in Fig. 5, where the magnitude of the time-varying gain is indicated in red and is in good agreement with the envelope of the time-domain data y. As the model is a distribution, also uncertainty bounds are available, i.e.

$$\mathrm{p}\big(|G(j\omega_e, t) - \hat{G}(j\omega_e, t)| < \sqrt{-\log(1-\alpha)}\sigma_{\hat{G}}(j\omega_e, t)\big) = \alpha, \quad (55)$$

where $\sigma^2_{\hat{G}}(j\omega_e, t) = k_{\hat{G}_{\omega_e}}(t, t)$.

5.2 Hyper parameter estimation

Again, hyper parameters must be estimated for the prior knowledge. This is done by using the marginal likelihood, similarly to Section 4.4.

$$p(Y_w \mid \Theta) = \frac{1}{\sqrt{(2\pi)^{N_w}|K_w|}} \exp\Big(-\frac{1}{2} Y_w^H K_w^{-1} Y_w\Big), \quad (56)$$

where

$$K_w = F_w D K D^H F_w^H + \sigma_V^2 I. \quad (57)$$

The loglikelihood yields

$$\log p(Y_w \mid \Theta) = -\frac{1}{2} Y_w^H K_w^{-1} Y_w - \frac{1}{2}\log|K_w| - \frac{N_w}{2}\log 2\pi. \quad (58)$$

The hyper parameters $\hat{\Theta}$ are then found as

$$\hat{\Theta} = \arg\max_{\Theta} \log p(Y_w \mid \Theta). \quad (59)$$

Note that computing the matrix product

$$K_f = F_w D K D F_w^H \quad (60)$$

is extremely time consuming when N is large. Hence, to make the hyper parameter estimation faster we propose to use the analytical version of this function which we compute now. The squared exponential kernel in the frequency domain is given by

$$K(j\omega, j\omega') = \mathrm{cov}\{G(j\omega_e, j\omega), G(j\omega_e, j\omega')\} \quad (61)$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} k(t, t') e^{-j(\omega t - \omega' t')} dt dt' \quad (62)$$
$$= a\sqrt{\pi l^2} \exp\big(-(\frac{l\omega}{2})^2\big)\delta(\omega - \omega'), \quad (63)$$

where δ represents the Dirac function. This formula shows that there is no correlation between different frequencies and that the variance is the highest for frequencies near 0 Hz. By all means, these continuous Fourier transforms are only valid when infinitely long and continuous time data is available, which is never the case in practice. Hence it is only an approximation. Further it can be shown that

$$K_{Y_a}(j\omega, j\omega') = \mathrm{cov}\{Y_a(j\omega), Y_a(j\omega')\} \quad (64)$$
$$= K(j(\omega - \omega_e), j(\omega' - \omega_e)), \quad (65)$$

and hence K_f can be evaluated from $K_{Y_a}(j\omega)$ which is much faster than computing the matrix products, however, an error is introduced but this can be afforded to obtain reasonable values of the hyper parameters. Further speed up of the matrix inversion will be investigated in future work, for example by using reduced-rank Gaussian process regression (Solin and Särkkä, 2014).

5.3 Numerical improvements

Gaussian process regression does not scale well. When working with long data records, i.e. when N is large, the matrices in the equations for the posterior distribution become very large. Hence, computing matrix products is not recommended. For the multiplications with the DFT matrix F_w the fast Fourier transform can be used and note that multiplying a frequency domain kernel with the diagonal matrix D boils down to shifting the kernel vertically and horizontally over $k_e = \omega_e T/2\pi$ elements.

6. EXTENSION TO BROADBAND EXCITATIONS

6.1 Multisine framework

If one wants to know the time-varying gain in a discrete set of frequencies, the usual excitation signal is a random phase multisine, defined as

$$u(t) = \sum_{\omega_k \in \mathbb{W}} u_k \sin(\omega_k t + \varphi_k), \quad (66)$$

where the set $\{u_k\}$ are the amplitudes, $\{\varphi_k\}$ are the random phases and $\mathbb{W}$ is the discrete set of excited frequencies. The output signal $y(t) = G\{u(t)\}$ is then computed using linearity,

$$y(t) = \sum_{\omega_k \in \mathbb{W}} u_k |G(j\omega_k, t)| \sin\big(\omega_k t + \varphi_k + \angle G(j\omega_k, t)\big). \quad (67)$$

Hence, the response $y(t)$ contains the information of the time-varying gains at all the excited frequencies $\omega_k \in \mathbb{W}$. When measuring multiple periods of the multisine, skirts will be visible in the DFT spectrum around the excited frequencies. Hence, one can use the method described in Section 5 to obtain the TV-FRF, by applying it to all the excited frequencies. Technical details will be conveyed in future work and results are shown next.

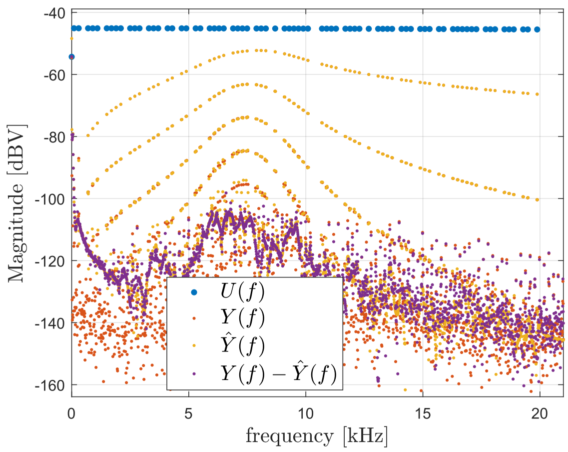

Fig. 6. Reconstructed output spectrum using the estimated TV-FRF on validation data.

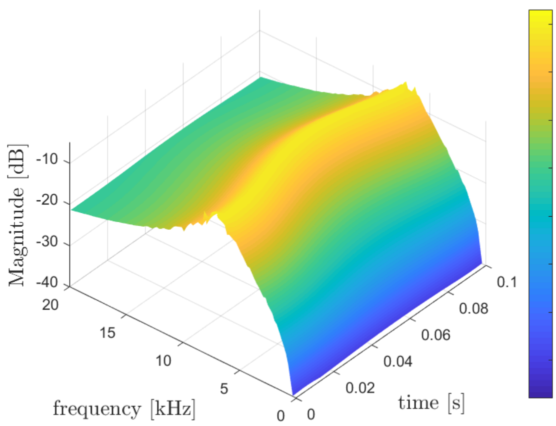

Fig. 7. Estimated TV-FRF $\hat{G}(j\omega, t)$, obtained from the benchmark data.

6.2 Results on measurements

In this Section, results of the identification are shown on measured benchmark data from Lataire et al. (2015). Here, a multisine was applied to a parameter varying electronic circuit. A sinusoidal scheduling parameter has been applied to the system. The input and output spectra are shown in Fig. 6. The TV-FRF is estimated using the three excited frequencies algorithm, as briefly described in Section 6.1, and is shown in Fig. 7. One can now reconstruct the output signal of a validation data set using the estimated model $\hat{G}(j\omega_k, t)$ and (67). Fig. 6 shows this reconstructed output spectrum compared to the measured validation output spectrum, hence showing the model accuracy. The skirt around 0 Hz is intentionally not modelled. The errors over frequency are reasonable, which validates the model.

7. CONCLUSION

In this paper a new algorithm was developed in order to identify a time-varying gain from input-output data when a sine was used as an excitation signal. This was done by using Gaussian process regression instead of the classical basis function approach. This new method has both advantages and disadvantages.

The main advantage of Gaussian process regression is that no model order should be selected. This makes the identification more accessible for non-experts. Another important advantage is that the model is a distribution of continuous functions, this means that uncertainty bounds are available.

However, these advantages come at a price. In this Gaussian process approach hyper parameters $\Theta = \{l, a, \sigma_V\}$ have to be estimated. This estimation boils down to maximising the marginal likelihood given by (58). This is an optimisation problem in three dimensions. Furthermore, the marginal likelihood is not necessarily convex. Hence, nonlinear optimisation tools must be used and finding a global maximum is not ensured. Also, Gaussian Gaussian processes do not scale well, which makes the identification slow.

The described method offers a new perspective to the identification of linear time-varying systems. The ultimate objective is to make the model order selection automatic, and hence to facilitate the process for the user. Still, there is room for improvement regarding the optimisation and the computational efficiency.

REFERENCES

Kwakernaak, H. and Sivan, R. (1991). Modern signals and systems. *NASA STI/Recon Technical Report A*, 91.

Lataire, J., Louarroudi, E., Pintelon, R., and Rolain, Y. (2015). Benchmark data on a linear time- and parameter-varying system. *IFAC-PapersOnLine*, 48(28), 1477–1482.

Lataire, J. and Pintelon, R. (2011). Frequency-domain weighted non-linear least-squares estimation of continuous-time, time-varying systems. *IET Control Theory & Applications*, 5(7), 923–933.

Lataire, J., Pintelon, R., and Louarroudi, E. (2012). Non-parametric estimate of the system function of a time-varying system. *Automatica*, 48(4), 666–672.

Ljung, L. and Söderström, T. (1983). *Theory and practice of recursive identification*. MIT press.

Pintelon, R., Louarroudi, E., and Lataire, J. (2015). Non-parametric time-variant frequency response function estimates using arbitrary excitations. *Automatica*, 51, 308–317.

Solin, A. and Särkkä, S. (2014). Hilbert space methods for reduced-rank gaussian process regression. *arXiv preprint arXiv:1401.5508*.

Williams, C.K. and Rasmussen, C.E. (2006). *Gaussian Processes for Machine Learning*, volume 2. MIT Press Cambridge, MA.

Zadeh, L.A. (1950). Frequency analysis of variable networks. *Proceedings of the IRE*, 38(3), 291–299.