VUB VRIJE
UNIVERSITEIT
BRUSSEL

Transfer function and transient estimation by Gaussian process regression in the frequency domain

Lataire, John; Chen, Tianshi

Link to publication

Transfer function and transient estimation by Gaussian process regression in the frequency domain

John Lataire [a] and Tianshi Chen [b]

[a]*Vrije Universiteit Brussel (VUB), Dept. ELEC, Pleinlaan 2, 1050 Brussels, Belgium*

[b]*School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China*

Abstract

Inspired by the recent promising developments of Bayesian learning techniques in the context of system identification, this paper proposes a Transfer Function estimator, based on Gaussian process regression. Contrary to existing kernel-based impulse response estimators, a frequency domain approach is adopted. This leads to a formulation and implementation which is seamlessly valid for both continuous- and discrete-time systems, and which conveniently enables the selection of the frequency band of interest. A pragmatic approach is proposed in an output error framework, from sampled input and output data. The transient is dealt with by estimating it simultaneously with the transfer function.

Modelling the transfer function and the transient as Gaussian processes allows for the incorporation of relevant prior knowledge on the system, in the form of suitably designed kernels. The SS (Stable Spline) and DC (Diagonal Correlated) kernels from the literature are translated to the frequency domain, and are proven to impose the stability of the estimated transfer function. Specifically, the estimates are shown to be stable rational functions in the frequency variable. The hyperparameters of the kernel are tuned via marginal likelihood maximisation.

The comparison between the proposed method and three existing methods from the literature – the regularised finite impulse response (RFIR) estimator, the Local Polynomial Method (LPM), and the Local Rational Method for Frequency Response Function estimation – is illustrated on simulations on multiple case studies.

Key words: Transfer function estimation, Gaussian Process Regression, Frequency domain, System identification, Regularised Least Squares

1 Introduction

When considering the identification of Linear Time Invariant (LTI) systems, an important initial step is the non-parametric estimation of their Transfer Functions (TF) [6], [16, Chapters 2–7]. It provides the user with insight into the dynamic behaviour of the system, even before any attempt is made to determine a parametric model.

The estimation of a transfer function ought to consider that the available signals are confined to a finite time interval. This is typically handled by including initial conditions (time domain) or an additional transient (frequency domain) in the estimation process. The transient takes into account the fact that the input and output signals are not necessarily periodic, or that their periodicity does not match the length of the measurement window, as explained later on.

The Frequency Response Function (FRF) of a system is defined in [16, Chapter 2] as the evaluation of its TF – a

continuous function – at a discrete set of frequencies. FRF estimation has been studied extensively in the past, starting with tools for spectral analysis [1,2,23]. These tools aim at suppressing the transient by applying carefully designed windows. Alternatively, [26] applies a frequency dependent smoothing procedure to the Empirical Transfer Function Estimate (ETFE) to suppress the transient. More recent work makes use of an intrinsic property of the transient to estimate it simultaneously with the FRF, yielding much better results. Namely, both the FRF and the transient are known to be smooth functions of the frequency [16, Appendix 6.B]. Therefore, their estimation can be performed via smoothers. By following this point of view, the references [19,20,24] use a local polynomial smoother, and will be referred to as the Local Polynomial Method (LPM), while [8] discusses the Local Rational Method (LRM), which uses a local rational function as a smoother. It is worth to note that both the LPM and the LRM provide a set of local models centred around the bins of the DFT (Discrete Fourier Transform), for which the interpolation in-between the DFT bins remains an open question. Consequently, the stability of the LPM and LRM estimates is also undefined.

Recently, new results for LTI system identification have been reported [14,12,4] on the estimation of impulse responses, which is the time domain equivalent of the TF. The impulse response estimation is formulated as a Gaussian process regression problem, which can also be interpreted as a regularization method. More specifically, the impulse response is modeled as a real and zero mean Gaussian process with suitably chosen and tuned covariance (often called kernel) functions. The impulse response estimate is then given by the conditional mean of the Gaussian process conditioned on the given data. This method will be denoted by RFIR (Regularised least squares for estimating the Finite Impulse Response). It has two unique features which were not present in earlier methods. The first one is that the kernel function is designed to embed the prior knowledge, e.g., stability and smoothness of the impulse response, into the estimation problem. The second one is that the model complexity is tuned in a continuous way and handled by maximizing the marginal likelihood of the hyper-parameters used to parameterize the kernel function. This approach is known to enable an automatic trade-off between the model fit and the model complexity [7,21], and as pointed out in [13], is more reliable than existing complexity measures, such as the Akaike's criterium (AIC) or cross validation, especially for small data sets.

In this paper, Gaussian process regression is applied directly to the TF and the transient estimation. That is, the estimation is formulated in the frequency domain. The resulting estimate will be denoted by GPTF. A particular attention is deserved to the fact that the TF and the transient are complex valued functions, but that at some frequencies – at 0 Hz and at the Nyquist frequency for discrete time systems – they should be real valued. For that

reason, the method will be developed in the context of mixed real/complex Gaussian processes. Next, properties of the associated frequency domain kernels, applicable to LTI systems, will be derived and a sufficient condition on the kernel will be formulated to impose the stability of the GPTF estimate. Then, it will be shown how the time domain kernels – Stable Spline and Diagonal Correlated – proposed in [14,12,4] are transformed to the frequency domain and satisfy the condition of stability.

It will be shown that the RFIR estimate in [14,12,4,13] and the GPTF in the frequency domain are dual to each other, under specific conditions. However, from a practical point of view, the frequency-domain formulation is shown to give a more appealing implementation than the RFIR when working with continuous-time systems. This is because the explicit computation of the convolution between the input and the impulse response is circumvented. A second advantage of the GPTF over the RFIR is that it allows the estimation to be performed in a limited frequency band. The main advantages of the GPTF over the LPM and the LRM are that the estimated transfer function is guaranteed to be stable, and that it is expressed as a continuous function of the frequency.

The remaining part of this paper is organised as follows. Gaussian processes for regression of mixed real and complex valued functions are developed in Section 2. The formulation of the TF estimation problem is given in Section 3, and is rewritten as a Bayesian regression problem in Section 4. The choice and construction of kernels in the frequency domain is discussed in Section 5, and the duality with regularised impulse response estimation is given in Section 6. The Gaussian process TF estimator is compared with the LPM, the LRM and the RFIR in Section 7 on case studies. Section 8 concludes this paper. The appendix provides the proofs of the Lemma and the Theorems.

2 Real/complex Gaussian distributions

Transfer Functions (TFs) will be modelled as Gaussian processes. Since a TF takes both real (at 0 Hz and at the Nyquist frequency for discrete time systems) and complex values, it cannot be modelled as either a real or a complex random variable. In particular, treating a real Gaussian random variable as a complex one leads to a singular covariance matrix, see Remark 1. This prompts us to introduce the so-called real/complex Gaussian (RCG) distribution to model the TF.

Definition 1 (RCG distribution) *A random vector*

$$Z = \begin{bmatrix} Z_r^T & Z_c^T \end{bmatrix}^T, \quad Z_r \in \mathbb{R}^{n_r}, Z_c \in \mathbb{C}^{n_c}. \quad (1)$$

is said to be real/complex Gaussian distributed, if $\begin{bmatrix} Z_r^T & \Re Z_c^T & \Im Z_c^T \end{bmatrix}^T$ *is Gaussian distributed, where* $\Re$, $\Im$

denote the real and imaginary parts respectively, and the superscript T denotes the transpose of a vector.

We introduce below its probability density function and derive its characteristic parameters. Define

$$Z_{\text{re}} = \left[Z_{\text{r}}^T \ \Re Z_{\text{c}}^T \ \Im Z_{\text{c}}^T \right]^T, \ z_{\text{re}} = \left[z_{\text{r}}^T \ \Re z_{\text{c}}^T \ \Im z_{\text{c}}^T \right]^T,$$
$$z_{\text{r}} \in \mathbb{R}^{n_{\text{r}}}, z_{\text{c}} \in \mathbb{C}^{n_{\text{c}}} \tag{2}$$

The probability density function of Z_{re} is described by

$$p(z_{\text{re}}) = \frac{1}{\sqrt{\det 2\pi\Gamma_{\text{re}}}} \dots$$
$$\times \exp\left(-\frac{1}{2}(z_{\text{re}} - m_{\text{re}})^H \ \Gamma_{\text{re}}^{-1} \ (z_{\text{re}} - m_{\text{re}})\right) \tag{3}$$

where $m_{\text{re}} = \mathbb{E}\{Z_{\text{re}}\}$ and $\Gamma_{\text{re}} = \mathbb{E}\left\{(Z_{\text{re}} - m_{\text{re}})(Z_{\text{re}} - m_{\text{re}})^T\right\}$, and the superscript H denotes the Hermitian transpose of a vector. Define

$$\left[m_{\text{r}}^T \ m_{\text{c}}^T \right]^T = \mathbb{E}\{Z\} \tag{4a}$$
$$\Gamma_{\text{r}} = \mathbb{E}\left\{(Z_{\text{r}} - m_{\text{r}})(Z_{\text{r}} - m_{\text{r}})^T\right\} \tag{4b}$$
$$\Gamma_{\text{rc}} = \mathbb{E}\left\{(Z_{\text{r}} - m_{\text{r}})(Z_{\text{c}} - m_{\text{c}})^H\right\} \tag{4c}$$
$$\Gamma_{\text{c}} = \mathbb{E}\left\{(Z_{\text{c}} - m_{\text{c}})(Z_{\text{c}} - m_{\text{c}})^H\right\} \tag{4d}$$
$$C_{\text{c}} = \mathbb{E}\left\{(Z_{\text{c}} - m_{\text{c}})(Z_{\text{c}} - m_{\text{c}})^T\right\} \tag{4e}$$

then it is easy to verify that

$$m_{\text{re}} = \left[m_{\text{r}}^T \ \Re m_{\text{c}}^T \ \Im m_{\text{c}}^T \right]^T, \tag{5}$$

$$\Gamma_{\text{re}} = \begin{bmatrix} \Gamma_{\text{r}} & \Re\Gamma_{\text{rc}} & -\Im\Gamma_{\text{rc}} \\ \Re\Gamma_{\text{rc}}^H & \Re(\Gamma_{\text{c}} + C_{\text{c}})/2 & \Im(C_{\text{c}} - \Gamma_{\text{c}})/2 \\ \Im\Gamma_{\text{rc}}^H & \Im(\Gamma_{\text{c}} + C_{\text{c}})/2 & \Re(\Gamma_{\text{c}} - C_{\text{c}})/2 \end{bmatrix}. \tag{6}$$

Now, define $\tilde{z} = \left[z_{\text{r}}^T \ z_{\text{c}}^T \ z_{\text{c}}^H \right]^T$. It holds that

$$z_{\text{re}} = M\tilde{z}, \ \text{with} \ M = \begin{bmatrix} I_{n_{\text{r}}} & 0 & 0 \\ 0 & I_{n_{\text{c}}}/2 & I_{n_{\text{c}}}/2 \\ 0 & -jI_{n_{\text{c}}}/2 & jI_{n_{\text{c}}}/2 \end{bmatrix} \tag{7}$$

In turn, using (7) in (3) yields

$$p(z_{\text{re}}) = \frac{1}{\sqrt{2^{n_{\text{r}}} \pi^{n_{\text{r}} + 2n_{\text{c}}} \det\Gamma}} \dots$$
$$\times \exp\left(-\frac{1}{2} (\tilde{z} - m)^H \Gamma^{-1} (\tilde{z} - m)\right) \tag{8}$$

with (an overline denotes a complex conjugate)

$$m = \begin{bmatrix} m_{\text{r}} \\ m_{\text{c}} \\ \overline{m_{\text{c}}} \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \Gamma_{\text{r}} & \Gamma_{\text{rc}} & \overline{\Gamma_{\text{rc}}} \\ \Gamma_{\text{rc}}^H & \Gamma_{\text{c}} & C_{\text{c}} \\ \overline{\Gamma_{\text{rc}}^H} & C_{\text{c}}^H & \overline{\Gamma_{\text{c}}} \end{bmatrix}, \tag{9}$$

which is true by noting $\Gamma_{\text{re}} = M\Gamma M^H$ and $\det\Gamma_{\text{re}} = 2^{-2n_{\text{c}}} \det\Gamma$. When $n_{\text{c}} = 0$ in (1), Z becomes real Gaussian distributed and (8) (same as (3)) is its probability density. When $n_{\text{r}} = 0$ in (1), Z becomes complex Gaussian distributed and (8) is referred to as its probability density in the literature, see e.g., [11],[25, Section 2.3].

Notation 1 (Augmented vector and covariance) *In the following, Γ in (8) will be called the augmented covariance matrix of Z in the sense of $\Gamma = \mathbb{E}\left\{(\tilde{Z} - m)(\tilde{Z} - m)^H\right\}$, where*

$$\tilde{Z}^T = \left[Z_{\text{r}}^T \ Z_{\text{c}}^T \ Z_{\text{c}}^H \right]. \tag{10}$$

Accordingly, $\tilde{Z}$ is called the augmented vector associated with Z. The matrix C_{c} is known as the relation matrix of Z_c; see e.g., [25, Section 2.3.1].

Notation 2 *The real/complex normally distributed Z is denoted by*

$$Z \sim \mathscr{RCN}(m, \Gamma). \tag{11}$$

The corresponding $\Gamma_{\text{r}}, \Gamma_{\text{rc}}, \Gamma_{\text{c}}, C_{\text{c}}$ in (4) can be identified from Γ according to (9) when n_{r} and n_{c} are given.

Since the definition of the RCG distribution is essentially based on that of the joint Gaussian distribution of its real and imaginary parts, many results valid for the real Gaussian distribution translate straightforwardly to the real/complex case.

Theorem 1 *Consider $A \sim \mathscr{RCN}(m_A, \Gamma_A)$, $B \sim \mathscr{RCN}(m_B, \Gamma_B)$. The following properties hold:*

Property 1 (Independent sum) *Assume A, B are independent and have the same dimensional real and complex part (see (1)), then*

$$A + B \sim \mathscr{RCN}(m_A + m_B, \Gamma_A + \Gamma_B). \tag{12}$$

Property 2 (Element-wise product) *Assume U is a given real/complex vector and has the same dimensional real and complex part as A, then the element-wise (or Hadamard) product of A, U (denoted by $U \odot A$ below) is*

$$U \odot A \sim \mathscr{RCN}\left(\tilde{U} \odot m_A, (\tilde{U}\tilde{U}^H) \odot \Gamma_A\right), \tag{13}$$

3

where $\tilde{U}$ is the augmented vector associated with U.

Property 3 (Conditional distribution) *Assume A and B are jointly real/complex Gaussian distributed (it is said so if A_re and B_re as defined in (2) are jointly Gaussian distributed). Let*

$$\Gamma_{AB} = \mathbb{E}\left\{(\tilde{A} - m_A)(\tilde{B} - m_B)^H\right\}, \tag{14}$$

where $\tilde{A}$ and $\tilde{B}$ are the augmented vectors associated with A and B, respectively. Then the conditional distribution of A given B is real/complex distributed as $A|B \sim \mathcal{RCN}\left(m_{A|B}, \Gamma_{A|B}\right)$ with

$$m_{A|B} = m_A + \Gamma_{AB}\Gamma_B^{-1}(\tilde{B} - m_B), \tag{15}$$
$$\Gamma_{A|B} = \Gamma_A - \Gamma_{AB}\Gamma_B^{-1}\Gamma_{BA} \tag{16}$$

Proof. The proof is similar to the complex Gaussian case, see [25, Section 2.3] and thus is omitted.

Remark 1 *If a real random vector is treated as a complex one, then the associated augmented covariance matrix will be singular. For example, let $X \in \mathbb{R}$ be a zero-mean random variable. If X is treated as a complex random variable, then its associated augmented covariance matrix is given by $\mathbb{E}\left\{[X \ \overline{X}]^T[X \ \overline{X}]\right\} = \mathbb{E}\left\{[X \ X]^T[X \ X]\right\}$, and is clearly singular.*

2.1 Real/complex Gaussian processes

Similarly to *real* Gaussian processes, the RCG distribution can be generalised to the real/complex Gaussian process (RCGP). First introduce the following notation.

Definition 2 (Real/complex Gaussian process) *A complex random function $\zeta(k)$ over the argument domain $k \in \mathbb{R}$ is an RCGP if for every finite set $\{k_1, \cdots, k_n\} \subset \mathbb{R}$, $\zeta(k_1), \cdots, \zeta(k_n)$ are jointly RCG distributed. The subset $\mathbb{K}_\mathfrak{R} \subset \mathbb{R}$ indicates where $\zeta(k)$ is real, i.e. $k \in \mathbb{K}_\mathfrak{R} \implies \zeta(k) \in \mathbb{R}$.*

The RCGP $\zeta(k)$ is completely determined by $\mathbb{K}_\mathfrak{R}$ and its first and second order moment functions, given by the mean function m, the covariance function K and the relation function C, defined as (for $k, k' \in \mathbb{R}$)

$$m(k) = \mathbb{E}\{\zeta(k)\} \tag{17a}$$
$$K(k, k') = \mathbb{E}\left\{(\zeta(k) - m(k))\overline{(\zeta(k') - m(k'))}\right\} \tag{17b}$$
$$C(k, k') = \mathbb{E}\{(\zeta(k) - m(k))(\zeta(k') - m(k'))\}. \tag{17c}$$

Notation 3 *The real/complex Gaussian process $\zeta(k)$, with mean function m, covariance function K and relation function C is denoted as:*

$$\zeta(k) \sim \mathcal{RCGP}(m, K, C) \mid \mathbb{K}_\mathfrak{R} \tag{18}$$

and $\mathbb{K}_\mathfrak{R}$ indicates where $\zeta(k)$ is real.

Notation 4 (Vectorized scalar function) *Let a scalar function $\zeta(k)$ and an ordered set $\text{k} = \{k_1, k_2, \ldots, k_n\}$ be defined. Evaluating $\zeta(\text{k})$ is defined as*

$$\zeta(\text{k}) = \begin{bmatrix} \zeta(k_1) \ \zeta(k_2) \ \ldots \ \zeta(k_n) \end{bmatrix}^T \tag{19}$$

A similar notation will be used for scalar functions with two arguments $K(k, l)$:

$$K(\text{k}, \text{l}) = \begin{bmatrix} K(k_1, l_1) & \ldots & K(k_1, l_n) \\ \vdots & \ddots & \vdots \\ K(k_n, l_1) & \ldots & K(k_n, l_n) \end{bmatrix} \tag{20}$$

The stochastic vector $\zeta(\text{k})$ with $\text{k} \subset \mathbb{R}$ is RCG distributed, whose mean is $m(\text{k})$ and whose augmented covariance matrix Γ is computed according to (9), by use of (17) and (4), taking into account $\mathbb{K}_\mathfrak{R}$. In what follows, RCGPs will be used to model transfer functions.

3 Context of system identification

3.1 System assumptions

Consider the causal, discrete- or continuous-time LTI system, with input/output description given by

$$y_\circ(t) = (h * u)(t) \tag{21a}$$
$$y(t) = y_\circ(t) + v(t) \tag{21b}$$

where the convolution $(h * u)(t)$ is defined as

$$(h * u)(t) = \begin{cases} \sum_{n=0}^{\infty} u((t - n)T_\text{s})h(nT_\text{s}) & \text{discrete-time} \\ \int_0^\infty u(t - \tau)h(\tau)d\tau & \text{continuous-time} \end{cases}$$

Here, $y(t)$, $v(t)$ and $y_\circ(t)$ are the measured output, measurement noise, and convolution of the impulse response $h(\cdot)$ and input $u(\cdot)$ at time instant t, respectively. Without loss of generality, assume that $y(t)$ is measured at $t = 0, T_\text{s}, \cdots, (N - 1)T_\text{s}$ where $T_\text{s} = 1/f_\text{s}$ is the sampling period, and f_s is the sampling frequency. We make the following assumption on the system.

Assumption 1 (Stability) *Assume that the considered system is stable, i.e. its impulse response $h(t)$ is absolutely integrable (continuous time) or absolutely summable (discrete time).*

Denote $U(k)$, $Y(k)$, $Y_\circ(k)$ and $V(k)$ the N-point DFTs, at frequency bin k, of $u(t)$, $y(t)$, $y_\circ(t)$ and $v(t)$, *viz.*:

Definition 3 (N-point DFT) *The N-point DFT, at frequency bin k, of a sampled signal $x(nT_s)$, $n = 0, \cdots, N-1$ is given by*

$$X(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(nT_s) e^{\frac{-j2\pi kn}{N}}, \qquad k \in \mathbb{Z}. \qquad (22)$$

Property 4 *For the N-point DFT $X(k)$ it holds that*

$$X(k) = \overline{X(-k)} \qquad (23a)$$
$$X(k + pN) = X(k) \qquad for\ p \in \mathbb{Z} \qquad (23b)$$
$$X(0) \in \mathbb{R} \qquad (23c)$$
$$X(N/2) \in \mathbb{R} \qquad for\ even\ N. \qquad (23d)$$

Notation 5 *Denote Ω a generalised frequency variable, i.e. for continuous time $\Omega = j\omega$, and for discrete time $\Omega = e^{j\omega T_s}$, for $\omega \in \mathbb{R}$. Then, Ω_k is defined as*

$$\Omega_k = j\omega_k = \frac{j2\pi k f_s}{N} \qquad for\ continuous\ time \qquad (24)$$

$$\Omega_k = e^{j\omega_k T_s} = e^{\frac{j2\pi k}{N}} \qquad for\ discrete\ time. \qquad (25)$$

with $k \in \mathbb{R}$. Note that, for $k \in \mathbb{Z}$, Ω_k corresponds to the kth bin of an N-point DFT.

Property 5 *For a continuous-time, windowed signal $u(t)$, the N-point DFT $U(k)$ is related with the sampled Fourier transform $U_{[0,NT_s]}(j\omega_k) = \int_0^{NT_s} u(t)e^{-j\omega_k t}\ dt$ as follows, for $k = 0, 1, \cdots, \lfloor N/2 \rfloor$ [9, Section 7.4], [3, Chapter 6]:*

$$U(k) = \frac{f_s}{\sqrt{N}} U_{[0,NT_s]}(j\omega_k) + \delta_u(k), \qquad (26)$$

where $\lfloor N/2 \rfloor$ rounds $N/2$ to the nearest integer towards minus infinity and $\delta_u(k)$ is the alias error. This also applies to $y_\circ(t)$.

Assumption 2 *Assume that, for the continuous-time signals $u(t)$ and $y_\circ(t)$, the N-point DFTs are alias error free, viz.: $\delta_u(k) = 0$, $\delta_{y_\circ}(k) = 0$.*

Remark 2 *For the discrete-time, windowed signal $u(t)$, the Fourier transform $U_{[0,N-1]}(e^{j\omega_k T_s}) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u(t)e^{-j\omega_k T_s t}$, $k \in \mathbb{Z}$, is equal to its N-point DFT. This also applies to $y_\circ(t)$.*

Now, the time domain description of the system (21) is transformed to the frequency domain as follows. From Appendix A, [16, Section 6.3.2] and [18], and under Assumption 2 and Remark 2, the DFTs $Y_\circ$ and U satisfy the following relation, for $k = 0, 1, \cdots, \lfloor N/2 \rfloor$

$$Y_\circ(k) = G(\Omega_k)U(k) + T(\Omega_k) \qquad (27a)$$
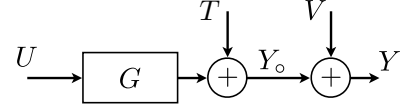$$Y(k) = Y_\circ(k) + V(k), \qquad (27b)$$

Fig. 1. Output error schematic w.r.t. the N-point DFT spectra. The input spectrum U is assumed to be known exactly, the exact output spectrum $Y_\circ$ contains a transient T. The measured output spectrum Y is disturbed by the noise with spectrum V. Arguments were omitted for convenience.

These expressions are depicted in Fig. 1.

The transfer function $G(\Omega)$ is computed as the Fourier transform – denoted by $\mathcal{F}$ (either continuous or discrete) – of the impulse response, *viz.* $G(\Omega) = \mathcal{F}\{h(t)\}$. In (27a), $T(\Omega_k)$ is the transient, which depends on the difference $u(t) - u(t + NT_s)$, for $t < 0$, and on the impulse response of the system, see Lemma 1.

Remark 3 *A necessary condition for Assumption 2 to be valid is that the sampling frequency f_s is greater than or equal to the Nyquist sampling rate [3, Chapter 5]. However, for non-periodic windowed signals, a small residual alias error always remains, see [16, Appendix 6.F]. Nevertheless, it has been shown in [15] that this error is, to a great extent, captured by the estimated transient $T(j\omega_k)$ and, thus, has negligible influence on the estimated transfer function. In addition, this error decreases for an increased distance between the upper bound of the considered frequency band and the Nyquist frequency. For these reasons, $\delta_{y_\circ}(k)$ and $\delta_u(k)$ will be neglected in this paper.*

Property 6 *The Fourier transform $G(\Omega)$, of the real function $h(t)$, explicitly defined as (continuous- and discrete-time respectively)*

$$G(j\omega) = \int_0^\infty h(t)e^{-j\omega t}\ dt \qquad (28a)$$

$$G(e^{j\omega T_s}) = \sum_{n=0}^\infty h(nT_s)e^{-j\omega n T_s} \qquad (28b)$$

is a complex function satisfying ($k \in \mathbb{R}$)

$$G(\Omega_k) = \overline{G(\Omega_{-k})} \qquad (29a)$$
$$G(\Omega_0) \in \mathbb{R}. \qquad (29b)$$

For the discrete time case specifically, it holds that

$$G(\Omega_{N/2}) = G(e^{-j\pi}) \in \mathbb{R} \qquad (30)$$
$$G(\Omega_{k+pN}) = G(\Omega_k) \qquad p \in \mathbb{Z}. \qquad (31)$$

This property also applies to $T(\Omega)$. In that perspective,

introduce the set $\mathbb{K}_{\Re}$ where $G(\Omega_k)$ and $T(\Omega_k)$ are real:

$$\mathbb{K}_{\Re} = \begin{cases} \{0\} & \text{(continuous time)} \\ \{0, \pm N/2, \pm 2(N/2), \cdots\} & \text{(discrete time)} \end{cases} \tag{32}$$

This yields $G(\Omega_k), T(\Omega_k) \in \mathbb{R}$ for $k \in \mathbb{K}_{\Re}$.

Remark 4 (Continuous frequency dependence) *From Definition 3, the N-point DFT spectra are only defined at $k \in \mathbb{Z}$. Nevertheless, $G(\Omega_k)$ and $T(\Omega_k)$ are defined for $k \in \mathbb{R}$ as well (also for discrete time systems), since they are continuous functions of the frequency variable Ω.*

Assumption 3 *Assume that the sampled disturbing noise $v(nT_s)$ ($n \in \mathbb{Z}$), with DFT spectrum $V(k)$, is white and stationary. As a consequence, the N-point DFT of the noise is uncorrelated, has a constant variance, and its relation matrix is 0 (i.e. the noise is complex circular), viz.*

$$\mathbb{E}\left\{V(k)\overline{V(k')}\right\} = \delta_{kk'}\sigma_v^2, \qquad k, k' \in \{0, 1, \ldots, \lfloor N/2 \rfloor\}$$

$$\mathbb{E}\{V(k)V(k')\} = 0, \qquad k, k' \in \{1, \ldots, \lceil N/2 \rceil - 1\}.$$

3.2 Problem formulation

The goal of this paper is to obtain an estimate of the transfer function $G(\Omega)$ and the transient $T(\Omega)$ as continuous functions of Ω, under Assumptions 1-3, given the exact input and measured output N-point DFT spectra $U(k)$ and $Y(k)$, and taking into account properties 4 and 6.

Remark 5 (Available frequency band) *The knowledge of the N-point DFT $U(k)$ at $k \in \{0, 1, \cdots, \lfloor N/2 \rfloor\}$ allows to compute $U(k)$ for $k \in \mathbb{Z}$. (Recall that $U(N - k) = \overline{U(k)}$, and that $U(pN + k) = U(k)$ for $p \in \mathbb{Z}$.) However, due to the periodicity of the N-point DFT, and the non-periodicity of the continuous-time Fourier transform, (27) is invalid for $|k| > N/2$ for the continuous-time case. Thus, the available frequency band is confined to $\{0, 1, \cdots, \lfloor N/2 \rfloor\}$.*

Remark 6 (Frequency band of interest) *Very often, not the whole available frequency band (i.e. $\{0, 1, \cdots, \lfloor N/2 \rfloor\}$) is of interest. For instance, only a single resonance of the system could be investigated, or the excitation could be restricted to low frequencies to avoid alias errors. Then, the data set for the estimation of the transfer function may be confined to those frequencies that are of interest, i.e. a subset of $\{0, 1, \cdots, \lfloor N/2 \rfloor\}$.*

The following section reaches the goal formulated above by modelling $G(\Omega)$ and $T(\Omega)$ as RCGPs (thus taking Property 6 into account), and by estimating them in a Bayesian framework in the frequency band of interest.

4 Bayesian TF and transient regression

The aim of regression in a Bayesian framework by use of Gaussian processes [21, Chapter 2] is formulated as follows. Given observed data points of joint Gaussian processes and *a priori* mean, covariance and relation functions, determine the *a posteriori* mean, covariance and relation functions. This will be applied to the regression of the TF G and the transient T. To this end, consider the following a priori information, consistent with Notation 3.

Assumption 4 (A priori information) *Assume that the TF G and the transient T are RCGPs over $k \in \mathbb{R}$:*

$$G(\Omega_k) \sim \mathcal{RCGP}(0, \alpha_G K, \alpha_G C) \mid \mathbb{K}_{\Re} \tag{33a}$$
$$T(\Omega_k) \sim \mathcal{RCGP}(0, \alpha_T K, \alpha_T C) \mid \mathbb{K}_{\Re} \tag{33b}$$

where $\alpha_G \geq 0$, $\alpha_T \geq 0$, and $\mathbb{K}_{\Re}$ are defined in (32), and K and C are well-defined covariance function and relation function, respectively. Moreover, assume that G and T are independent.

As can be seen in Assumption 4, G and T have structural resemblance, which will be justified in Section 5.3. The specific choice of K and C will be discussed in Section 5.

The MAP estimates $\hat{G}$ and $\hat{T}$ will be constructed in the following three steps:

1) Write the output spectrum as an RCG distribution in Section 4.1,
2) Compute the joint covariances of G and Y, and T and Y at the DFT-frequencies in the frequency band of interest in Section 4.2,
3) Compute the expectations of G and T at the estimation frequencies, conditioned on Y in Section 4.3.

4.1 The output spectrum as a Gaussian distribution

Consistent with Remark 6, denote $\mathrm{k} = \{k_1, k_2, \cdots, k_n\} \subset \{0, \cdots, \lfloor N/2 \rfloor\}$ the DFT-frequency indices that lie in the frequency band of interest. Consistent with Notation 4, denote $G(\Omega_{\mathrm{k}}) = [G(\Omega_{k_1}), \cdots, G(\Omega_{k_n})]^T$. Also $T(\Omega_{\mathrm{k}})$, $K(\Omega_{\mathrm{k}}, \Omega_{\mathrm{k}})$ and $C(\Omega_{\mathrm{k}}, \Omega_{\mathrm{k}})$ are defined similarly. According to Assumption 4 and Definition 2, the a priori distributions of $G(\Omega_{\mathrm{k}})$ and $T(\Omega_{\mathrm{k}})$ are RCG:

$$G(\Omega_{\mathrm{k}}) \sim \mathcal{RCN}(0, \alpha_G \Gamma) \tag{34}$$
$$T(\Omega_{\mathrm{k}}) \sim \mathcal{RCN}(0, \alpha_T \Gamma) \tag{35}$$

where Γ is constructed according to (9). Moreover, Γ can be represented by the use of the covariance and relation

functions K and C, in the following form:

$$\Gamma = \begin{bmatrix} K(\Omega_{k_r}, \Omega_{k_r}) & K(\Omega_{k_r}, \Omega_{k_c}) & \overline{K(\Omega_{k_r}, \Omega_{k_c})} \\ K(\Omega_{k_c}, \Omega_{k_r}) & K(\Omega_{k_c}, \Omega_{k_c}) & C(\Omega_{k_c}, \Omega_{k_c}) \\ \overline{K(\Omega_{k_c}, \Omega_{k_r})} & C(\Omega_{k_c}, \Omega_{k_c})^H & \overline{K(\Omega_{k_c}, \Omega_{k_c})} \end{bmatrix} \quad (36)$$

$$k_r = k \cap \mathbb{K}_{\Re}, \quad k_c = k \setminus \mathbb{K}_{\Re}. \quad (37)$$

In agreement with Definition 1, the cardinalities of k_r and k_c will be denoted by n_r and n_c respectively. The output spectrum $Y(k)$ of the LTI system in (27) is described as an RCG distribution as follows.

Theorem 2 *For G and T Gaussian processes as given in (33), the output spectrum $Y(k)$ is an RCG distributed vector, viz.*

$$Y(k) \sim \mathscr{RCN}(0, \Gamma_Y) \quad (38a)$$

$$\text{with} \quad \Gamma_Y = \alpha_G(\widetilde{U(k)}\widetilde{U(k)}^H) \odot \Gamma + \alpha_T \Gamma + \sigma_v^2 I. \quad (38b)$$

Proof. Follows from the model equations (27), Properties 1 and 2 of RCG distributions, and Assumptions 3 and 4. □

Remark 7 *It should be noted that the transient $T(\Omega_k)$ may not always be present in (27). As can be seen from Lemma 1 and Appendix A, if the input signal $u(t)$ is periodic with period length NT_s, then the transient T will vanish from (27), i.e., $T(\Omega_k) = 0$ in (27). In this case, the estimation problem is simplified and we only need to estimate G.*

The noise standard deviation σ_v is not known a priori. Just like α_G and α_T, it will be handled as a hyperparameter in Section 5.4.

4.2 The joint covariances

Consider that the transfer function G and the transient T should be estimated at the generalised frequency Ω. From (33), $G(\Omega)$ and $T(\Omega)$ are also RCG distributed. Moreover, from (27) and (33), $G(\Omega)$ and $Y(k)$, and $T(\Omega)$ and $Y(k)$ are jointly RCG distributed, respectively. The joint covariances with $Y(k)$ are given by

$$\Gamma_{G^*Y} \equiv \mathbb{E}\left\{\widetilde{G(\Omega)} \ \widetilde{Y(k)}^H\right\} = \Gamma_{G^*G} \text{ diag}\left(\widetilde{U(k)}^H\right) \quad (39a)$$

$$\Gamma_{T^*Y} \equiv \mathbb{E}\left\{\widetilde{T(\Omega)} \ \widetilde{Y(k)}^H\right\} = \Gamma_{T^*T} \quad (39b)$$

where $\Gamma_{G^*G} = \mathbb{E}\left\{\widetilde{G(\Omega)} \ \widetilde{G(\Omega_k)}^H\right\}$ is computed via (33a), $\Gamma_{T^*T} = \mathbb{E}\left\{\widetilde{T(\Omega)} \ \widetilde{T(\Omega_k)}^H\right\}$ is computed via (33b), and $\text{diag}\left(\widetilde{U(k)}^H\right)$ is a diagonal matrix with $\widetilde{U(k)}^H$ being its main diagonal vector.

4.3 Regression and prediction of the TF and the transient

Theorem 3 *The MAP estimates of the TF and the transient are the a posteriori means*

$$\hat{G}(\Omega) \equiv m_{G^*|Y} \qquad \hat{T}(\Omega) \equiv m_{T^*|Y} \quad (40)$$

of the following posterior distribution

$$G(\Omega)|Y(k) \sim \mathscr{RCN}\left(m_{G^*|Y}, \Gamma_{G^*|Y}\right) \quad (41a)$$

$$T(\Omega)|Y(k) \sim \mathscr{RCN}\left(m_{T^*|Y}, \Gamma_{T^*|Y}\right) \quad (41b)$$

$$\text{with} \qquad m_{G^*|Y} = \Gamma_{G^*Y}\Gamma_Y^{-1}\widetilde{Y(k)}, \quad (41c)$$

$$m_{T^*|Y} = \Gamma_{T^*Y}\Gamma_Y^{-1}\widetilde{Y(k)} \quad (41d)$$

$$\Gamma_{G^*|Y} = \Gamma_{G^*G^*} - \Gamma_{G^*Y}\,\Gamma_Y^{-1}\,\Gamma_{G^*Y}^H \quad (41e)$$

$$\Gamma_{T^*|Y} = \Gamma_{T^*T^*} - \Gamma_{T^*Y}\,\Gamma_Y^{-1}\,\Gamma_{T^*Y}^H \quad (41f)$$

with
$$\Gamma_{G^*G^*} = \mathbb{E}\left\{\widetilde{G(\Omega)} \ \widetilde{G(\Omega)}^H\right\}, \ \Gamma_{T^*T^*} = \mathbb{E}\left\{\widetilde{T(\Omega)} \ \widetilde{T(\Omega)}^H\right\}.$$

Proof. The a posteriori distribution follows from (38), (39), and Property 3. □

The estimator given by (40) will be referred to as the GPTF (Gaussian Process Transfer Function) estimator.

Remark 8 *Consistent with Remark 4, k is restricted to the DFT frequencies. Nonetheless, Ω is allowed to be a non-DFT frequency, thus allowing for evaluations of $\hat{G}(\Omega)$ and $\hat{T}(\Omega)$ in-between DFT frequencies.*

5 Kernel construction and model tuning

The design of covariance and relation functions K and C is a central issue for TF estimation with Gaussian process regression. A natural way is to use the duality between the TF and the impulse response function in Property 6, and derive the corresponding covariance and relation functions based on the ones for regularised impulse response estimation in the literature, e.g. [14,12,4].

Specifically, assume that $h(t)$ is a zero mean Gaussian process with covariance function $\text{cov}(h(t), h(s)) = \alpha_G P(t, s)$, where $t, s \in [0, +\infty)$ for continuous-time systems and $t, s = 0, T_s, 2T_s, \cdots$, for discrete-time systems, and $P(t, s)$ is a kernel suitable for impulse response estimation introduced in the literature, e.g., [14,12,4]. Since the Fourier transform is a linear transform of $h(\cdot)$ and preserves the Gaussianity of $h(\cdot)$, $G(\Omega)$ is a Gaussian process as well [10, p. 309], [22], [7, Section 5.4.1]. In particular, and consistent with (17), $G(\Omega)$ is an RCGP with mean function

$$\mathbb{E}\{G(\Omega)\} = \mathscr{F}\{\mathbb{E}\{h(t)\}\} = 0 \quad (42)$$

and covariance and relation functions ($k, l \in \mathbb{R}$):

$$\alpha_G K(\Omega_k, \Omega_l) = \mathbb{E}\left\{ G(\Omega_k)\overline{G(\Omega_l)} \right\} \tag{43}$$

$$\alpha_G C(\Omega_k, \Omega_l) = \mathbb{E}\left\{ G(\Omega_k) G(\Omega_l) \right\} = \alpha_G K(\Omega_k, \Omega_{-l}). \tag{44}$$

Property 7 (Stability) *A sufficient condition on the kernel $K(\Omega_k, \Omega_l)$ to guarantee the stability of the estimated TF in* (40) *is that the transfer function*

$$G_{\text{gen}}(\Omega) = K(\Omega, \Omega_{k_{\text{r}}})r + K(\Omega, \Omega_{k_{\text{c}}})c + K(\Omega, \Omega_{-k_{\text{c}}})\bar{c} \tag{45}$$

is stable for any $c \in \mathbb{C}$, $r, k_{\text{c}} \in \mathbb{R}$, $k_{\text{r}} \in \mathbb{K}_{\Re}$.

Proof. Noticing (43) and (44), (40) can be rewritten as

$$\hat{G}(\Omega) = \left[K(\Omega, \Omega_{k_{\text{r}}})\ K(\Omega, \Omega_{k_{\text{c}}})\ K(\Omega, \Omega_{-k_{\text{c}}}) \right] \gamma \tag{46a}$$

$$\text{with } \gamma = \alpha_G \operatorname{diag}\left(\widetilde{U(\text{k})}^H \right) \Gamma_Y^{-1} \widetilde{Y(\text{k})} \tag{46b}$$

From the particular structure of (46b) it follows that γ can be written as $\gamma = \left[\gamma_{\text{r}}^T\ \gamma_{\text{c}}^T\ \gamma_{\text{c}}^H \right]^T$, where $\gamma_{\text{r}} \in \mathbb{R}^{n_r}$, $\gamma_{\text{c}} \in \mathbb{C}^{n_c}$. Consequently, γ has the same real/complex dimensionality as $\widetilde{Y(\text{k})}$. Expanding (46a) gives

$$\hat{G}(\Omega) = \sum_{i=1}^{n_r} K(\Omega, \Omega_{k_{\text{r}i}})\gamma_{\text{r}i} \dots$$
$$+ \sum_{i=1}^{n_c} \left(K(\Omega, \Omega_{k_{\text{c}i}})\gamma_{\text{c}i} + K(\Omega, \Omega_{-k_{\text{c}i}})\bar{\gamma}_{\text{c}i} \right) \tag{47}$$

with $\gamma_{\text{r}i}$, $\gamma_{\text{c}i}$, $k_{\text{r}i}$ and $k_{\text{c}i}$ the ith elements of γ_{r}, γ_{c}, k_{r} and k_{c} respectively. A generic term of this sum is given by $G_{\text{gen}}(\Omega)$ in (45). Therefore, if $G_{\text{gen}}(\Omega)$ is a stable transfer function, then $\hat{G}(\Omega)$ is stable as well. $\square$

5.1 Kernels for continuous-time systems

For this case, $\Omega = j\omega$ and $G(j\omega)$ is given in (28a). The covariance and relation functions in (43) and (44) take the following form:

$$K(j\omega_k, j\omega_l) = \int_0^\infty \int_0^\infty P(t,s) e^{-j\omega_k t} e^{j\omega_l s}\, \mathrm{d}t\, \mathrm{d}s \tag{48}$$

$$C(j\omega_k, j\omega_l) = K(j\omega_k, -j\omega_l) \tag{49}$$

Due to (49) – consistent with (44) – we will only provide explicit expressions of the covariance function $K(j\omega_k, j\omega_l)$. The Stable Spline (SS) and Diagonal Correlated (DC) kernels in [14,12,4] are given by:

$$P_{\text{SS}}(t,s) = \begin{cases} \frac{e^{-2\beta t}}{2}\left(e^{-\beta s} - \frac{e^{-\beta t}}{3} \right) & t \geq s \geq 0 \\ \frac{e^{-2\beta s}}{2}\left(e^{-\beta t} - \frac{e^{-\beta s}}{3} \right) & 0 \leq t < s \end{cases} \tag{50}$$

$$P_{\text{DC}}(t,s) = e^{-\alpha|t-s|} e^{-\beta(t+s)/2}, \quad t, s \geq 0 \tag{51}$$

where $\alpha, \beta > 0$ are the so-called hyperparameters. The corresponding frequency domain covariance functions are given in the following.

Theorem 4 *The SS and DC kernels for estimating impulse responses in [14,12,4] correspond to the following covariance functions for estimating transfer functions:*

$$K_{\text{SS}}(j\omega_k, j\omega_l) = \frac{1}{2} \cdot \frac{1}{3\beta + j(\omega_k - \omega_l)} \cdots$$
$$\times \left(\frac{1}{2\beta - j\omega_l} + \frac{1}{2\beta + j\omega_k} - \frac{1}{3(3\beta - j\omega_l)} - \frac{1}{3(3\beta + j\omega_k)} \right) \tag{52}$$

$$K_{\text{DC}}(j\omega_k, j\omega_l) = \frac{1}{\beta + j(\omega_k - \omega_l)} \cdots$$
$$\times \left(\frac{1}{\alpha + \beta/2 + j\omega_k} + \frac{1}{\alpha + \beta/2 - j\omega_l} \right) \tag{53}$$

In addition, when using these SS and DC kernels, the MAP estimate of the TF in (40) *is a stable rational function in $j\omega$.*

5.2 Kernels for discrete-time systems

For this case, $\Omega = e^{j\omega T_s}$ and $G(e^{j\omega T_s})$ is given in (28b). The covariance function in (43) takes the following form:

$$K(e^{j\omega_k T_s}, e^{j\omega_l T_s})$$
$$= \sum_{n=0}^\infty \sum_{n'=0}^\infty P(nT_s, n'T_s) e^{-j\omega_k n T_s} e^{j\omega_l n' T_s} \tag{54}$$

Similar to the continuous-time case, we also derive the covariance functions of the SS and DC kernels for estimating transfer functions of discrete-time systems.

Theorem 5 *The SS and DC kernels for estimating impulse responses in [14,12,4] correspond to the following covariance functions for estimating transfer functions:*

$$K_{\text{SS}}(e^{j\omega_k T_s}, e^{j\omega_l T_s}) = \left\{ \frac{1}{2} \frac{1 - e^{-4\beta - j(\omega_k - \omega_l)T_s}}{(1 - e^{-2\beta - j\omega_k T_s})(1 - e^{-2\beta + j\omega_l T_s})} \cdots \right.$$
$$\left. - \frac{1}{6} \frac{1 - e^{-6\beta - j(\omega_k - \omega_l)T_s}}{(1 - e^{-3\beta - j\omega_k T_s})(1 - e^{-3\beta + j\omega_l T_s})} \right\} \frac{1}{1 - e^{-3\beta - j(\omega_k - \omega_l)T_s}} \tag{55}$$

$$K_{\text{DC}}(e^{j\omega_k T_s}, e^{j\omega_l T_s}) = \frac{1}{1 - e^{-\beta - j(\omega_k - \omega_l)T_s}} \cdots$$
$$\times \frac{1 - e^{-2\alpha - \beta - j(\omega_k - \omega_l)T_s}}{(1 - e^{-\alpha - \beta/2 - j\omega_k T_s})(1 - e^{-\alpha - \beta/2 + j\omega_l T_s})} \tag{56}$$

In addition, when using the SS and DC kernels, the MAP estimate of the TF in (40) *is a stable rational function in $e^{-j\omega T_s}$.*

5.3 Kernels for the transient $T(\Omega)$

As has been discussed in e.g., [1,23] and [16, Section 6.3.2], there exists a structural resemblance between

$G(\Omega)$ and $T(\Omega)$. In particular, we have the following lemma useful for the design of kernels for $T(\Omega)$.

Lemma 1 *Consider the system (21). Let $h^*(t)$ denote the response associated with the transient $T(\Omega)$, i.e, $T(\Omega) = \mathscr{F}\{h^*(t)\}$. Then*

$$h^*(t) = \begin{cases} c_T(h * u_d)(t) & \text{for } t \geqslant 0 \\ 0 & \text{for } t < 0 \end{cases} \tag{57}$$

where $c_T = f_s/\sqrt{N}$ for the continuous-time case, $c_T = 1$ for the discrete-time case, and

$$u_d(t) = \begin{cases} u(t) - u(t + NT_s) & \text{for } t < 0 \\ 0 & \text{for } t \geqslant 0. \end{cases} \tag{58}$$

Remark 9 *Lemma 1 reveals the dependence of the transient $T(\Omega_k)$ on the input signal $u(t)$ and on the system's impulse response $h(t)$. However, this dependence involves an unmeasured (and thus unknown) part of the input signal, namely $u(t)$ for $t < 0$. For that reason, and for computational convenience, this dependence is not included as a priori knowledge in Assumption 4.*

The next result follows immediately from Lemma 1.

Theorem 6 *Consider the system (21). Assume that $h(t)$ is a zero mean Gaussian process which is independent of $u_d(t)$ and with covariance function $\text{cov}(h(t), h(s)) = \alpha_G P(t, s)$ where for continuous time systems $t, s \in [0, +\infty)$ and for discrete time systems $t, s = 0, T_s, 2T_s, \cdots$. Then $h^*(t)$ is a zero mean Gaussian process with covariance function*

$$\text{cov}(h^*(t), h^*(s)) = c_T^2 \alpha_G \times \tag{59a}$$
$$\int_0^\infty \int_0^\infty \mathbb{E}\{u_d(-\tau) u_d(-\tau')\} P(t + \tau, s + \tau') \, d\tau' \, d\tau$$

for continuous-time case and

$$\text{cov}(h^*(t), h^*(s)) = c_T^2 \alpha_G \times \tag{59b}$$
$$\sum_{n=1}^\infty \sum_{n'=1}^\infty \mathbb{E}\{u_d(-nT_s) u_d(-n'T_s)\} P(nT_s + t, n'T_s + s)$$

for discrete-time case.

Remark 10 *Equation (59) shows that the covariance function of $h^*(t)$ depends on that of $h(t)$ and on the properties of the input $u(t)$. This finding immediately translates to the frequency domain and is valid for the covariance functions of $T(\Omega)$ and $G(\Omega)$, and the property of $U(k)$. In (59), the covariance of $h^*(t)$ (equivalently the covariance of $T(\Omega)$) may be hard to compute. From a computational point of view, it is convenient to notice the following case. If $u_d(t)$ is white noise with mean zero and variance σ_u^2, and $P(t, s)$ is the SS or the DC kernel, then it is easy to verify that the*

covariance function of $h^*(t)$ is the same as that of $h(t)$ (up to a scaling factor), i.e.,

$$\text{cov}(h^*(t), h^*(s)) = c_T^2 \alpha_G \begin{cases} \sigma_u^2 P_{\text{DC}}(t, s)/c_{\text{DC}}(\beta) & \text{(DC)} \\ \sigma_u^2 P_{\text{SS}}(t, s)/c_{\text{SS}}(\beta) & \text{(SS)} \end{cases} \tag{60}$$

where for continuous time system, $c_{\text{DC}}(\beta) = \beta$, $c_{\text{SS}}(\beta) = 3\beta$, and for discrete-time system $c_{\text{DC}}(\beta) = \frac{e^{-\beta}}{1-e^{-\beta}}$, $c_{\text{SS}}(\beta) = \frac{e^{-3\beta}}{1-e^{-3\beta}}$. Due to this observation, we assume for computational convenience in the paper that $T(\Omega)$ and $G(\Omega)$ have the same kind of covariance function but with independent scaling hyperparameters α_G and α_T in (33). The simulation in Section 7 shows that this assumption leads to satisfying results for tested examples.

5.4 Hyperparameter tuning

The kernels used in Sections 5.1 and 5.2 (and any other commonly used kernels) depend on some tuneable parameters, also called hyperparameters. In the DC kernel (51), the hyperparameters are α and β. In addition, the weights α_G and α_T of the kernels, and the noise variance σ^2 in the output covariance Γ_Y in (38) must be determined. The quality of $\hat{G}(\Omega)$ and $\hat{T}(\Omega)$ depends on these hyperparameters, such that an appropriate criterion must be devised to tune them.

Denote by θ the vector of hyperparameters (e.g. for the DC kernel: $\theta^T = [\alpha \ \beta \ \alpha_G \ \alpha_T \ \sigma^2]$). In this paper, tuning θ is done via the maximisation of the log marginal likelihood $\log p(\widetilde{Y(k)}|\theta)$ [21, Section 5.4.1] of the output spectrum, which is computed as:

$$\log p\left(\widetilde{Y(k)}\Big|\theta\right) = -\frac{1}{2}\widetilde{Y(k)}^H \Gamma_Y^{-1}(\theta)\widetilde{Y(k)} \cdots$$
$$-\frac{1}{2}\log|\Gamma_Y(\theta)| - \frac{n_r}{2}\log 2\pi - n_c \log \pi. \tag{61}$$

Equation (61) expresses the *likelihood* that the observed output spectrum has been generated by the Gaussian process determined by the covariance function with the given hyperparameters θ. The term *marginal* stems from the fact that the estimates $\hat{G}(\Omega)$ and $\hat{T}(\Omega)$ are not appearing in the expression because they have been marginalised out of it. Maximising the marginal likelihood has been shown to automatically provide a trade-off between the uncertainty of the estimates and their complexity. Alternatively to *marginal likelihood* (which is often used in a Bayesian framework), (61) may be called the *evidence* or the *empirical Bayes*, depending on the context. Other criteria for hyperparameter tuning exist, e.g. leave-one-out cross-validation [21, Section 5.4.2].

The maximization of (61) can be done via a gradient based approach, where the partial derivative w.r.t. the ith

hyperparameter is given by

$$\frac{\partial}{\partial \theta_i} \log p\left(\widetilde{Y(\mathbf{k})}|\theta\right) = \frac{1}{2}\text{tr}\left(\left(\mathring{\alpha}\mathring{\alpha}^H - \Gamma_Y^{-1}\right)\frac{\partial \Gamma_Y}{\partial \theta_i}\right) \quad (62\text{a})$$

$$\text{with} \quad \mathring{\alpha} = \Gamma_Y^{-1}\widetilde{Y(\mathbf{k})} \quad (62\text{b})$$

Remark 11 *The hyperparameters determine to a great extent the model complexity of the estimate. Their optimisation is highly related to model order selection in the context of the parametric identification of a system model. The most important conceptual difference is that hyperparameter tuning is a* continuous *optimisation, whereas model order selection is* discrete*, and prone to be cursed by dimensionality.*

6 Equivalence with impulse response estimation by Gaussian process regression

To investigate the equivalence, consider causal LTI systems described by (21). The impulse response estimation problem based on a collection of measured output samples $y(t), t = 0, \cdots, (N-1)T_s$ and input samples $u(t)$ was studied in [14,12,4]. The applied method can be interpreted as Gaussian process regression. Assume that $v(t)$ in (21) is zero mean white Gaussian distributed with variance σ^2 and $h(t)$ is a zero mean Gaussian process, independent of $v(t)$, with covariance function $\text{cov}(h(t), h(s)) = \alpha_G P(t,s)$ where $t, s \in [0, \infty)$ and $t, s = 0, T_s, 2T_s, \cdots$, for continuous and discrete time systems, respectively, and $P(t,s)$ can be the SS or DC kernel.

Since $(h * u)(t)$ is a linear transformation of $h(\cdot)$ and preserves Gaussianity of $h(\cdot)$ for given $u(\cdot)$, $y_\circ(t)$ is a zero mean Gaussian process with covariance function

$$\begin{aligned}\Gamma_{y_\circ}(t,s) &= \text{cov}(y_\circ(t), y_\circ(s)) = (\Gamma_{hy_\circ}(\cdot,s) * u)(t), \\ \Gamma_{hy_\circ}(t,s) &= \text{cov}(h(t), y_\circ(s)) = \alpha_G(P(t,\cdot) * u)(s).\end{aligned} \quad (63)$$

Since $y_\circ(t)$ and $v(t)$ are Gaussian and independent of each other, $y(t)$ and $h(t)$ are jointly Gaussian distributed and the conditional distribution $h(t)|y(0)\cdots, y((N-1)T_s)$ is thus Gaussian. In [14,12,4], the estimate of the impulse response $\hat{h}(t)$ is the conditional mean $\mathbb{E}\{h(t)|y(0)\cdots, y((N-1)T_s)\}$:

$$\hat{h}(t) = \Gamma_{hy_\circ}(t, \mathbf{t})\left(\Gamma_{y_\circ}(\mathbf{t}, \mathbf{t}) + \sigma^2 I_N\right)^{-1} y(\mathbf{t}) \quad (64)$$

where $\mathbf{t} = \{0, T_s, \cdots, (N-1)T_s\}$. Then, the following result holds.

Theorem 7 *Consider the system (21). The impulse response estimate (64) is equivalent to the transfer function estimate in (40), i.e., $\mathscr{F}\{\hat{h}(t)\} = \hat{G}(\Omega)$, if*

(A1) for the continuous-time system (21), $u(t)$ is band-limited and periodic with fundamental period

$T_0 = NT_s$, *and the sampling period T_s is chosen such that the Shannon-Nyquist condition is satisfied, and the frequency band of interest consists of the DFT frequencies $\mathbf{k} = \{0, 1, \ldots, \lfloor N/2 \rfloor\}$;*

(A2) for discrete-time system (21), $u(t)$ is periodic with fundamental period $T_0 = NT_s$, and the frequency band of interest consists of the DFT frequencies $\mathbf{k} = \{0, 1, \ldots, \lfloor N/2 \rfloor\}$.

7 Comparison with other TF estimators

This section further compares the GPTF with the Regularised Finite Impulse Response (RFIR) estimation [14,12,4], the Local Polynomial Method (LPM) [19,20,24], and the Local Rational Method (LRM) [5,8] on simulation results. The following preliminary observations are made.

- In contrast to the RFIR, the implementation of the GPTF estimator is straightforwardly compatible for both discrete time and continuous time systems, as was highlighted in Section 5.1, whereas the RFIR is restricted to discrete time systems in practice.
- The LPM and the LRM estimate the TF and the transient in (27a) by modelling them as polynomials and rational functions respectively in local frequency windows (i.e. the polynomial/rational estimation is performed in a sliding frequency band). An important distinction between the kernel based methods (GPTF and RFIR) and the local methods (LPM and LRM) is that the latter provide estimates *at the DFT bins $\Omega_k, k \in \mathbb{Z}$ only*, whereas the GPTF and RFIR provide estimates which are *continuous* in Ω. As a consequence, the stability of the estimates from the LRM and the LPM cannot be determined.
- The computational load of the LPM and the LRM is $\mathbb{O}(N)$, whereas that of GPTF and RFIR is $\mathbb{O}(N^3)$ (a square matrix of size $N \times N$ must be inverted via Cholesky factorization), if no approximation is made. See [21, Chapter 8] for approximations.

7.1 Simulation setup

Simulations on continuous-time systems are considered. The sampling period is $T_s = 1\,\text{s}$. The input signal is periodic random noise, and is generated as a sum of sines with random amplitudes and phases, and such that the highest frequency of the sines is smaller than the Nyquist frequency. This ensures that the input and output signals satisfy the Shannon-Nyquist theorem. The period length of the input signal is made longer than NT_s, such that a transient is present. For Sections 7.2 and 7.3, the systems were simulated with the Matlab command ode45, while for Section 7.4 a frequency domain steady-state simulation is performed, which is then truncated in the time such that a transient is present.

The considered estimators are listed in Table 1 (not all estimators are considered in each case study). Monte Carlo simulations (100 runs) are performed. Each Monte Carlo run consists of randomising the excitation signal $u(t)$ and the disturbing noise $v(t)$. The figures will show the magnitude of the estimated TF from a single Monte Carlo run, and the Mean Squared Error (MSE) at each frequency, *viz.*:

$$\text{MSE}(\Omega) = \frac{1}{100} \sum_{m=1}^{100} |\hat{G}_m(\Omega) - G(\Omega)|^2, \qquad (65)$$

with $\hat{G}_m(\Omega)$ the estimated TF at the frequency Ω from the mth Monte Carlo run. The frequency axes of the figures are labeled ω, with ω and Ω related as in Notation 5. For the LPM and the LRM, ω is evaluated at the DFT bins, giving estimates which are discrete in the frequency. It is evaluated at a denser grid for the GPTF estimates.

For each Monte Carlo run, the hyperparameters of the GPTF and RFIR are optimised, as explained in Section 5.4. For the LPM and LRM estimators, two structural parameters must be chosen: the orders of the local models (polynomials for LPM and rational functions for LRM), and the local bandwidth. For LPM, the two parameters were chosen as the minimisers of the true MSE via a grid based optimisation. Since the true MSE of an estimate is not known in practice, this yields optimistic results for the LPM estimator. For LRM, the two parameters suggested in [8] were used. A complete discussion on the parameter selection for LPM and LRM estimators is outside the scope of this paper. The implementation of the LPM has been obtained from the authors of [16], and that of the LRM from the authors of [8] and [5].

Remark 12 *The simulation results in the subsections below compare the TF estimates based on their MSEs. Other criteria for comparison may be important depending on the application at hand, as for instance the correlation length of the resulting estimate, and the quality of the estimated noise model [17]. Note that the LPM is the only discussed estimator to produce a non-parametric noise model, concurrently with the TF estimate [19] (not shown in this paper).*

7.2 Limited frequency band

In Fig. 2, estimates are shown of the TF of a resonating second order continuous time system. The resonating pole, located at $-5 \times 10^{-3} \pm 0.31j$ had a damping coefficient of 1.6%. White Gaussian disturbing noise was added to the output, s.t. the Signal to Noise Ratio (SNR) was 20 dB. Estimates of the GPTF with DC kernel (●), the LPM (▲▲), and the LRM (++) are shown, and the black line is the true TF. The RFIR is not included because an implementation that allows the estimation in a limited frequency band is not available. The training samples for the GPTF were lying between the two grey vertical lines, that is, in

a limited frequency band. This way, 65 frequency domain samples were used. For the LPM, a 6th order polynomial was used, and the local bandwidth was 15 bins. For the LRM, as suggested in [8], rational functions of order 2/2 were used, and the local bandwidth was 17 bins.

One observes that GPTF captures the resonance peak better than LPM but worse than LRM. In this case, LRM gives the lowest MSE, which is not surprising since the true model is, in fact, a rational function of order 2.
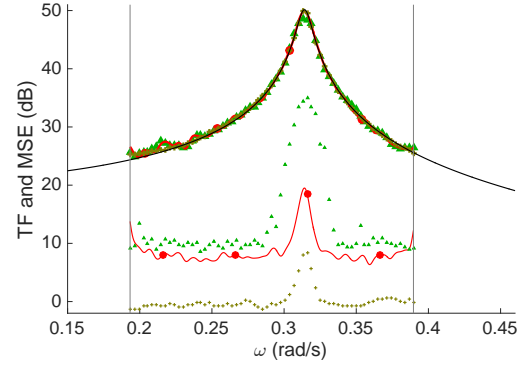

Fig. 2. Comparing GPTF with LPM and LRM in a limited frequency band on a resonating system. Thick lines and markers: estimates. Thin lines and markers: mean squared errors. Marker conventions are given in Table 1. The gray vertical lines are the upper and lower bounds of the considered frequency band for the GPTF estimator.

7.3 Low-pass excitation

In Fig. 3, the results are presented on a 4th order continuous time system, with poles located at $-0.0513 \pm 0.3142j$ and $-0.0101 \pm 2.5133j$. The excited frequency band is upper bounded to 0.94 rad/s. Only one of the two complex poles lies inside the excited frequency band. No disturbing noise was added (i.e. $V(k) = 0$). 78 frequency domain samples were present in the excited frequency band, 178 in the not excited band. For the LPM, a 10th order polynomial was used, and the local bandwidth was 23 bins. For the LRM, rational functions of order 2/2 were used, and the local bandwidth was 17 bins, as suggested in [8].

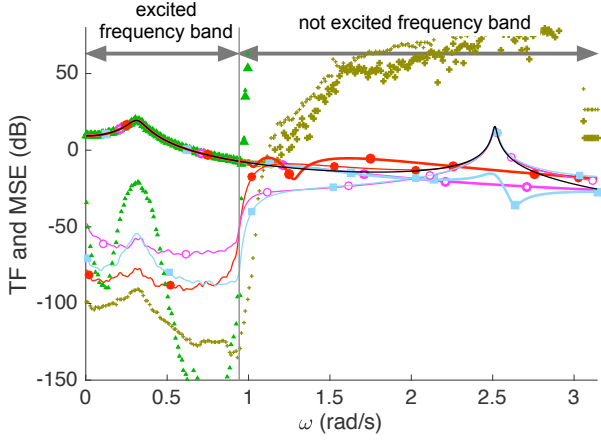Inside the *excited frequency band*, one observes that:

Fig. 3. Non-persistent excitation. Thick lines and markers: estimates. Thin lines and markers: MSE. Marker conventions in Table 1. The gray vertical line is the upper bound of the considered frequency band for the GPTF estimators.

- the advantage of the fact that the LPM is a local method is visible: the fairly high MSE at the resonance frequency does not prevent it to reach a very low MSE at other frequencies. Moreover, GPTF (—●— and —○—) and RFIR (—■—) perform better than the LPM (▲▲) but worse than the LRM (++). In this case, the LRM gives the lowest MSE, which is not surprising since the true model can locally be approximated very well by a rational function of order 2.
- GPTF with DC kernel (—●—) performs better than RFIR with DC kernel (—■—), because the estimation of the first was confined to the excited frequency band, while the RFIR took the full frequency band into account. In addition, GPTF with SS kernel (—○—) performs worse than GPTF with DC kernel.

Inside the *non-excited frequency band*, none of the estimates is able to capture the second resonance. Besides, one observes that the local methods (LPM and LRM) and Gaussian process regression methods (GPTF and RFIR) behave very differently, due to the following.

- For Gaussian process regression methods, the extrapolation in the non-excited frequency band depends on both the data in the excited frequency band and the smoothness and stability assumption embedded in the kernel used. For this case, the resulting estimate in the non-excited band has the correct order of magnitude.
- In the non-excited band, the local methods do not have access to the excited band. Therefore, their estimates in the non-excited band are based on data from a non-persistent excitation, yielding very poor results.

7.4 High order system, large and small SNR

The GPTF with DC kernel and the LRM are applied to a 30th order system, for multiple values of the SNR:

0 dB, 20 dB and 50 dB. The poles and zeroes are plotted in Fig. 4, right. The least damped pole, given by $-0.075 \pm 0.64j$, had a damping coefficient of 11.6%. For the LRM, rational functions of order 2/2 were used, and the local bandwidth was 17 bins, as suggested in [8]. The considered frequency band for the GPTF (0-2.44 rad/s) contains 200 frequency domain samples.

The results are plotted in Fig. 4, left. One observes that:

- for the cases with the SNR equal to 0 and 20 dB, the GPTF (—●—) gives the best results overall. This is thanks to the optimized tuning of the bias-variance trade-off, see Section 5.4. Note that an increased error of the GPTF is observed in the vicinity of the resonance peak of the least damped pole, around 0.64 rad/s. This is not the case for the LRM (++).
- for the case with 50 dB SNR, the LRM (++) and the GPTF (—●—) have virtually equal MSEs inside the considered band.

7.5 Averaged MSE for estimates over all frequencies

The averaged MSE for the estimates over all the frequencies in the considered frequency bands (delimited by grey
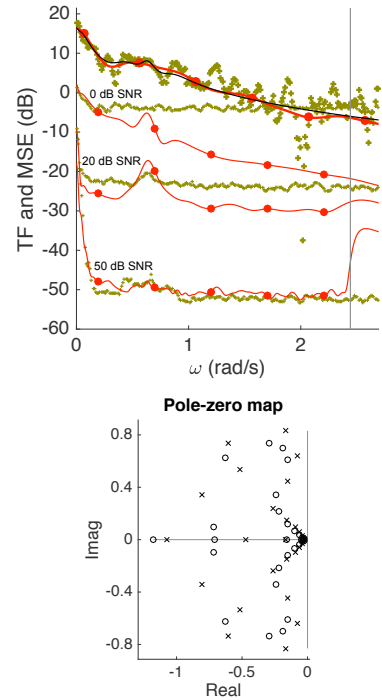

Fig. 4. Left: Results of the GPTF with DC kernel and the LRM on a 30th order system, for multiple SNR cases. The thick lines and markers are estimates for the 0 dB SNR case, the thin lines and markers are the MSEs averaged over 100 realizations, for the SNR cases as tagged. Marker conventions in Table 1. The gray vertical line is the upper bound of the considered frequency band for the GPTF estimator.
Right: Poles (x) and zeroes (o) of the considered system.

vertical lines in the plots), is summarised in Table 2.

Table 2
Averaged MSE of all estimates (in dB). For Fig. 4, the SNR is given between brackets.

Fig nr	2	3	4 (0 dB)	4 (20 dB)	4 (50 dB)
GPTF (DC)	9.919	-72.33	-9.202	-24.76	-39.65
GPTF (TC)	10.05	-71.1	-10.14	-25.13	-37.23
GPTF (SS)	37.94	-60.03	-12.38	-24.41	-39.93
RFIR (DC)	-	-65.78	-10.1	-22.45	-32.93
LPM	23.76	-32.18	-5.529	-20.03	-33
LRM	0.794	-98.21	-3.644	-22.33	-30.59

This table also includes results of the GPTF with the popular TC (Tuned-Correlated) kernel [4] (obtained by letting $\alpha = \beta/2$ in the DC kernel in (51)). These results were very close to that of the GPTF with DC kernel and thus were not included in the previous figures.

One observes quite similar results as before: LPM does not work as well as LRM and GPTF; LRM gives the smallest averaged MSE for Fig. 2 and Fig. 3; GPTF gives the smallest averaged MSE for different choices of SNR for Fig. 4.

Finally, we stress that there is room to improve for both LRM and the GPTF. For LRM, the current MSE could be further lowered if more suitable orders of the rational model and local bandwidths are selected. For GPTF, instead of DC and SS kernels, which can better capture the exponentially decaying behavior, kernels that can better capture the resonant behavior could be used.

8 Conclusions

The estimation of the Transfer Function (TF) of a causal LTI system from known input and measured output DFT spectra has been formulated as a Gaussian process regression problem. To that end, the multivariate real/complex Gaussian (RCG) distribution was introduced, where the stochastic variables were predefined to be either real or complex. Properties of the RCG distribution, which are required for this Bayesian regression, have been derived, including the conditional expectation.

The construction of kernels, required for Gaussian process regression, has been formulated in the frequency domain. Time domain kernels from the literature, that have been used for kernel-based impulse response estimation, were translated to the frequency domain for continuous- and discrete-time systems, and the resulting estimates were proven to be stable rational functions in the appropriate frequency variables. Whereas the current formulation of the method can only handle stable systems, the extension to unstable systems is foreseen for future work.

The equivalence was shown between the TF estimation via Gaussian processes (GPTF) and the time domain Regularised Finite Impulse Response (RFIR) estimation under suitable conditions. Compared with the RFIR, the GPTF has the additional advantage that the estimation can be restricted to a limited frequency band, and that its implementation for continuous-time systems is more appealing. Compared with the Local Polynomial Method (LPM), it was observed that the prior knowledge – introduced via the DC kernel – of the kernel based methods (GPTF and the RFIR), led to a more accurate estimate of stable LTI transfer functions. Also, these kernel based methods take advantage of a Bayesian tuning algorithm for bias-variance trade-off. From a comparison with the Local Rational Method (LRM), it is clear that there is room for improvement for the design of kernels for the estimation of systems which exhibit lowly damped resonances.

In the perspective of future work, formulating the estimation of TFs with Gaussian processes in the frequency domain opens up the possibility of introducing prior knowledge which is more easily formulated in the frequency domain (like phase margin, resonance frequencies and damping, etc.).

Appendix

A Proof of (27a)

Equation (27a), and a closed form expression for $T(\Omega)$ will be derived for the *continuous-time* case, starting from (21a). Denote $Y_{\circ[0,NT_s]}(j\omega)$ the continuous-time Fourier transform of $y_\circ(t)$ in the rectangular window $t \in [0, NT_s]$:

$$
\begin{aligned}
Y_{\circ[0,NT_s]}(j\omega) &= \int_0^{NT_s} \int_0^\infty h(\tau)u(t-\tau)\,d\tau e^{-j\omega t}\,dt \\
&= \int_0^\infty h(\tau)e^{-j\omega\tau} \int_0^{NT_s} u(t)e^{-j\omega t}\,dt\,d\tau + T'(j\omega) \\
&= G(j\omega)U_{[0,NT_s]}(j\omega) + T'(j\omega)
\end{aligned}
\tag{A.1}
$$

$$
T'(j\omega) = \int_0^\infty h(\tau)e^{-j\omega\tau} \cdots
\tag{A.2}
$$
$$
\times \left[\int_{-\tau}^0 u(t)e^{-j\omega t} - \int_{NT_s-\tau}^{NT_s} u(t)e^{-j\omega t} \right] dt\,d\tau
$$

By Property 5 and under Assumption 2, (A.1) is equivalent to (27a) when evaluated at the DFT frequencies ω_k, $k = 0, 1, \ldots, \lfloor N/2 \rfloor$, with

$$
T(j\omega_k) = T'(j\omega_k)f_s/\sqrt{N}.
\tag{A.3}
$$

13

B Proof of Lemma 1

First, recall the definition of $u_d(t)$ in (58). Then note that $e^{-j\omega_k NT_s} = 1$, and compute $T(j\omega_k)$ via (A.3) and (A.2):

$$
\begin{aligned}
T(j\omega_k) &= c_T \int_0^\infty h(\tau) e^{-j\omega_k \tau} \int_{-\tau}^0 u_d(t) e^{-j\omega_k t} \, dt \, d\tau \\
&= c_T \int_0^\infty \int_0^\tau h(\tau) u_d(t-\tau) e^{-j\omega_k t} \, dt \, d\tau \\
&= c_T \int_0^\infty \int_t^\infty h(\tau) u_d(t-\tau) \, d\tau e^{-j\omega_k t} \, dt \quad \text{(B.1)} \\
&= \int_0^\infty h_*(t) e^{-j\omega_k t} \, dt \quad \text{(B.2)}
\end{aligned}
$$

This proves the structural resemblance between $G(j\omega_k)$ and $T(j\omega_k)$. The proof for the discrete-time is analogous, and yields

$$
\begin{aligned}
T(e^{j\omega_k T_s}) &= \sqrt{N}^{-1} \sum_{t=0}^\infty h^*(t) e^{\frac{-j2\pi k t}{N}} \\
h^*(t) &= \sum_{n=t+1}^\infty h(nT_s) u_d((t-n)T_s).
\end{aligned} \quad \text{(B.3)}
$$

C Proof of Theorems 4 and 5

Since it is straightforward to get the representation of the SS and DC kernels in the frequency domain for estimating the transfer function, we only give the proof regarding the stability issue below.

C.1 DC kernel

For the continuous time we have that $\Omega = j\omega$ and $\Omega_{k_c} = jv$. Applying (45) to (53) yields

$$
\begin{aligned}
&G_{\text{gen}}(j\omega) = \cdots \\
&\frac{1}{\beta + j(\omega - v)} \cdot \left(\frac{1}{\alpha + \beta/2 + j\omega} + \frac{1}{\alpha + \beta/2 - jv} \right) c \cdots \\
&+ \frac{1}{\beta + j(\omega + v)} \cdot \left(\frac{1}{\alpha + \beta/2 + j\omega} + \frac{1}{\alpha + \beta/2 + jv} \right) \bar{c}
\end{aligned}
$$

Exchanging $j\omega$ for the Laplace variable s and working out yields:

$$
\begin{aligned}
G_{\text{gen}}(s) &= 2 \frac{(\beta + s)\Re c - v\Im c}{(\alpha + \beta/2 + s) D(s,v)} \cdots \\
&+ \frac{1}{(\alpha + \beta/2)^2 + v^2} \frac{B_{\text{DC}}(s,v)}{D(s,v)} \quad \text{(C.1)}
\end{aligned}
$$

which is a rational function in s, with

$$
\begin{aligned}
D(s,v) &= s^2 + 2\beta s + (\beta^2 + v^2) \quad \text{(C.2)} \\
B_{\text{DC}}(s,v) &= 2(\beta + s)\Re[c(\alpha + \beta/2 + jv)] \cdots \\
&\quad - 2v\Im[c(\alpha + \beta/2 + jv)] \quad \text{(C.3)}
\end{aligned}
$$

For $\alpha, \beta > 0$, all the poles of $G_{\text{gen}}(s)$ lie in the left half plane, such that $G_{\text{gen}}(s)$ is stable. The term $K(\Omega, \Omega_{k_r})r$ in (45) was not included, but is a special case of the above and, thus, is a stable rational function as well.

For the discrete time we have that $\Omega = e^{j\omega T_s}$. Denote $z = e^{j\omega T_s}$, $v = e^{j\omega_{k_c} T_s}$, $B = e^{-\beta}$ and $A = e^{-\alpha - \beta/2}$. Applying (45) to (56) yields

$$
\begin{aligned}
G_{\text{gen}}(z^{-1}) &= \frac{1}{(1 - Az^{-1})(1 - Av)(1 - Av^{-1})} \cdots \\
&\times \frac{C(z^{-1}, v)}{(1 - Bz^{-1}(v + v^{-1}) + B^2 z^{-2})} \quad \text{(C.4)}
\end{aligned}
$$

which is a rational function in z^{-1}. ($C(z^{-1}, v)$ is a polynomial in z^{-1}). The roots of $G_{\text{gen}}(z^{-1})$ are A and $Be^{\pm j\omega_{k_c}}$. Since $0 < A, B < 1$, these roots lie inside the unit circle, yielding a stable estimate.

C.2 SS kernel

Similarly as for the DC kernel, equation (45) in Property 7 is computed for the continuous time SS kernel (52), giving the following rational function in the Laplace variable s:

$$
\begin{aligned}
G_{\text{gen}}(s) &= \frac{1}{2D} \left(\frac{[(6\beta + 2s)\Re c - 2v\Im c](7\beta + 2s)}{3(2\beta + s)(3\beta + s)} \cdots \right. \\
&\left. + (6\beta + 2s)\Re A - 2v\Im A \right) \quad \text{(C.5)}
\end{aligned}
$$

with

$$
\begin{aligned}
D &= s^2 + 6\beta s + 9\beta^2 + v^2 = (s + 3\beta + jv)(s + 3\beta - jv) \\
A &= c \frac{7\beta - 2jv}{3(2\beta - jv)(3\beta - jv)} \quad \text{(C.6)}
\end{aligned}
$$

For $\beta > 0$, all the poles of $G_{\text{gen}}(s)$ lie in the left half plane, such that $G_{\text{gen}}(s)$ is stable.

For the discrete time SS kernel, again the same reasoning can be applied (plug (55) into (45)), showing that the poles of the rational function $G_{\text{gen}}(z^{-1})$ are $e^{-2\beta}$, $e^{-3\beta}$ and $e^{-3\beta \pm j\omega_{k_c} T_s}$, all of which are inside the unit circle, yielding stable estimates.

D Proof of Theorem 7

Before proceeding to the proof, some facts of real-valued continuous-time or discrete-time signal $u(t)$ are summarised. For convenience, we specify the N-point DFT of $\{u(nT_s)\}_{n=0}^{N-1}$ as $\{U(k) = \sum_{n=0}^{N-1} u(nT_s) e^{\frac{-j2\pi n k}{N}} / \sqrt{N}\}_{k=0}^{N-1}$ and the N-point unitary DFT matrix $F(n+1, k+1) = e^{\frac{-j2\pi nk}{N}} / \sqrt{N}$, $n, k = 0, \cdots, N - 1$.

Lemma 2 *Consider a real-valued continuous-time or discrete-time signal $u(t)$. Denote the N-point DFT of $\{u(nT_s)\}_{n=0}^{N-1}$ by $\{U(k)\}_{k=0}^{N-1}$. Then the next results hold:*

(1) There exists an N-dimensional sorting matrix S such that S is orthogonal and $\tilde{U} = SFu(t)$, where $\tilde{U}$ is the augmented vector of $U = \begin{bmatrix} U(0) & \cdots & U(\lfloor N/2 \rfloor) \end{bmatrix}^T$.

(2) for continuous-time $u(t)$, if assumption (A1) of Theorem 7 is satisfied, the N-point DFT of $\{u(nT_s - \tau)\}_{n=0}^{N-1}$ with $\tau \in \mathbb{R}$ and $\tau \geq 0$ is

$$\begin{cases} e^{-j\frac{2\pi k}{NT_s}\tau}U(k), & k = 0, \cdots, \lfloor N/2 \rfloor \\ e^{j\frac{2\pi(N-k)}{NT_s}\tau}U(k), & k = \lfloor N/2 \rfloor + 1, \cdots, N-1 \end{cases}$$

(3) for discrete-time $u(t)$, if assumption (A2) of Theorem 7 is satisfied, the N-point DFT of $\{u((n-\tau)T_s)\}_{n=0}^{N-1}$ with $\tau \in \mathbb{Z}$ and $\tau \geq 0$ is $\{e^{-j\frac{2\pi k}{N}\tau}U(k)\}_{k=0}^{N-1}$.

Proof. Part (1) follows by the definition of $\tilde{U}$ and $U(k) = \overline{U(N-k)}$, $k = \lfloor N/2 \rfloor + 1, \cdots, N-1$. Part (2) follows by the shift property of continuous-time Fourier series, and the definition and properties of DFT of $u(nT_s)$ and $u(nT_s - \tau)$ under assumption (A1) of Theorem 7. Part (3) follows from the shift property of DFT. $\square$

Now we prove Theorem 7. First, due to the assumptions (A1) and (A2), the transient $T(\Omega_k)$ will vanish from (27), i.e., $T(\Omega_k) = 0$ in (27). In this case, α_T should be set to 0 when computing $\hat{G}(\Omega)$ in (40).

Noting (64) and using Part (1) of Lemma 2, we have

$$\mathscr{F}\{\hat{h}(t)\} = \mathscr{F}\{\Gamma_{hy_\circ}(t,t)\}F^{-1} \cdots \\ \times S^T(SF\Gamma_{y_\circ}(t,t)F^{-1}S^T + \sigma^2 I_N)^{-1}SFy(t)$$

where $\mathscr{F}\{\Gamma_{hy_\circ}(t,t)\}$ shall be understood componentwise. Noting $SFy(t) = \tilde{Y}$ and comparing with (40), it is enough to prove

$$\Gamma_{G^*Y} = \mathscr{F}\{\Gamma_{hy_\circ}(t,t)\}F^{-1}S^T \tag{D.1}$$
$$\Gamma_Y = SF\Gamma_{y_\circ}(t,t)F^{-1}S^T + \sigma^2 I_N \tag{D.2}$$

First consider (D.1) for the continuous-time case. Noting Part (2) of Lemma 2 and $U(-k) = \overline{U(k)}$ yields

$$\Gamma_{hy_\circ}(t,t)F^{-1} \\ = \int_0^\infty \alpha_G P(t,\tau)[u(-\tau) \cdots u((N-1)T_s - \tau)]F^{-1}\,\mathrm{d}\tau \\ = \int_0^\infty \alpha_G P(t,\tau)[1 \cdot \overline{U(0)} \cdots e^{j\frac{2\pi\lfloor N/2 \rfloor}{NT_s}\tau}\overline{U(\lfloor N/2 \rfloor)} \\ e^{-j\frac{2\pi(\lceil N/2 \rceil - 1t)}{NT_s}\tau}\overline{U(\lfloor N/2 \rfloor + 1)} \cdots e^{-j\frac{2\pi}{NT_s}\tau}\overline{U(N-1)}]\,\mathrm{d}\tau$$

Noting this equation and $\overline{U(N-k)} = U(k)$ yields

$$\mathscr{F}\{\Gamma_{hy_\circ}(t,t)\}F^{-1}S^T \\ = \alpha_G[K(\Omega, 0)\overline{U(0)} \cdots K(\Omega, j\frac{2\pi\lfloor N/2 \rfloor}{NT_s})\overline{U(\lfloor N/2 \rfloor)} \cdots \\ K(\Omega, -j\frac{2\pi(\lceil N/2 \rceil - 1)}{NT_s})U(\lceil N/2 \rceil - 1) \cdots \\ \cdots K(\Omega, -j\frac{2\pi}{NT_s})U(1)]S^T$$

Then (D.1) follows from $\alpha_G K(\Omega, j\frac{2\pi k}{NT_s}) = \mathbb{E}\left\{G(\Omega)\overline{G(j\frac{2\pi k}{NT_s})}\right\}$, $k = -\lfloor N/2 \rfloor + 1, \cdots, \lfloor N/2 \rfloor$, and the property of the sorting matrix S. It is apparent that the discrete-time case for (D.1) can be derived in a similar way by replacing integrals with series in the computation of convolutions and noting Part (3) of Lemma 2.

As for (D.2), it can be proved in a similar but more tedious way as (D.1). The details are omitted.

Acknowledgement

References

[1] J. Antoni and J. Schoukens. A comprehensive study of the bias and variance of frequency-response-function measurements: Optimal window selection and overlapping strategies. *Automatica*, 43(10):1723 − 1736, 2007.

[2] J.S. Bendat and A. G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons, Inc., New York, NY, 2 edition, 1993.

[3] E. Oran Brigham. *The Fast Fourier Transform and Its Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[4] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and Gaussian processes – revisited. *Automatica*, 48(8):1525 − 1535, 2012.

[5] E. Geerardyn, T. A. E. Oomen, and J. Schoukens. Enhancing $\mathscr{H}_\infty$ norm estimation using local LPM/LRM modeling: Applied to an AVIS. In *Proc. of the 19th IFAC World Congress*, pages 10856–10861, Cape Town, South Africa, 24-29 Aug 2014 2014.

[6] L. Ljung. On the estimation of transfer functions. *Automatica*, 21(6):677 − 696, 1985.

[7] D. J. C. MacKay. Introduction to gaussian processes. In C. M. Bishop, editor, *Neural Networks and Machine Learning*. Springer-Verlag, 1998.

[8] T. McKelvey and Guérin. Non-parametric frequency response estimation using a local rational model. In *proc. of the 16th IFAC Symposium on System Identification*, pages 49–54, Brussels, Belgium, July 2012.

[9] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals & Systems*. Prentice-Hall International, inc., New Jersey, 2nd edition, 1983.

[10] A Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Education, 2002.

[11] B. Picinbono. Second-order complex random vectors and normal distributions. *Signal Processing, IEEE Transactions on*, 44(10):2637–2640, 1996.

[12] G. Pillonetto, A. Chiuso, and G. De Nicolao. Prediction error identification of linear systems: A nonparametric gaussian regression approach. *Automatica*, 47(2):291 – 305, 2011.

[13] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657 – 682, 2014.

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

[15] R. Pintelon and J. Schoukens. Identification of continuous-time systems using arbitrary signals. *Automatica*, 33(5):991 – 994, 1997.

[16] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. John Wiley, 2nd edition, Mar 2012.

[17] R. Pintelon and J. Schoukens. Non-parametric techniques in system identification. In J. Baillieul and T. Samad, editors, *Encyclopedia of Systems and Control*. Springer-Verlag, London, 2015.

[18] R. Pintelon, J. Schoukens, and G. Vandersteen. Frequency domain system identification using arbitrary signals. *IEEE Transactions on Automatic Control*, 42(12):1717 – 1720, Dec. 1997.

[19] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbé. Estimation of nonparametric noise and FRF models for multivariable systems–Part I: Theory. *Mechanical Systems and Signal Processing*, 24(3):573 – 595, 2010.

[20] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbé. Estimation of nonparametric noise and FRF models for multivariable systems–part ii: Extensions, applications. *Mechanical Systems and Signal Processing*, 24(3):596 – 616, 2010.

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

[22] S. Särkkä. Linear operators and stochastic partial differential equations in gaussian process regression. In Timo Honkela, W. Duch, Mark Girolami, and Samuel Kaski, editors, *Artificial Neural Networks and Machine Learning – ICANN 2011*, volume 6792 of *Lecture Notes in Computer Science*, pages 151–158. Springer Berlin Heidelberg, 2011.

[23] J. Schoukens, Y. Rolain, and R. Pintelon. Analysis of windowing/leakage effects in frequency response function measurements. *Automatica*, 42(1):27 – 38, 2006.

[24] J. Schoukens, G. Vandersteen, K. Barbé, and R. Pintelon. Nonparametric preprocessing in system identification: a powerful tool. *European Journal of Control*, 15(3-4):260–274, May-Aug 2009. 10th European Control Conference, Budapest, Hungary, AUG 23-26, 2009.

[25] Peter J. Schreier and Louis L. Scharf. *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.

[26] Anders Stenman, Fredrik Gustafsson, Daniel E. Rivera, Lennart Ljung, and Tomas McKelvey. On adaptive smoothing of empirical transfer function estimates. *Control Engineering Practice*, 8(11):1309 – 1315, 2000.