The Analysis of Variance Error Part II:Fundamental Principles.

Brett Ninness* Håkan Hjalmarsson†

Abstract

This paper presents the theoretical underpinnings for a preceding companion work in which new improved accuracy quantifications for noise induced estimation errors were presented. In particular, via the ideas of reproducing kernels and orthonormal parameterisations of the subspaces they represent, this paper develops new methods for evaluating certain quadratic forms in inverse Toeplitz matrices that are instrumental to the quantification of variance error. Additionally, new results on the convergence rates of generalised Fourier expansions are derived and then employed to derive necessary and sufficient conditions for the accuracy of the quantifications of this paper, the preceding companion paper, and the work by other authors.

Technical Report EE2002021, Department of Electrical and Computer Engineering, University of Newcastle,AUSTRALIA

1 Introduction

The motivations for quantifying noise induced estimation errors - so called 'variance errors' - especially in the frequency domain, have been presented in the preluding companion paper [18], to which we refer the reader to avoid repetition here.

In that work, the main result was to present, prove and then profile the utility of a new variance error quantifications for Box–Jenkins and related Output–Error model structures, which in the latter case is given as (see the companion paper [18] for a definition of the quantities involved)

$$\text{Var}\{G(e^{j\omega},\widehat{\theta}^n_N)\} \approx \frac{2}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \tag{1}$$

Compared to the pre-existing and well known quantification[14, 16, 13]

$$\text{Var}\{G(e^{j\omega},\widehat{\theta}^n_N)\} \approx \frac{m}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)} \tag{2}$$

the new expression (1) differs in several key ways. In particular, for the Output-Error case considered in (1), it is frequency dependent in a manner dictated by the poles $\{\xi_0,\cdots,\xi_{m-1}\}$ of the estimated transfer function $G(q,\widehat{\theta}^n_N)$. Also, it contains a new factor of 2 in the quantification.

*This work was supported by the Australian Research Council. This author is with the School of of Electrical Engineering & Computer Science, University of Newcastle, Australia and can be contacted at email:`brett@ee.newcastle.edu.au` or FAX: +61 49 21 69 93. Part of this work was completed while visiting S3-Automatic Control, The Royal Institute of Technology, Stockholm, Sweden.

†This author is with the Department of Sensors, Signals and Systems (Automatic Control), The Royal Institute of Technology, S-100 44 Stockholm, Sweden and may be contacted at email:`hakan.hjalmarsson@s3.e.kth.se` or FAX: +46 8 790 7329

The enhanced accuracy afforded by these modifications was illustrated via extensive empirical study in the preceding paper [18], after that same work presented and proved the main convergence results that led to the new quantification (1) together with certain generalisations applicable to closed loop data collection.

With this in mind, the purpose of the present paper is threefold. Firstly, while the preceding work has established the new quantification (1), it has not elucidated the theoretical background behind its improved performance relative to the existing expression (2). This paper therefore provides an examination of the accuracy limiting aspects of the existing variance error quantification (2) and the commensurate accuracy enhancing aspects associated with new quantifications such as (1).

Secondly, and motivated by this first point, this work derives new analysis tools that circumvent the difficulties exposed here in the examination of (1), and which lead to certain key results that are central to the proof of the main results of the preluding work [18].

Finally, this paper provides a theoretical analysis to inform assessments of the accuracy of any variance error quantifications, such as those of the companion paper [18] and pre-existing ones such as (2). They are obtained by employing the new theoretical tools developed in this paper.

There are two key principles underlying and motivating the theoretical developments of this paper. Firstly, prior work [14, 16, 13] leading to (2) has employed classical results on the asymptotics of Toeplitz matrices and Fourier series. However, as will be established here in §7, the accuracy of these asymptotics for finite (and hence practical) model orders depends crucially on the smoothness of underlying functions.

In recognition of this, the paper here as well as the companion [18] use certain generalised Toeplitz and Fourier asymptotics that are with respect to certain rational orthonormal bases *specifically adapted* to the functions being reconstructed. This strategy leads to the term $\sum_{k=0}^{m-1}(1-|\xi_k|^2)|e^{j\omega}-\xi_k|^{-2}$ in (1), which as already illustrated in [18] is a major contribution towards the accuracy of (1), and this factor has previously been introduced in a different but related context in [20].

Secondly, in the Output–Error and Box–Jenkins cases, previous work [14, 13] has implicitly approximated a full rank (or near full rank) matrix (the Hessian of the expected identification-cost) with a quantity close to a rank-one form. In order to avoid the approximation errors attendant to this strategy, we replace it with arguments relying on theory associated with reproducing kernels. This leads to the numerator factor of 2 in the expression (1).

Readers who are only interested in the main results on quantifying variance error, but not their genesis, are advised to only consider the preluding work [18], since the paper at hand is completely devoted to explaining the theoretical basis for the results already presented in [18], and hence providing the core technical results required in the proofs of the main results of [18].

This treatment begins in §2 with a presentation of some the key new technical ideas to be employed, and which relate to rational orthonormal bases and associated reproducing kernels. This is followed by the discussion in §3 providing a synopsis of pre-existing work leading to quantifications such as (2). This is then analysed in §4 so as to explain the need for the $\sum_{k=0}^{m-1}(1-|\xi_k|^2)|e^{j\omega}-\xi_k|^{-2}$ term in (1). The precise mechanics of how this term enters are discussed in §5. The following §6 then provides further analysis of pre-existing methods leading to (2), and in doing so establishes how the factor of 2 arises in the new quantification (1), before culminating with the main technical result of the paper, which is the cornerstone of the new results of the companion work [18].

Finally, given that convergence results leading to (1) have already been formally proven in [18], then in a certain sense, the utility of the new quantifications in [18], such as (1), has already been established. However, those proofs have relied on further results of this paper. Furthermore, establishing the quantification (1), requires a passage involving the use of an asymptotic convergence result at a finite truncation. While the validity of this has been illustrated empirically in [18], it is important that it also be established theoretically. The final §7 addresses this issue by establishing necessary and sufficient conditions for given convergence rate bounds of the new approximations such as (1), to the actual variances they seek to quantify. As well, necessary conditions for existing quantifications such as (2) are also provided, and it is significant that these necessary conditions seem difficult to satisfy in practice.

2 Technical Preliminaries

The key new mathematical ideas that are discussed and then employed and developed in this paper and in the proofs of the preluding companion work [18] involve certain rational orthonormal bases together with associated ideas of generalised Toeplitz matrices, generalised Fourier series convergence and, perhaps most importantly, the concept of an associated 'Reproducing Kernel'. The purpose of this section is to present these underlying ideas which underpin the theoretical developments and analysis of this paper and [18].

2.1 An Orthonormal Basis

The most important and fundamental technical tool to be employed is a particular formulation of rational functions $\{\mathcal{B}_0(z), \mathcal{B}_1(z), \cdots, \mathcal{B}_{m-1}(z)\}$ defined by a choice of poles $\{\xi_0, \xi_1, \cdots, \xi_{m-1}\}$ all contained within the open unit disk $\mathbf{D} \triangleq \{z \in \mathbf{C} : |z| < 1\}$ as

$$\mathcal{B}_k(z) \triangleq \frac{\sqrt{1-|\xi_k|^2}}{z-\xi_k}\prod_{\ell=0}^{k-1}\left(\frac{1-\overline{\xi}_\ell z}{z-\xi_\ell}\right), \qquad \mathcal{B}_0(z) \triangleq \frac{\sqrt{1-|\xi_0|^2}}{z-\xi_0}. \tag{3}$$

These functions are orthonormal with respect to the inner product ($\bar{\cdot}$ denotes conjugation)

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)\overline{g(\omega)}\,\mathrm{d}\omega \tag{4}$$

and the usual trigonometric basis $\{e^{-j\omega}, \cdots, e^{-jm\omega}\}$ is obtained as a special case of $\xi_0 = \xi_1 = \cdots = \xi_{m-1} = 0$. In fact, the set $\{\mathcal{B}_k(1/z)\}_{k=0}^{\infty}$ forms a basis for $H_2(\mathbf{T})$ whenever $\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$ and where $\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}$ is the complex unit circle [17].

2.2 Generalised Toeplitz Matrices

The importance of these orthonormal basis functions (3) is in defining what is called a 'generalised' block Toeplitz matrix $M_n(F)$ as follows ($\cdot^\star$ denotes 'conjugate transpose')

Definition 2.1. *An $n \times n$ 'Generalised' Block-Toeplitz matrix is defined by a $p \times p$ positive definite matrix valued function $F(\omega)$ as*

$$M_n(F) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi}[\Gamma_m(e^{j\omega}) \otimes I_p]F(\omega)[\Gamma_m^\star(e^{j\omega}) \otimes I_p]\,\mathrm{d}\omega \tag{5}$$

where $mp = n$, I_p is a $p \times p$ identity matrix and

$$\Gamma_m(z) \triangleq [\mathcal{B}_0(z), \mathcal{B}_1(z), \cdots, \mathcal{B}_{m-1}(z)]^T \tag{6}$$

with $\{\mathcal{B}_k(z)\}$ given by (3). □

Here $\otimes$ is the Kronecker tensor product of matrices, which is defined when A is $m \times n$ and B is $\ell \times p$ according to the $m\ell \times np$ dimensional matrix $A \otimes B$ being given by

$$A \otimes B \triangleq \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}.$$

There are various useful identities applicable to this tensor product, but the only ones to be employed in this and the companion paper [18] are that $(A \otimes B)^T = A^T \otimes B^T$ and that for C and D of appropriate dimension $(A \otimes B)(C \otimes D) = AC \otimes BD$. For more detail on this topic of Kronecker product properties see [3].

The 'generalised' epithet in Definition 2.1 is derived from the fact that for the special case of $\xi_k = 0$ for all k, then $\{\mathcal{B}_k(e^{j\omega})\}$ reverts to the trigonometric basis $\{e^{-j\omega k}\}$ and in this case, $M_n(F)$ possesses a more familiar 'block-banded' Toeplitz matrix structure. This paper reserves the notation $T_n(F)$ for this special case of $M_n(F)$ as follows.

Definition 2.2. *An $n \times n$ Block-Toeplitz matrix is defined by a $p \times p$ positive definite matrix valued function $F(\omega)$ as*

$$T_n(F) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} [\Lambda_m(e^{j\omega}) \otimes I_p] F(\omega) [\Lambda_m^\star(e^{j\omega}) \otimes I_p] \, \mathrm{d}\omega \tag{7}$$

where $pm = n$ and

$$\Lambda_m(z) \triangleq [1, z, z^2, \cdots, z^{m-1}]^T. \tag{8}$$

□

2.3 Generalised Fourier Results

These Toeplitz matrix definitions realise their fundamental role in deriving approximations like (2) and (1) by way of frequency dependent quadratic forms, and the approximations actually arise by virtue of these forms representing convergent Fourier reconstructions. In the 'traditional' block-banded Toeplitz matrix case this principle appears as follows.

Theorem 2.1. *Provided $F(\omega)$ of dimension $p \times p$ is positive definite and (component-wise) continuous for all $\omega \in [-\pi, \pi]$, then*

$$\lim_{m \to \infty} \frac{1}{m} [\Lambda_m^\star(e^{j\omega}) \otimes I_p] T_{mp}^{-1}(F) [\Lambda_m(e^{j\lambda}) \otimes I_p] = \begin{cases} F^{-1}(\omega) & ; \omega = \lambda \\ 0 & ; \omega \neq \lambda \end{cases}$$

component-wise and uniformly on $\omega \in [-\pi, \pi]$. Here I_p is a $p \times p$ identity matrix.

Proof. See [11]. □

It is this result that underlies the approximation (2), and to derive the improved accuracy approximation (1) a generalised Toeplitz matrix form of it will be employed and which is given as follows.

Theorem 2.2. *Provided $F(\omega)$ of dimension $p \times p$ is positive definite and (component-wise) Lipschitz continuous of order $\alpha > 0$ for all $\omega \in [-\pi, \pi]$, then for any integer k (possibly negative)*

$$\lim_{m\to\infty} \frac{1}{\kappa_m(\omega)}[\Gamma_m^\star(e^{j\omega}) \otimes I_p] M_{mp}^{-k}(F)[\Gamma_m(e^{j\lambda}) \otimes I_p] = \begin{cases} F^{-k}(\omega) & ; \omega = \lambda \\ 0 & ; \omega \neq \lambda \end{cases}$$

component-wise and uniformly on $\omega \in [-\pi, \pi]$ where

$$\kappa_m(\omega) \triangleq \sum_{k=0}^{m-1} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2} \tag{9}$$

with $\Gamma_m(z)$ being defined in (6) and the poles $\{\xi_k\}$ in (9) deriving from (6) via (3).

Proof. See [20]. □

Note that the 'classical' result of Theorem 2.1 is a special case of Theorem 2.2 in which $\xi_k = 0$ for all k. Associated with the use of this result is an idea introduced in [21] that arbitrary symmetric $mp \times mp$ matrices A_{mp} and B_{mp} are denoted as being 'asymptotically equivalent' by the notation $A_{mp} \sim B_{mp}$ as $m \to \infty$ if

$$\lim_{m\to\infty} \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p](A_{mp} - B_{mp})^\star (A_{mp} - B_{mp})[\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} = 0, \qquad \forall \omega \in [-\pi, \pi]. \tag{10}$$

The utility of this definition arises via the simple decomposition

$$\begin{aligned} \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p] A_{mp} [\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} &= \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p] B_{mp} [\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} + \\ &\quad \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p](A_{mp} - B_{mp})[\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} \end{aligned}$$

so that by using the Cauchy–Schwartz inequality on the last term, if (10) is satisfied then as $m \to \infty$ the quadratic forms in A_{mp} and B_{mp} converge to the same value for any ω.

As is indicated by the external reference to proofs of the preceding results, they have been developed elsewhere for other applications. In particular, Theorem 2.2 was developed in [20] by extending scalar ($p = 1$) results given in [21] and for the purpose of analysing pre-filtered FIR and ARX structures.

The application of Theorem 2.2 here and in [18] to the examination of Output-Error and Box–Jenkins model structures is new and because of this, the conclusions of this paper (and its companion) such as variance error being model structure dependent are also new. Furthermore, the employment here of reproducing kernel principles to derive variance quantifications is also new.

2.4 Reproducing Kernels

The idea of what is called a 'Reproducing Kernel' for a space will prove to be a vital tool that allows for the direct simplification of complicated quantities via what is essentially a geometric principle.

To explain these key ideas, consider a space X_m defined by a number of orthonormal basis elements $\{\mathcal{B}_0, \cdots, \mathcal{B}_{m-1}\}$ as

$$X_m \triangleq \text{Span}\left\{\mathcal{B}_0(e^{j\omega}), \cdots, \mathcal{B}_{m-1}(e^{j\omega})\right\}$$

and suppose that $f(z) \in X_m$. Then since the mapping $f \mapsto f(e^{j\omega})$ is a (bounded) linear functional $X_m \to \mathbf{C}$, then it is a consequence of the Riesz Representation Theorem [22, 23] that for any fixed ω, a further element $\varphi_m(\lambda,\omega) \in X_m$ exists such that

$$f(e^{j\omega}) = \langle f(\lambda), \varphi_m(\lambda,\omega)\rangle \qquad \forall f \in X_m. \tag{11}$$

Here, and in what follows, in any inner product the implied integration according to the definition (4) will be over the common argument (in the above case, λ). Since the function $\varphi_m(\lambda,\omega)$ 'reproduces' arbitrary elements f of X_m at a particular point ω, it is termed [5] the 'Reproducing Kernel' for the space X_m.

There are further important properties of this kernel which, in fact, stem directly from the reproducing one. In particular, $\varphi_m(\lambda,\omega)$ is 'Hermitian Symmetric' in that, since for any fixed ω the kernel $\varphi_m(\lambda,\omega) \in X_m$, then

$$\varphi_m(\lambda,\omega) = \langle \varphi_m(\zeta,\omega), \varphi_m(\zeta,\lambda)\rangle = \overline{\langle \varphi_m(\zeta,\lambda), \varphi_m(\zeta,\omega)\rangle} = \overline{\varphi_m(\omega,\lambda)}.$$

This then implies that $\varphi_m(\lambda,\omega)$ is the *unique* element in X_m that has the property (11), since if another function $H_m(\lambda,\omega)$ also satisfied (11), then it would hold that

$$H_m(\lambda,\omega) = \overline{H_m(\omega,\lambda)} = \overline{\langle H_m(\zeta,\lambda), \varphi_m(\zeta,\omega)\rangle} = \langle \varphi_m(\zeta,\omega), H_m(\zeta,\lambda)\rangle = \varphi_m(\lambda,\omega).$$

This last uniqueness property will be particularly vital in later developments in §6. Although the reproducing kernel is unique, there may (of course) be many different ways of expressing it. One obvious one is to use the orthonormal basis $\{\mathcal{B}_k\}$ for X_m to express the quantity as

$$\varphi_m(\lambda,\omega) = \sum_{k=0}^{m-1} \mathcal{B}_k(e^{j\lambda})\overline{\mathcal{B}_k(e^{j\omega})}. \tag{12}$$

This can be verified by noting that if $f = \sum_\tau c_\tau \mathcal{B}_\tau$ for some constants c_τ, then

$$\begin{aligned} \left\langle \sum_{\tau=0}^{m-1} c_\tau \mathcal{B}_\tau(e^{j\lambda}), \sum_{k=0}^{m-1} \mathcal{B}_k(e^{j\lambda})\overline{\mathcal{B}_k(e^{j\omega})} \right\rangle &= \sum_{\tau=0}^{m-1} c_\tau \sum_{k=0}^{m-1} \mathcal{B}_k(e^{j\omega}) \langle \mathcal{B}_\tau(e^{j\lambda}), \mathcal{B}_k(e^{j\lambda})\rangle \qquad (13) \\ &= \sum_{\tau=0}^{m-1} c_\tau \mathcal{B}_\tau(e^{j\omega}) = f(e^{j\omega}). \end{aligned}$$

3 Pre-Existing Work

With this technical background in place, the paper now moves to a synopsis of the pre-existing work [14, 13] on variance quantification, which has lead to (2), and which this and the companion paper [18] have sought to improve upon via new quantifications such as (1).

Additionally, a key purpose of the ensuing discussion of this section is to allow a diagnosis for the inaccuracy of (2) as illustrated in all of figures 1–11 of the preluding paper [18], and to also provide a framework via which the new results such as (1) may be established and elucidated.

To proceed with this, the general Box–Jenkins model structure studied here and in the companion work [18] is given as

$$y_t = G(q,\theta^n)u_t + H(q,\theta^n)e_t = \frac{B(q,\theta^n)}{A(q,\theta^n)}u_t + \frac{C(q,\theta^n)}{D(q,\theta^n)}e_t \tag{14}$$

where the numerator and denominator polynomials are of the form

$$\begin{aligned} A(q,\theta^n) &= q^m + a_{m-1}q^{m-1} + \cdots + a_1 q + a_0, && (15)\\ B(q,\theta^n) &= b_{m-1}q^{m-1} + b_{m-2}q^{m-2} + \cdots + b_1 q + b_0, && (16)\\ D(q,\theta^n) &= q^m + d_{m-1}q^{m-2} + \cdots + d_1 q + d_0, && (17)\\ C(q,\theta^n) &= q^m + c_{m-1}q^{m-2} + \cdots + c_1 q + c_0 && (18)\end{aligned}$$

and $\theta^n \in \mathbf{R}^n$ (with $n = 4m$) is defined as

$$\theta^n = [a_0, b_0, d_0, c_0, a_1, b_1, d_1, c_1, \cdots, a_{m-1}, b_{m-1}, d_{m-1}, c_{m-1}].$$

At times, in sections that follow, it will sometimes be expedient to consider the simplified Output-Error situation in which a fixed noise model $H(q,\theta^n) = 1$ is taken. This implies that the c_k and d_k terms indicated in θ^n are not present and hence $n = 2m$.

With this in mind, note that the sensitivity of the transfer functions $G(q,\theta^n)$, $H(q,\theta^n)$ to changes in their parameterisation θ^n can be measured as

$$\left[\frac{\mathrm{d}G(q,\theta^n)}{\mathrm{d}\theta^n}, \frac{\mathrm{d}H(q,\theta^n)}{\mathrm{d}\theta^n}\right] = [\Lambda_m(q) \otimes I_4]\Omega(q,\theta^n)Z(q,\theta^n) \tag{19}$$

where $\Lambda_m(q)$ has been defined in (8) and

$$\Omega(q,\theta^n) \triangleq \begin{bmatrix} A^{-1}(q,\theta^n)I_2 & \varnothing \\ \varnothing & D^{-1}(q,\theta^n)I_2 \end{bmatrix}, \tag{20}$$

$$Z(q,\theta^n) \triangleq \begin{bmatrix} -G(q,\theta^n) & 0 \\ 1 & 0 \\ 0 & -H(q,\theta^n) \\ 0 & 1 \end{bmatrix}. \tag{21}$$

Use of equation (12) of [18] then leads to an expression for the gradient of the prediction error $\varepsilon_t(\theta^n) \triangleq y_t - \widehat{y}_t(\theta^n)$ as a filtered version of $\zeta_t(\theta^n) \triangleq [u_t, \varepsilon_t(\theta^n)]^T$ according to

$$\psi_t(\theta^n) = H^{-1}(q,\theta^n)[\Lambda_m(q) \otimes I_4]\Omega(q,\theta^n)Z(q,\theta^n)\zeta_t(\theta^n). \tag{22}$$

The point of this formulation is that it allows a frequency domain interpretation of noise induced estimation errors. Specifically, recall from (8) of [18] that under quite mild assumptions, there exists a $\theta^n_\circ$ defined in (7) of [18] such that

$$\sqrt{N}(\widehat{\theta}^n_N - \theta^n_\circ) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_n), \qquad \text{as } N \to \infty \tag{23}$$

where the covariance matrix P_n is defined in terms of two other matrices R_n and Q_n as $P_n \triangleq R_n^{-1}Q_nR_n^{-1}$.

Furthermore, if the true system is in the model structure defined by (14) so that a true parameter vector exists and is equal to the asymptotic estimate $\theta^n_\circ$, then $\varepsilon_t(\theta^n_\circ) = e_t$ (actually, much less than this has been assumed in Theorem 4.1 of [18], but it will unnecessarily distract to consider this for the moment) so that use of equation (10) of [18], Parseval's Formula, and input assumption 4.1 of [18] leads to (with obvious compacting of notation involving $\theta^n_\circ$)

$$\begin{aligned} R_n &= \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \mathbf{E}\left\{\psi_t(\theta^n_\circ)\psi_t^T(\theta^n_\circ)\right\} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} [\Lambda_m(e^{j\omega}) \otimes I_4]\Omega_\circ(e^{j\omega})\frac{Z_\circ(e^{j\omega})\Phi_\zeta(\omega,\theta^n_\circ)Z_\circ^\star(e^{j\omega})}{|H(e^{j\omega},\theta^n_\circ)|^2}\Omega_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega}) \otimes I_4]\mathrm{d}\omega. \end{aligned} \tag{24}$$

The block-Toeplitz matrix notation of Definition 2.2 then allows R_n to be compactly expressed (with $p = 4$ in that definition) as

$$R_n = T_n\left(\Omega(e^{j\omega},\theta_\circ^n)\frac{Z(e^{j\omega},\theta_\circ^n)\Phi_\zeta(\omega,\theta_\circ^n)Z(e^{j\omega},\theta_\circ^n)^\star}{|H(e^{j\omega},\theta_\circ^n)|^2}\Omega(e^{j\omega},\theta_\circ^n)^\star\right). \tag{25}$$

Also, again under the assumption of $\varepsilon_t(\theta_\circ^n) = e_t$, then

$$Q_n \triangleq \lim_{N\to\infty}\frac{\sigma^2}{N}\sum_{t=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ^n)\psi_t^T(\theta_\circ^n)\right\} = \sigma^2 R_n$$

so that the matrix P_n quantifying the parameter space variability of $\widehat{\theta}_N^n$ via (23) is in fact expressible as a block Toeplitz matrix associated with a particular spectral density as follows

$$P_n = \sigma^2 R_n^{-1} = \sigma^2 T_n^{-1}\left(\Omega(e^{j\omega},\theta_\circ^n)\frac{Z(e^{j\omega},\theta_\circ^n)\Phi_\zeta(\omega,\theta_\circ^n)Z(e^{j\omega},\theta_\circ^n)^\star}{|H(e^{j\omega},\theta_\circ^n)|^2}\Omega(e^{j\omega},\theta_\circ^n)^\star\right). \tag{26}$$

This frequency domain formulation of P_n is a key ingredient underlying the pre-existing methods of [14, 16, 13, 15] that arrive at the approximation (2). A second fundamental idea in that work is to relate the parameter space variability quantified by P_n to frequency domain variability of $G(e^{j\omega},\widehat{\theta}_N^n)$ and $H(e^{j\omega},\widehat{\theta}_N^n)$ via the Taylor expansion (14) of [18] as

$$\Pi^T(e^{j\omega},\widehat{\theta}_N^n) - \Pi^T(e^{j\omega},\theta_\circ^n) = Z^T(e^{j\omega},\theta_\circ^n)\Omega^T(e^{j\omega},\theta_\circ^n)[\Lambda_m^T(e^{j\omega})\otimes I_4](\widehat{\theta}_N^n - \theta_\circ^n) + o(\|\widehat{\theta}_N^n - \theta_\circ^n\|^2) \tag{27}$$

Therefore, using (8) of [18] and the previous Toeplitz matrix formulation of P_n provides

$$\sqrt{N}\left[\begin{array}{c} G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n) \\ H(e^{j\omega},\widehat{\theta}_N^n) - H(e^{j\omega},\theta_\circ^n)\end{array}\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Delta_n(\omega)), \qquad \text{as } N\to\infty \tag{28}$$

where using (26)

$$\Delta_n(\omega) = \sigma^2 Z_\circ^\star(e^{j\omega})\Omega_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_4]T_n^{-1}\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right)[\Lambda_m(e^{j\omega})\otimes I_4]\Omega_\circ(e^{j\omega})Z_\circ(e^{j\omega}) \tag{29}$$

There is now a potential difficulty in continuing the analysis of this expression in that the so-called 'symbol' $\Omega_\circ Z_\circ\Phi_u Z_\circ^\star\Omega_\circ^\star/|H_\circ^n|^2$ defining the above Toeplitz matrix is, by construction, singular. On the other hand, as will be shown in the following equation (31), a approximating step employed in [14] requires the inversion of this symbol. To circumvent this problem, the previous work [14] has defined a perturbed matrix (note the symbol of T_n^{-1})

$$\begin{aligned}\Delta_n(\omega,\delta) \;\triangleq\;& \sigma^2 Z_\circ^\star(e^{j\omega})\Omega_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_4]\,\times \\ & T_n^{-1}\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ|^2}\Omega_\circ^\star + \delta I_4\right)[\Lambda_m(e^{j\omega})\otimes I_4]\Omega_\circ(e^{j\omega})Z_\circ(e^{j\omega})\end{aligned} \tag{30}$$

to argue that $\Delta_n(\omega) = \lim_{\delta\to 0}\Delta_n(\omega,\delta)$, and that this strategy can be interpreted as one of modifying $V_N(\theta)$ to provide a 'regularised' estimate.

The final principle underlying the analysis of existing work such as [14, 16, 10] is then that the quadratic form defining $\Delta_n(\omega)$, by virtue of being formulated in terms of inverses of Toeplitz matrices parameterised by spectral densities, can be viewed as an m'th order Fourier reconstruction of the inverse of the spectral density. Specifically, by Theorem 2.1

$$\lim_{m\to\infty}\frac{1}{m}\Delta_{4m}(\omega,\delta) \triangleq \Delta(\omega,\delta) = \sigma^2 Z_\circ^\star(e^{j\omega})\Omega_\circ^\star(e^{j\omega})\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ|^2}\Omega_\circ^\star + \delta I_4\right)^{-1}\Omega_\circ(e^{j\omega})Z_\circ(e^{j\omega}). \tag{31}$$

Therefore, using the Matrix Inversion Lemma [1] twice (and dropping the ω dependence for readability)

$$\begin{aligned}\Delta(\omega,\delta) &= \sigma^2 Z_\circ^\star\Omega_\circ^\star\left[\delta^{-1}I - \delta^{-1}\Omega_\circ Z_\circ\left[|H_\circ|^2\Phi_\zeta^{-1} + \delta^{-1}Z_\circ^\star\Omega_\circ^\star\Omega_\circ Z_\circ\right]^{-1}\delta^{-1}Z_\circ^\star\Omega_\circ^\star\right]\Omega_\circ Z_\circ \\ &= \sigma^2|H_\circ|^2 Z_\circ^\star\Omega_\circ^\star\Omega_\circ Z_\circ\left[\Phi_\zeta Z_\circ^\star\Omega_\circ^\star\Omega_\circ Z_\circ + \delta|H_\circ|^2 I\right]^{-1} \end{aligned}\tag{32}$$

Consequently, works such as [14] conclude that

$$\lim_{\delta\to 0}\lim_{m\to\infty}\frac{1}{m}\Delta_{4m}(\omega,\delta) = \sigma^2|H(e^{j\omega},\theta_\circ^n)|\Phi_\zeta^{-1}(\omega) \tag{33}$$

Finally, assuming that the convergence with increasing m has approximately occurred for finite m leads to the approximation

$$\Delta_{4m}(\omega) \approx m\sigma^2|H(e^{j\omega},\theta_\circ^n)|\Phi_\zeta^{-1}(\omega) \tag{34}$$

and in a similar vein, assuming that convergence of (28) has approximately occurred for finite N provides the approximation

$$\mathrm{Cov}\{\Pi(e^{j\omega},\widehat{\theta}_N^n)\} \approx \frac{1}{N}\Delta_{4m}(\omega).$$

Combining these expressions then furnishes the overall estimate

$$\mathrm{Cov}\{\Pi(e^{j\omega},\widehat{\theta}_N^n)\} \approx \frac{m}{N}\sigma^2|H(e^{j\omega},\widehat{\theta}_N^n)|\Phi_\zeta^{-1}(\omega) \tag{35}$$

which leads to the existing quantification (2).

4 Genesis of Impaired Approximation

The point of the preceding synopsis of [14] and related works such as [16, 10] was to now isolate why, as illustrated in figures 1–11 of the companion paper [18], it may occur that (2) suffers from poor accuracy. In fact, there are two main influences that act to degrade the fidelity of (2).

The first of them, explained in this section, arises since (2) and (35) are predicated on the convergence in (33) having approximately occurred for finite model order m so that (34) can be concluded. However, as we shall explain, whether this approximate convergence has occurred can be problematic.

To provide more detail, and in the interests of most clearly exposing the idea, consider the simplified Output-Error case in which only a dynamics model $G(q,\theta^n)$ is estimated. This implies that $H(q,\theta_\circ^n) = 1$ and

$$\Omega(q,\theta^n) = \frac{1}{A(q,\theta^n)}, \qquad Z(q,\theta^n) = \begin{bmatrix} -G(q,\theta^n) \\ 1 \end{bmatrix}, \qquad \Phi_\zeta(\omega) = \Phi_u(\omega). \tag{36}$$

Therefore, the limiting value of (33) arises from Fourier series reconstruction via (31) of a matrix valued function

$$\frac{Z_\circ(e^{j\omega})\Phi_u(\omega)Z_\circ^\star(e^{j\omega})}{|A_\circ(e^{j\omega})|^2} + \delta I_2. \tag{37}$$

As is well known [4], the rate of classical Fourier series convergence is governed by the smoothness of the function being reconstructed. Furthermore, the variation (and hence non-smoothness) of (37) can be significantly *increased* due to division by the $|A_\circ(e^{j\omega})|^2$ term.

To see this, suppose for simplicity that all the zeroes of $A_\circ(z)$ are in the left half plane. Then (see figure 1) $|A_\circ(e^{j0})| \leq \eta^m$ for some $\eta < 1$ and $|A_\circ(e^{j\pi})| \geq \gamma^m$ for some $\gamma > 1$ so that division of a function by $|A_\circ(e^{j\omega})|^2$ can magnify the maximum value of (37) by a factor of $\gamma^{2m} \gg 1$ and shrink the minimum values by a factor of $\eta^{2m} \ll 1$.

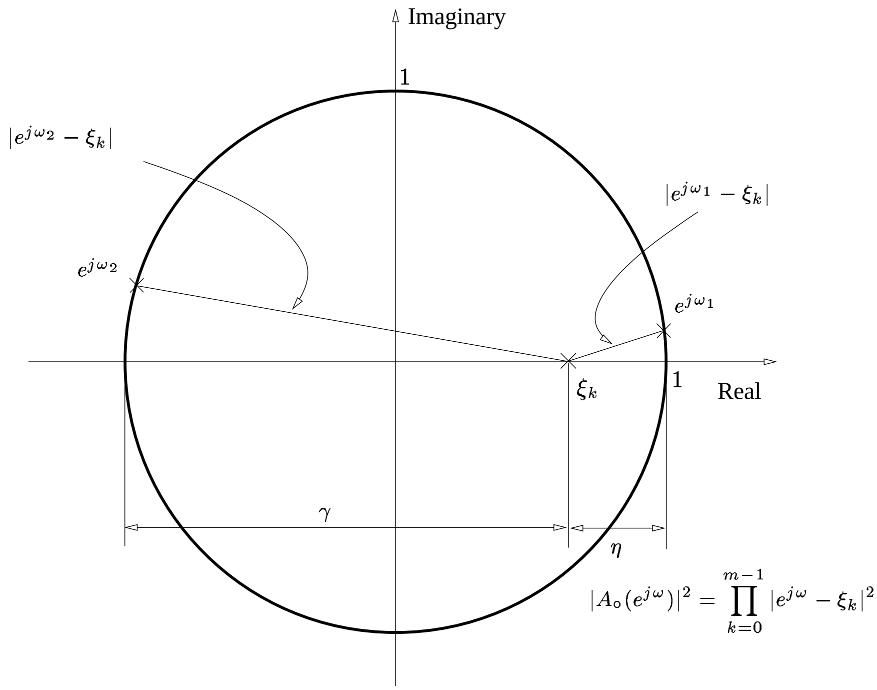

Figure 1: *Graphical illustration of how the magnitude variation of $A_\circ(e^{j\omega})$ depends on ω and hence increases for increasing order m.*

Therefore, as the model order m grows, the function (37) being implicitly Fourier reconstructed in (31) can develop greater variation, which necessitates more terms in its Fourier expansion before approximate convergence can be assumed in the step (34) leading to (2). However, a key point is that the number of terms in the implicit Fourier reconstruction (31) is also given by the quantity m.

That is, a circular situation arises in which just as the approximation order grows with m, the function being approximated develops greater variation with increasing m, and hence also requires a higher reconstruction order m in order to attain accurate approximation.

The net result is that, depending on the problem setting, it is problematic as to whether Fourier convergence can be assumed to hold in such a way that the approximation (35) via (34) can be concluded.

The material in the following §7 is devoted to to more firmly establishing this point by showing that a necessary and sufficient condition for a given convergence rate of (33) and hence a given accuracy of the quantification (2) is a commensurate Lipschitz smoothness restriction on quantities equal, or related to (37).

5 Initial Resolution: A Change of Basis

As precursed via the plot of the modified expression (1) in figures 1–11 of [18], the Fourier series convergence difficulties raised in §4 are not insurmountable. The key idea behind providing a solution is to change the orthonormal basis involved in the Fourier series to one that is adapted to the function (37) being reconstructed.

In particular, again focusing on the simplified Output–Error case for the sake of clarity, the problem of quantifying $\Delta_n(\omega)$ in (29) may be re-parameterised in terms of the orthonormal basis $\{\mathcal{B}_k(z)\}$ introduced in §2 after it is 'adapted' to the underlying denominator polynomial $A(z,\theta_\circ^n)$ in the sense that the zeros $\{\xi_0,\xi_1,\cdots,\xi_{m-1}\}$ of $A(z,\theta_\circ^n)$ are used as the poles of $\{\mathcal{B}_k(z)\}$ according to the formulation (3).

The use of this basis to obtain the approximation (1) appearing as the improved dashed line approximation in figures 1–11 of [18] now involves the analysis invented in [14] up until (29) being retained. However, since the poles $\{\xi_k\}$ of the bases (3) are chosen the same as the zeros of $A(z,\theta_\circ^n)$, then using the definition (6) it holds that for some non-singular $2m\times 2m$ matrix J_m, the matrix $\Lambda_m(z)$ appearing in (29), and defined in (8) in terms of a trigonometric basis $\{e^{j\omega n}\}$, is expressible as

$$A^{-1}(z,\theta_\circ^n)\Lambda_m(z) = J_m\Gamma_m(z).$$

Therefore

$$A^{-1}(z,\theta^n)(\Lambda_m(z)\otimes I_2) = J_m\Gamma_m(z)\otimes I_2 = (J_m\otimes I_2)(\Gamma_m(z)\otimes I_2). \tag{38}$$

Further developments then depend on using the generalised Toeplitz matrix $M_n(F)$ of Definition 2.1 which via (38) is related to the conventional 'banded' Toeplitz matrix $T_n(F)$ of Definition 2.2 via

$$T_n\left(\frac{F}{|A_\circ^2|}\right) = (J_m\otimes I_2)M_n(F)(J_m^T\otimes I_2) \tag{39}$$

and hence the quantity $\Delta_n(\omega,\delta)$ in (30) previously analysed via Fourier theory with respect to the trigonometric basis becomes

$$\begin{aligned}\Delta_{2m}(\omega,\delta) &= \frac{\sigma^2}{|A_\circ(e^{j\omega})|^2}Z_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_2](J_m^{-T}\otimes I_2)M_n^{-1}\left(Z_\circ\Phi_u Z_\circ^\star+\delta|A_\circ|^2 I_2\right)\times\\ &\quad (J_m^{-1}\otimes I_2)[\Lambda_m(e^{j\omega})\otimes I_2]Z_\circ(e^{j\omega})\\ &= \sigma^2 Z_\circ^\star(e^{j\omega})[\Gamma_m^\star(e^{j\omega})\otimes I_2]M_n^{-1}\left(Z_\circ\Phi_u Z_\circ^\star+\delta|A_\circ|^2 I_2\right)[\Gamma_m(e^{j\omega})\otimes I_2]Z_\circ(e^{j\omega}). \end{aligned}\tag{40}$$

In [21, 20] a generalised Fourier theory involving the generalised basis (3) is developed, for which the most pertinent result in the current context is the convergence result of Theorem 2.2. Applying it together with the well known Matrix Inversion Lemma [9] provides the conclusion

$$\lim_{m\to\infty}\frac{\Delta_{2m}(\omega,\delta)}{\kappa_m(\omega)} = \sigma^2 Z_\circ^\star(e^{j\omega})\left[Z_\circ(e^{j\omega})\Phi_u(\omega)Z_\circ^\star(e^{j\omega})+\delta|A_\circ(e^{j\omega})|^2 I_2\right]^{-1}Z_\circ(e^{j\omega}) \tag{41}$$

$$= \frac{\sigma^2 Z_\circ^\star(e^{j\omega})Z_\circ(e^{j\omega})}{\Phi_u(\omega)Z_\circ^\star(e^{j\omega})Z_\circ(e^{j\omega})+\delta|A_\circ(e^{j\omega})|^2} \tag{42}$$

so that

$$\lim_{\delta\to 0}\lim_{m\to\infty}\frac{\Delta_{2m}(\omega,\delta)}{\kappa_m(\omega)} = \frac{\sigma^2}{\Phi_u(\omega)}. \tag{43}$$

Therefore, following the same line of argument established in [13] that led to (35) provides

$$\mathbf{E}\left\{|G(e^{j\omega},\widehat{\theta}^n_N) - G(e^{j\omega},\theta^n_\circ)|^2\right\} \approx \frac{\sigma^2}{N}\frac{\kappa_m(\omega)}{\Phi_u(\omega)} = \frac{1}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \tag{44}$$

which, aside from a numerator factor of 2, is the new approximation (1).

A vital point here is that, by virtue of using the basis (3), the implicit Fourier reconstruction operating here (via (43)), is of a function $Z_\circ(e^{j\omega})\Phi_u(\omega)Z^\star_\circ(e^{j\omega})$ *whose Lipschitz smoothness is constant with respect to Fourier reconstruction length* m.

This property, which implies that the quality of the approximation improves monotonically with increasing m, is a key factor that imbues (44) with improved accuracy when compared with (2).

For example, as just shown, it is via recognising the importance of monotonic convergence that the term $\sum_{k=0}^{m-1}(1-|\xi_k|^2)|e^{j\omega}-\xi_k|^{-2}$ is introduced in (1), which by its frequency dependent nature is the key element in quantifying the variability shown in, for example, figure 1 of [18] as 'low-pass' rather than 'high-pass'.

6 Further Resolution: The Reproducing Kernel

There is a further essential point to recognise in deriving the improved approximation (1). Again it involves critical examination of Fourier convergence, and its resolution leads to the final ingredient that imbues (1) with enhanced accuracy; the factor of 2 in the numerator and which has not been explained by the developments of the previous sections.

To prescience the ensuing developments, when employing an asymptotic in model order m result for a finite m, it is vital to distinguish between cases where the finite m is expected to be equal to, or higher than, the true model order.

This is necessary since a key step in existing analysis involves progressing from (30) to (31) by using the Fourier convergence result of Theorem 2.1 to argue that (again, for sake of clarity, we are considering the simpler Output-Error case whereby (36) applies)

$$\lim_{m\to\infty}\frac{1}{m}Z^\star_\circ(e^{j\omega})[\Lambda^\star_m(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ\Phi_u Z^\star_\circ}{|A_\circ|^2}+\delta I_2\right)[\Lambda_m(e^{j\omega})\otimes I_2]Z_\circ(e^{j\omega}) = \left[\frac{Z_\circ(e^{j\omega})\Phi_u(\omega)Z^\star_\circ(e^{j\omega})}{|A_\circ(e^{j\omega})|^2}+\delta I_2\right]^{-1}. \tag{45}$$

The pre-existing analysis in [14] leading to the quantification (2) then involves assuming that the above equality holds approximately for the finite model order m actually used.

However if the chosen finite m is less than or equal to the true model order, then the Toeplitz matrix, and hence the whole matrix on the left hand side of (45) will be full rank, even for $\delta = 0$. At the same time, when $\delta = 0$, the matrix on the right hand side of (45) is only of rank one. It is therefore unlikely, as employed in [14], to be a good approximation to the full-rank quantity on the left hand side.

For example, as will be presently shown, when $\Phi_u(\omega) = \gamma$ a constant and m is less than or equal to the true model order, then the following *exact* equality holds, even for finite m

$$Z^\star_\circ(e^{j\omega})[\Lambda^\star_m(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{\gamma Z_\circ Z^\star_\circ}{|A_\circ|^2}\right)[\Lambda_m(e^{j\omega})\otimes I_2]Z_\circ(e^{j\omega}) = \frac{2}{\gamma}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \tag{46}$$

The authors consider that the ability to exactly evaluate a quantity as seemingly complex as the left hand side of (46) is a most surprising result!

Nevertheless, it is by recognising it that the important factor of 2 arises in the new quantification (1). In particular, note that this exact expression (46) is quite different to the approximate value of 1 which is argued via the existing analysis [14] as summarised in (30)-(32).

The key to establishing quantifications such as (46) is the use of the reproducing kernel ideas presented in §2.4. To explain these principles, consider the subspace

$$X_{2m} = \text{Span}\left\{\mathcal{B}_0(z), \mathcal{B}_1(z), \cdots, \mathcal{B}_{m-1}(z), G(z,\theta_\circ^n))\mathcal{B}_0(z), \cdots, G(z,\theta_\circ^n)\mathcal{B}_{m-1}(z)\right\}$$

where the $\{\mathcal{B}_k(z)\}$ are the orthonormal basis functions defined in (3) with poles $\{\xi_k\}$ equal to those of $G(z,\theta_\circ^n)$. Therefore, with $Z_\circ(z) \triangleq Z(z,\theta_\circ^n)$ given by (21), all the basis elements of X_{2m}, when collected in a vector, may be written using the definition (6) as

$$[\Gamma_m(z) \otimes I_2]Z_\circ(z).$$

Furthermore, via the argument from equation (38) onwards, the study of the left hand side of (45) can (and should) be replaced by one in which the underlying basis is adapted to the poles of $G(z,\theta_\circ^n)$ via the study of the equivalent quantity

$$[\Gamma_m^\star(e^{j\omega}) \otimes I_2]M_n^{-1}\left(Z_\circ\Phi_u Z_\circ^\star + \delta|A_\circ|^2 I_2\right)[\Gamma_m(e^{j\omega}) \otimes I_2].$$

Now, by Lemma B.1 of the Appendix, the generalised Toeplitz matrix in the above expression is full rank even for $\delta = 0$ provided that $G(q,\theta_\circ^n)$ contains no pole-zero cancellations. Therefore, in any analysis in which the model order m of interest is anticipated to be less than or equal to the true one, then that same analysis should also assume that $M_n(Z_\circ\Phi_u Z_\circ^\star)$ is full rank.

With these ideas in mind, note that by the definition (5) of $M_n^{-1}(Z_\circ\Phi_u Z_\circ^\star)$ then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} Z_\circ^\star(e^{j\omega})[\Gamma_m^\star(e^{j\omega})\otimes I_2]M_n^{-1}(Z_\circ\Phi_u Z_\circ^\star)[\Gamma_m(e^{j\lambda})\otimes I_2]Z_\circ(e^{j\lambda})\Phi_u(\lambda)Z_\circ^\star(e^{j\lambda})[\Gamma_m^\star(e^{j\lambda})\otimes I_2]\,\mathrm{d}\lambda$$

$$= Z_\circ^\star(e^{j\omega})[\Gamma_m^\star(e^{j\omega}) \otimes I_2]M_n^{-1}(Z_\circ\Phi_u Z_\circ^\star)M_n(Z_\circ\Phi_u Z_\circ^\star) = Z_\circ^\star(e^{j\omega})[\Gamma_m^\star(e^{j\omega}) \otimes I_2].$$

Therefore, defining the (weighted by Φ_u) inner product between two functions $f, g \in X_{2m}$ as

$$\langle f, g\rangle \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)\overline{g(\omega)}\Phi_u(\omega)\,\mathrm{d}\omega \tag{47}$$

indicates that the function

$$\varphi_{2m}(\lambda,\omega) \triangleq Z_\circ^\star(e^{j\omega})[\Gamma_m^\star(e^{j\omega}) \otimes I_2]M_n^{-1}(Z_\circ\Phi_u Z_\circ^\star)[\Gamma_m(e^{j\lambda}) \otimes I_2]Z_\circ(e^{j\lambda}) \tag{48}$$

has the property that for any $f \in X_{2m}$

$$f(\omega) = \langle f(\lambda), \varphi_{2m}(\lambda,\omega)\rangle. \tag{49}$$

However, as shown in §2.4, the Reproducing Kernel for the space X_{2m} with respect to the inner product (47) is the unique function in X_{2m} with the property (49). Furthermore, the space X_{2m} can also be written as

$$X_{2m} = \text{Span}\left\{\mathcal{B}_0(z), \mathcal{B}_1(z), \cdots, \mathcal{B}_{m-1}(z), \mathcal{B}_m(z), \cdots, \mathcal{B}_{2m-1}(z)\right\}$$

where, since the poles of $\{\mathcal{B}_0, \cdots, \mathcal{B}_{m-1}\}$ are the poles of $G(z, \theta_\circ^n)$, then the poles of $\{\mathcal{B}_m, \cdots, \mathcal{B}_{2m}\}$ are taken to be those same poles repeated. Furthermore, again using the definition (5), this time with $p = 1$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_{2m}^\star(e^{j\omega}) M_n^{-1}(\Phi_u)\Gamma_{2m}(e^{j\lambda})\Gamma_{2m}^\star(e^{j\lambda})\Phi_u(\lambda)\,\mathrm{d}\lambda = \Gamma_{2m}^\star(e^{j\omega}).$$

Therefore, the function $\Gamma_{2m}^\star(\lambda)M_n^{-1}(\Phi_u)\Gamma_{2m}(\omega)$ also has the property (49). Consequently, by the uniqueness of the Reproducing Kernel, the expression

$$\varphi_{2m}(\lambda, \omega) = \Gamma_{2m}^\star(\lambda)M_n^{-1}(\Phi_u)\Gamma_{2m}(\omega)$$

is a formulation for that kernel which is alternative to (48). However, via Theorem 2.2

$$\lim_{m\to\infty} \frac{1}{\kappa_{2m}(\omega)}\Gamma_{2m}^\star(\omega)M_n^{-1}(\Phi_u)\Gamma_{2m}(\omega) = \frac{1}{\Phi_u(\omega)}. \tag{50}$$

Recognising that $\kappa_{2m}(\omega) = 2\kappa_m(\omega)$ since the poles in the basis definition are repeated, then establishes that

$$\lim_{m\to\infty} \frac{1}{\kappa_m(\omega)} Z_\circ^\star(e^{j\omega})[\Gamma_m^\star(e^{j\omega}) \otimes I_2]M_n^{-1}\left(Z_\circ\Phi_u Z_\circ^\star\right)[\Gamma_m(e^{j\omega}) \otimes I_2]Z_\circ(e^{j\omega}) = \frac{2}{\Phi_u(\omega)}. \tag{51}$$

Therefore, considering the analysis leading to (40), and provided that at a finite m of interest there are no pole-zero cancellations in $G(q, \theta_\circ^n)$, then (40) is most sensibly approximated for $\delta = 0$ as

$$\Delta_{2m}(\omega) \approx \frac{2\sigma^2}{\Phi_u(\omega)}\kappa_m(\omega) = \frac{2\sigma^2}{\Phi_u(\omega)}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}$$

which according to the principles discussed immediately after (42), leads to the new approximation (1) of this paper.

Now, as discussed in§5, the accuracy of the above approximation and hence that of the new quantification (1) depends on the length of the underlying Fourier reconstruction. Furthermore, the preceding argument has shown that the uniqueness of the Reproducing Kernel provides a means for re-expressing the complicated quantity (40) in a simpler form via the geometric principle of recognising it as a subspace invariant, and the subspace itself is exactly *twice* that of the model order m. Not only is this the reason for the factor of 2 appearing in the new approximation (1), it also implies that the underlying Fourier reconstruction is of length twice that of the model order m. This explains the perhaps unexpected accuracy of the 'low-order' quantifications illustrated in the simulation examples of [18].

In fact, when $\Phi_u(\omega) = \gamma$ a constant, then by the orthonormality of the basis $\{\mathcal{B}_k\}$ used to construct the matrix M_n, there is an exact equality

$$\varphi_{2m}(\omega, \omega) = \Gamma_{2m}^\star(\omega)M_n^{-1}(\gamma)\Gamma_{2m}(\omega) = \frac{1}{\gamma}\Gamma_{2m}^\star(\omega)\Gamma_{2m}(\omega) = \frac{2}{\gamma}\sum_{k=0}^{m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}.$$

As well as establishing the previously claimed result (46), this also indicates that in this case of white input excitation, the accuracy of the new variance quantification (1) is 'exact' for a finite true model order m, in the sense that it depends only on the data length N and not on the model order m.

While, for the sake of most clearly exposing these new reproducing kernel ideas, they have been developed for the simplified Output-Error case, the companion paper [18] has presented quantifications for the more sophisticated Box–Jenkins model situation. This required the employment of the following result, which is a generalisation of the preceding arguments, and therefore contains a formal presentation of them as a special case.

Theorem 6.1. *Consider the following factorisations*

$$\frac{G(z,\theta_\circ^n)}{A(z,\theta_\circ^n)}H^{-1}(z,\theta_\circ^n) = \frac{B_\dagger(z)}{A(z,\theta_\circ^n)A_\dagger(z)}H_\dagger^{-1}(z) \tag{52}$$

where $B_\dagger(z)$ is a polynomial, $H_\dagger(z)$ is rational and bi-proper and

$$A(z,\theta_\circ^n)A_\dagger(z) = (z-\xi_0)(z-\xi_1)\cdots(z-\xi_{2m-1}), \quad D(z,\theta_\circ^n)C(z,\theta_\circ^n) = (z-\eta_0)(z-\eta_1)\cdots(z-\eta_{2m-1}) \tag{53}$$

with all these zeros $\{\xi_k\}$ and $\{\eta_k\}$ are contained in the open unit disk $\mathbf{D}$ for any m. Use them to define the functions

$$\kappa_m(\omega) \triangleq \sum_{k=0}^{2m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}, \qquad \widetilde{\kappa}_m(\omega) \triangleq \sum_{k=0}^{2m-1}\frac{1-|\eta_k|^2}{|e^{j\omega}-\eta_k|^2}. \tag{54}$$

Then, provided that input assumptions 4.1 of [18] are satisfied and either $G(q,\theta_\circ^n)$ or $H(q,\theta_\circ^n)$ contain no pole zero cancellations for any model order m, and with $\varphi_m(\omega,\lambda)$ being defined as

$$\varphi_m(\lambda,\omega) \triangleq \frac{Z_\circ^\star(e^{j\omega})}{H(e^{-j\omega},\theta_\circ^n)}\,\Omega_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_4]T_n^{-1}\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right)[\Lambda_m(e^{j\lambda})\otimes I_4]\Omega_\circ(e^{j\lambda})\frac{Z_\circ(e^{j\lambda})}{H(e^{j\lambda},\theta_\circ^n)} \tag{55}$$

then

$$\lim_{m\to\infty}\varphi_m(\lambda,\omega)\cdot K_m^{-1}(\omega) = \begin{cases}\Phi_\zeta^{-1}(\omega) & ;\omega=\lambda\\ \varnothing & ;\omega\neq\lambda\end{cases} \tag{56}$$

where

$$K_m(\omega) \triangleq \begin{bmatrix}\kappa_m(\omega) & 0\\ 0 & \widetilde{\kappa}_m(\omega)\end{bmatrix}, \quad \Phi_\zeta(\omega) \triangleq \begin{bmatrix}\Phi_u(\omega) & \Phi_{ue}(\omega)\\ \overline{\Phi_{ue}(\omega)} & \sigma^2\end{bmatrix}.$$

Furthermore, in the special case of $\Phi_{ue}(\omega)=0$ then the non-asymptotic equality

$$\varphi_m(\omega,\omega) = \Phi_\zeta^{-1}(\omega)K_m(\omega) \tag{57}$$

holds for any m, except for the top left elements of the above 2×2 matrix valued functions. If, additionally, $\Phi_u(\omega)/H_\dagger(e^{j\omega})$ is a constant, then the equality in (57) hold for all elements.

Proof. See Appendix A. □

7 Convergence Rates for Variance Approximations

Central to the preceding discussions has been the principle that the convergence rate of (generalised) Fourier type reconstructions of functions depend critically on the 'smoothness' or 'variation' of those functions. Therefore, in the interests of minimising the error involved with using an asymptotic result at a finite truncation, the underlying basis should be chosen so as to maximise this smoothness; hence the introduction of the orthonormal basis (3) and associated generalised Toeplitz matrices.

The purpose of this section is to provide theoretical justification for this principle, together with an ensuing quantification of the convergence rates of variance quantifications.

To establish these results, the degree of 'variation' of a function will be characterised via its Lipschitz smoothness. Recall that a function f of a complex (or real) variable is 'Lipschitz continuous of degree α', which is denoted as $f \in \mathrm{Lip}(\alpha)$, if there exists a $C < \infty$ such that for some $\alpha > 0$ $|f(z) - f(\mu)| \le C|z-\mu|^\alpha$ for all z, μ.

Using this characterisation, according to Theorem C.1 it is possible to establish an upper bound on the rate of convergence of the generalised Fourier series defined in Theorem 2.2 which, as explained in the previous section, is central to the proof of Theorem 6.1 and which in turn, is the fundamental result underlying the work of the companion paper [18]. This then allows a quantification of the convergence rate of the variance error quantifications provided here and in [18].

Corollary 7.1. *Suppose that $\widehat{\theta}_N^n$ is calculated via the prediction error method defined in (6) of [18] and using the model structure (14). Then*

$$\sqrt{N}\begin{pmatrix} G(e^{j\omega},\widehat{\theta}_N^n) - G(e^{j\omega},\theta_\circ^n) \\ H(e^{j\omega},\widehat{\theta}_N^n) - H(e^{j\omega},\theta_\circ^n)\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_m(\omega)) \qquad \text{as } N\to\infty \tag{58}$$

where, provided that

1. *$\{u_t\}$ satisfies input-assumption 4.1 of [18];*
2. *Either $G(z,\theta_\circ^n)$ or $H(z,\theta_\circ^n)$ (or both) contains no pole-zero cancellation for the finite model order m of interest;*
3. *For this same finite m it holds that $\varepsilon_t(\theta_\circ^n) = e_t$ where $\{e_t\}$ is 'white noise' satisfying the assumptions of Corollary 4.1 of [18];*
4. *The cross spectrum $\Phi_{ue}(\omega) = 0$;*
5. *Under the factorisation*

$$\frac{G(z,\theta_\circ^n)}{A(z,\theta_\circ^n)}H^{-1}(z,\theta_\circ^n) = \frac{B_\dagger(z)}{A(z,\theta_\circ^n)A_\dagger(z)}H_\dagger^{-1}(z) \tag{59}$$

where $B_\dagger(z)$ is a polynomial, $H_\dagger(z)$ is rational and bi-proper and

$$A(z,\theta_\circ^n)A_\dagger(z) = (z-\xi_0)(z-\xi_1)\cdots(z-\xi_{2m-1}), \tag{60}$$

$$D(z,\theta_\circ^n)C(z,\theta_\circ^n) = (z-\eta_0)(z-\eta_1)\cdots(z-\eta_{2m-1}) \tag{61}$$

then the zeros $\{\xi_k\}$ and $\{\eta_k\}$ are contained in the open unit disk $\mathbf{D}$,

6. *For some $\alpha > 0$*

$$\frac{\Phi_u(\omega)}{|H_\dagger(e^{j\omega})|^2} \in \mathrm{Lip}(\alpha), \tag{62}$$

Then for some $C < \infty$ that does not depend on m,

$$\left|\Sigma_m(\omega)K_m^{-1}(\omega) - \sigma^2|H(e^{j\omega},\theta_\circ^n)|^2\begin{bmatrix}\Phi_u(\omega) & 0 \\ 0 & \sigma^2\end{bmatrix}^{-1}\right| \le C\frac{\log^4 m}{m^{\alpha/(2+\alpha)}} \tag{63}$$

where the above is to be interpreted component-wise in the 2×2 matrices involved and

$$K_m(\omega) \triangleq \begin{bmatrix}\kappa_m(\omega) & 0 \\ 0 & \widetilde{\kappa}_m(\omega)\end{bmatrix}, \qquad \kappa_m(\omega) \triangleq \sum_{k=0}^{2m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}, \qquad \widetilde{\kappa}_m(\omega) \triangleq \sum_{k=0}^{2m-1}\frac{1-|\eta_k|^2}{|e^{j\omega}-\eta_k|^2}. \tag{64}$$

Proof. The convergence (58) has already been established in Theorem 4.1 of [18]. Furthermore, under the condition of this Theorem that $\varepsilon_t(\theta_\circ^n) = e_t$, the proof of Theorem 4.1 of [18] establishes at equation A.9 of [18] that

$$\Sigma_m(\omega) = \sigma^2 |H(e^{j\omega}), \theta_\circ^n|^2 \varphi_m(\omega,\omega) = \sigma^2 |H(e^{j\omega}), \theta_\circ^n|^2 \begin{bmatrix} \varphi_m^{11}(\omega,\omega) & \varphi_m^{12}(\omega,\omega) \\ \varphi_m^{21}(\omega,\omega) & \varphi_m^{22}(\omega,\omega) \end{bmatrix} \quad (65)$$

with $\varphi(\lambda,\omega)$ being defined in (55) of Theorem 6.1. Furthermore, according to the assumption here that $\Phi_{ue}(\omega) = 0$, the reproducing kernel argument between equation (A.15) and (A.19) in the proof of Theorem 6.1 establishes the re-formulation

$$\varphi_m^{11}(\omega,\omega) = \frac{1}{|H_\dagger(e^{j\omega})|^2} \Gamma_{2m}^\star(e^{j\omega}) M_{2m}^{-1}\left(\frac{\Phi_u(\omega)}{|H_\dagger(e^{j\omega})|^2}\right) \Gamma_{2m}(e^{j\omega}). \quad (66)$$

Furthermore, by Corollary 5.2 of [21], there exists a $C < \infty$ such that

$$\frac{1}{\kappa_m(\omega)} \left| \Gamma_{2m}^\star(e^{j\omega}) \left[M_{2m}^{-1}\left(\frac{\Phi_u(\omega)}{|H_\dagger(e^{j\omega})|^2}\right) - M_{2m}\left(\frac{|H_\dagger(e^{j\omega})|^2}{\Phi_u(\omega)}\right)\right] \Gamma_{2m}(e^{j\omega}) \right| \le C \frac{\log^4 m}{m^{\alpha/(2+\alpha)}}. \quad (67)$$

Combining this with Theorem C.1 then establishes that

$$\left| \varphi_m^{11}(\omega,\omega) - \frac{1}{\Phi_u(\omega)} \right| \le C \frac{\log^4 m}{m^{\alpha/(2+\alpha)}}. \quad (68)$$

Via an identical argument

$$\left| \varphi_m^{22}(\omega,\omega) - \frac{1}{\sigma^2} \right| \le C \frac{\log^4 m}{m^{\alpha/(2+\alpha)}} \quad (69)$$

while under the assumption that $\Phi_{ue}(\omega) = 0$ equations (A.25) and (A.26) establish that $\varphi_m^{12}(\omega,\omega) = \varphi_m^{21}(\omega,\omega) = 0$. □

Therefore, the accuracy of the new variance quantifications of this paper and [18] increases monotonically with increasing m, and at a hyperbolic rate that increases with increasing smoothness of $\Phi_u/|H_\dagger|^2$. As remarked and illustrated in [18], often the factorisation (59) is not unique, and hence Corollary 7.1 indicates that when there is a choice of factorisations, the one maximising the Lipschitz smoothness of $\Phi_u/|H_\dagger|^2$ should be made. This was illustrated via simulation in §5 of [18] in figure 10(b).

Note that there are no such corresponding convergence rate quantifications available for pre-existing variance error quantifications such as (2). Part of the reason for this is that, although smoothness of underlying spectral densities is a sufficient condition for accurate variance quantification it is also, as the following result establishes, a necessary condition, and for the expression (2), the quantities involved generically violate this necessary condition.

Corollary 7.2. *Under the same conditions and definitions as made in Corollary 7.1, there exists a constant $C < \infty$ that is independent of m such that*

$$\left| \Sigma(\omega) m^{-1} - 2\sigma^2 |H(e^{j\omega}, \theta_\circ^n)|^2 \begin{bmatrix} \Phi_u(\omega) & 0 \\ 0 & \sigma^2 \end{bmatrix}^{-1} \right| \le \frac{C}{m^\alpha} \quad (70)$$

for some $\alpha > 0$ only if there exists a $K < \infty$ that is independent of m such that

$$|A(e^{j\omega}, \theta_\circ^n)|^2, |D(e^{j\omega}, \theta_\circ^n)|^4 \ge K \quad (71)$$

and if

$$\frac{\Phi_u(\omega)}{|A(e^{j\omega},\theta_\circ^n)H(e^{j\omega},\theta_\circ^n)|^2}, \frac{1}{|C(e^{j\omega},\theta_\circ^n)D(e^{j\omega},\theta_\circ^n)|^2} \in \mathrm{Lip}(\alpha). \tag{72}$$

Furthermore, there exists a constant $C < \infty$ that is independent of m such that

$$\left|\Sigma(\omega)K_m^{-1}(\omega) - \sigma^2|H(e^{j\omega},\theta_\circ^n)|^2 \begin{bmatrix} \Phi_u(\omega) & 0 \\ 0 & \sigma^2 \end{bmatrix}^{-1}\right| \le \frac{C}{m^\alpha} \tag{73}$$

for some $\alpha > 0$ only if there exists a $K < \infty$ that is independent of m such that

$$|H(e^{j\omega},\theta_\circ^n)|^2 \ge K \quad \textit{and} \quad \frac{\Phi_u(\omega)}{|H_\dagger(e^{j\omega})|^2} \in \mathrm{Lip}(\alpha). \tag{74}$$

Proof. Considering the top left element $\Sigma_m^{11}(\omega)$ of $\Sigma_m(\omega)$ first. Then according to (65), $\Sigma_m^{11}(\omega) = \sigma^2|H(e^{j\omega},\theta_\circ^n)|^2\varphi^{11}(\omega,\omega)$ with $\varphi^{11}(\lambda,\omega)$ being defined in (55) of Theorem 6.1. Furthermore, according to the assumption here that $\Phi_{ue}(\omega) = 0$, the argument between equation (A.15) and (A.19) establishes that $\varphi^{11}(\lambda,\omega)$ is the reproducing kernel for the space

$$X_m = \mathrm{Span}\left\{\frac{1}{A_\circ^2(z)}H_\circ^{-1}(z), \frac{z}{A_\circ^2(z)}H_\circ^{-1}(z), \cdots, \frac{z^{2m-1}}{A_\circ^2(z)}H_\circ^{-1}(z)\right\}$$

and with respect to the inner-product weighting Φ_u. However, by the definition of the Toeplitz matrix formulation (7)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Lambda_{2m}(e^{j\lambda})}{A_\circ(e^{j\lambda})H_\circ(e^{j\lambda})}\Phi_u(\lambda)\frac{\Lambda_{2m}^\star(e^{j\lambda})}{A_\circ(e^{-j\lambda})H_\circ(e^{-j\lambda})}T_n^{-1}\left(\frac{\Phi_u}{|A_\circ H_\circ|^2}\right)\frac{\Lambda_{2m}(e^{j\omega})}{A_\circ(e^{j\omega})H_\circ(e^{j\omega})}\,\mathrm{d}\lambda = \frac{\Lambda_{2m}(e^{j\omega})}{A_\circ(e^{j\omega})H_\circ(e^{j\omega})} \tag{75}$$

so that

$$\overline{\varphi_m^{11}(\lambda,\omega)} = \frac{\Lambda_{2m}^\star(e^{j\lambda})}{A_\circ(e^{-j\lambda})H_\circ(e^{-j\lambda})}T_n^{-1}\left(\frac{\Phi_u}{|A_\circ H_\circ|^2}\right)\frac{\Lambda_{2m}(e^{j\omega})}{A_\circ(e^{j\omega})H_\circ(e^{j\omega})}$$

is a re-formulation for that reproducing kernel. Therefore

$$\begin{aligned}\left|\frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|^2}{\Phi_u(\omega)} - \frac{\Sigma_m^{11}(\omega)}{2m}\right| &= \left|\frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|^2}{\Phi_u(\omega)} - \frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|^2\varphi_m^{11}(\omega,\omega)}{2m}\right| \\ &= \frac{\sigma^2}{|A(e^{j\omega},\theta_\circ^n)|^2}\times \\ &\quad \left|\frac{|A_\circ(e^{j\omega})H_\circ(e^{j\omega})|^2}{\Phi_u(\omega)} - \frac{1}{2m}\Lambda_{2m}^\star(e^{j\omega})T_n^{-1}\left(\frac{\Phi_u}{|A_\circ H_\circ|^2}\right)\Lambda_{2m}(e^{j\omega})\right|.\end{aligned}$$

However, by Theorem C.2 with $\xi_k = 0$ for all k

$$\left|\frac{|A_\circ(e^{j\omega})H_\circ(e^{j\omega})|^2}{\Phi_u(\omega)} - \frac{1}{2m}\Lambda_{2m}^\star(e^{j\omega})T_n^{-1}\left(\frac{\Phi_u}{|A_\circ H_\circ|^2}\right)\Lambda_{2m}(e^{j\omega})\right| \le \frac{C}{m^\alpha}$$

for some C independent of m only if $\Phi_u/|A_\circ H_\circ|^2 \in \mathrm{Lip}(\alpha)$. On the other hand, it has already been noted in (66) that another representation for the reproducing kernel is

$$\overline{\varphi_m^{11}(\lambda,\omega)} = \frac{1}{H_\dagger(e^{-j\lambda})H_\dagger(e^{j\omega})}\Gamma_{2m}^\star(e^{\lambda})M_{2m}^{-1}\left(\frac{\Phi_u(\omega)}{|H_\dagger(e^{j\omega})|^2}\right)\Gamma_{2m}(e^{j\omega}). \tag{76}$$

so that

$$\begin{aligned}\left|\frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|}{\Phi_u(\omega)} - \frac{\Sigma_m^{11}(\omega)}{\kappa_m(\omega)}\right| &= \left|\frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|}{\Phi_u(\omega)} - \frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|\varphi_m^{11}(\omega,\omega)}{\kappa_m(\omega)}\right| \\ &= \sigma^2|H(e^{j\omega},\theta_\circ^n)|^2 \times \\ &\quad \left|\frac{1}{\Phi_u(\omega)} - \frac{1}{|H_\dagger(e^{j\omega})|^2\kappa_m(\omega)}\Gamma_{2m}^\star(e^{j\omega})M_{2m}^{-1}\left(\frac{\Phi_u(\omega)}{|H_\dagger(e^{j\omega})|^2}\right)\Gamma_{2m}(e^{j\omega})\right|\end{aligned}$$

Application of Theorem C.2 with the ξ_k defined via the factorisations (59),(60) then completes the proof for $\Sigma_m^{11}(\omega)$. Similarly, considering the bottom right term $\Sigma_m^{22}(\omega)$ of $\Sigma_m(\omega)$, according to (65) $\Sigma_m^{22}(\omega) = \sigma^2|H(e^{j\omega},\theta_\circ^n)|^2\varphi^{22}(\omega,\omega)$ with $\varphi^{22}(\lambda,\omega)$ also being defined in (55) of Theorem 6.1. Furthermore, according to the assumption here that $\Phi_{ue}(\omega) = 0$, the argument between equation (A.22) to (A.22) establishes that $\sigma^2\overline{\varphi^{22}(\lambda,\omega)}$ is the reproducing kernel for the space

$$\widetilde{X}_m = \text{Span}\left\{\frac{1}{D_\circ^n(z)C_\circ^n(z)}, \frac{z}{D_\circ^n(z)C_\circ^n(z)}, \cdots, \frac{z^{2m-1}}{D_\circ^n(z)C_\circ^n(z)}\right\}.$$

Furthermore

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Lambda_{2m}(e^{j\lambda})}{C_\circ(e^{j\lambda})D_\circ(e^{j\lambda})}\frac{\Lambda_{2m}^\star(e^{j\lambda})}{C_\circ(e^{-j\lambda})D_\circ(e^{-j\lambda})}T_n^{-1}\left(\frac{1}{|C_\circ D_\circ|^2}\right)\frac{\Lambda_{2m}(e^{j\omega})}{C_\circ(e^{j\omega})D_\circ(e^{j\omega})}\,\mathrm{d}\lambda = \frac{\Lambda_{2m}(e^{j\omega})}{C_\circ(e^{j\omega})D_\circ(e^{j\omega})} \tag{77}$$

so that

$$\overline{\varphi_m^{22}(\lambda,\omega)} = \frac{\Lambda_{2m\sigma^2}^\star(e^{j\lambda})}{C_\circ(e^{-j\lambda})D_\circ(e^{-j\lambda})}T_n^{-1}\left(\frac{1}{|C_\circ D_\circ|^2}\right)\frac{\Lambda_{2m}(e^{j\omega})}{C_\circ(e^{j\omega})D_\circ(e^{j\omega})}$$

is a re-formulation for that reproducing kernel. Therefore

$$\begin{aligned}\left||H(e^{j\omega},\theta_\circ^n)|^2 - \frac{\Sigma_m^{22}(\omega)}{2m}\right| &= \left||H(e^{j\omega},\theta_\circ^n)|^2 - \frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|^2\varphi_m^{22}(\omega,\omega)}{2m}\right| \\ &= \frac{\sigma^2}{|D(e^{j\omega},\theta_\circ^n)|^4} \times \\ &\quad \left||C_\circ(e^{j\omega})D_\circ(e^{j\omega})|^2 - \frac{1}{2m}\Lambda_{2m}^\star(e^{j\omega})T_n^{-1}\left(\frac{1}{|C_\circ D_\circ|^2}\right)\Lambda_{2m}(e^{j\omega})\right|.\end{aligned}$$

However, by Theorem C.2 with $\xi_k = 0$ for all k

$$\left||C_\circ(e^{j\omega})D_\circ(e^{j\omega})|^2 - \frac{1}{2m}\Lambda_{2m}^\star(e^{j\omega})T_n^{-1}\left(\frac{1}{|C_\circ D_\circ|^2}\right)\Lambda_{2m}(e^{j\omega})\right| \leq \frac{C}{m^\alpha}$$

for some C independent of m only if $1/|C_\circ D_\circ|^2 \in \text{Lip}(\alpha)$. On the other hand, using the basis (A.23), another formulation for the reproducing kernel is

$$\sigma^2\overline{\varphi_m^{22}(\lambda,\omega)} = \widetilde{\Gamma}_{2m}^\star(e^\lambda)\widetilde{\Gamma}_{2m}(e^{j\omega}).$$

so that

$$\begin{aligned}\left||H(e^{j\omega},\theta_\circ^n)|^2 - \frac{\Sigma_m^{22}(\omega)}{\widetilde{\kappa}_m(\omega)}\right| &= \left||H(e^{j\omega},\theta_\circ^n)|^2 - \frac{\sigma^2|H(e^{j\omega},\theta_\circ^n)|^2\varphi_m^{22}(\omega,\omega)}{\widetilde{\kappa}_m(\omega)}\right| \\ &= |H(e^{j\omega},\theta_\circ^n)|^2\left|1 - \frac{\widetilde{\Gamma}_{2m}^\star(e^{j\omega})\widetilde{\Gamma}_{2m}(e^{j\omega})}{\widetilde{\kappa}_m(\omega)}\right| = 0.\end{aligned}$$

□

There are several important points to emerge from this result. Firstly, notice that (70) examines the accuracy of the variance approximation

$$\mathrm{Cov}\left[\begin{array}{c} G(e^{j\omega},\widehat{\theta}_N^n) \\ H(e^{j\omega},\widehat{\theta}_N^n)\end{array}\right] \approx 2\sigma^2\frac{m}{N}|H(e^{j\omega},\theta_\circ^n)|\Phi_\zeta^{-1}(\omega) \tag{78}$$

which differs from the pre-existing one of [14, 13], of which (2) is a special Output-Error case, by a factor of two. The reason for this is that without the factor of two, it appears to these authors that no conditions at all can be imposed under which an expression of the form $\sigma^2(m/N)H(e^{j\omega},\theta_\circ^n)|\Phi_\zeta^{-1}(\omega)$ can be expected to be a good variance approximation for large m.

Secondly, even with this factor of two, the *necessary* conditions (71), (72) for the accuracy of (78) seem impossible to satisfy, except for the special case of Output-Error model structure, will all poles beyond a fixed order $m_\star \ll m$ being at the origin. See §4 and especially figure 1 for further detail on this point.

By way of contrast, for the new approximation introduced in the companion paper [18] and studied here of

$$\mathrm{Cov}\left[\begin{array}{c} G(e^{j\omega},\widehat{\theta}_N^n) \\ H(e^{j\omega},\widehat{\theta}_N^n)\end{array}\right] \approx \frac{\sigma^2}{N}|H(e^{j\omega},\theta_\circ^n)|\Phi_\zeta^{-1}(\omega)K_m(\omega) \tag{79}$$

the necessary conditions (74) for the accuracy of (79) are quite easy to satisfy. For example, for the Output Error case, $K=1$, then $\Phi_u(\omega)\in\mathrm{Lip}(\alpha)$ is all that is required. Furthermore, according to the previous Theorem 7.1, this will then also be a sufficient condition for high accuracy of (79) at high model order m.

8 Conclusion

This paper has introduced certain new ideas and results that, while apparently quite abstract in isolation, were central to the improved estimation-error quantifications of the companion work, and are likely to have further application in other system identification and system theoretic scenarios. The central idea is to recognise that inverse quadratic forms in Toeplitz matrices are reproducing kernels for certain Hilbert spaces, and that since such kernels are unique, coupled with the fact that simpler representations for the kernels can be derived, then leads directly to the ability to quantify the quadratic forms, and hence to quantify estimation error.

The significance of these improved variance quantifications stems from the significance already enjoyed by the pre-existing ones. This indicates that there is now scope for further work examining the ramifications of the new results in the areas where the pre-existing ones have already been applied, such as clarifying the debate over equivalences between closed loop estimation errors, and experiment design for identification supporting closed loop control applications.

A Proof of Theorem 6.1

Proof. First, for compactness define

$$\Psi(e^{j\omega}) \triangleq [\Lambda_m(e^{j\omega})\otimes I_4]\Omega(e^{j\omega},\theta_\circ^n) \tag{A.1}$$

so that according to Definition 2.2

$$T_n\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{j\lambda})\frac{Z_\circ(e^{j\lambda})\Phi_\zeta(\lambda,\theta_\circ^n)Z_\circ^\star(e^{j\lambda})}{|H_\circ^n(e^{j\lambda})|^2}\Psi^\star(e^{j\lambda})\mathrm{d}\lambda$$

and therefore

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{j\lambda})\frac{Z_\circ(e^{j\lambda})\Phi_\zeta(\lambda,\theta_\circ^n)Z_\circ^\star(e^{j\lambda})}{|H_\circ^n(e^{j\lambda})|^2}\Psi^\star(e^{j\lambda})T_n^{-1}\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right)\Psi(\omega)\frac{Z(e^{j\omega})}{H_\circ^n(e^{j\omega})}\,\mathrm{d}\lambda = \Psi(\omega)\frac{Z_\circ(e^{j\omega})}{H_\circ^n(e^{j\omega})}. \tag{A.2}$$

Consequently, putting

$$\overline{\varphi_m(\lambda,\omega)} = \begin{bmatrix} \overline{\varphi_m^{11}(\lambda,\omega)} & \overline{\varphi_m^{12}(\lambda,\omega)} \\ \varphi_m^{21}(\lambda,\omega) & \varphi_m^{22}(\lambda,\omega) \end{bmatrix} \triangleq \frac{Z_\circ^\star(e^{j\lambda})}{H_\circ^n(e^{j\lambda})}\Psi^\star(e^{j\lambda})T_n^{-1}\left(\Omega_\circ\frac{Z_\circ\Phi_\zeta Z_\circ^\star}{|H_\circ^n|^2}\Omega_\circ^\star\right)\Psi(\omega)\frac{Z(e^{j\omega})}{H_\circ^n(e^{j\omega})} \tag{A.3}$$

and

$$\Phi_\zeta(\omega) = \begin{bmatrix} \overline{\Phi_u(\omega)} & \Phi_{u\varepsilon}(\omega) \\ \Phi_{u\varepsilon}(\omega) & \Phi_\varepsilon(\omega) \end{bmatrix} \tag{A.4}$$

then indicates that according to (A.2)

$$\Psi(e^{j\omega})\frac{Z_\circ(e^{j\omega})}{H_\circ^n(e^{j\omega})} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{j\lambda})\frac{Z_\circ(e^{j\lambda})}{H_\circ^n(e^{j\lambda})}\begin{bmatrix} \overline{\Phi_u(\lambda)} & \Phi_{u\varepsilon}(\lambda) \\ \Phi_{u\varepsilon}(\lambda) & \Phi_\varepsilon(\lambda) \end{bmatrix}\begin{bmatrix} \overline{\varphi_m^{11}(\lambda,\omega)} & \overline{\varphi_m^{12}(\lambda,\omega)} \\ \varphi_m^{21}(\lambda,\omega) & \varphi_m^{22}(\lambda,\omega) \end{bmatrix}\mathrm{d}\lambda. \tag{A.5}$$

Furthermore, note that

$$\begin{aligned}\Psi(z)\frac{Z_\circ(z)}{H_\circ^n(z)} &= [\Lambda_m(z)\otimes I_4]\begin{bmatrix} A_\circ^n(z)^{-1}I_2 & \varnothing \\ \varnothing & D_\circ^n(z,\theta^n)^{-1}I_2 \end{bmatrix}\begin{bmatrix} -G_\circ^n(z) & 0 \\ 1 & 0 \\ 0 & -H_\circ^n(z) \\ 0 & 1 \end{bmatrix}\frac{1}{H_\circ^n(z)} \\ &= \begin{bmatrix} -\frac{G_\circ^n(z)}{A_\circ^n(z)H_\circ^n(z)} & 0 \\ \frac{1}{A_\circ^n(z)H_\circ^n(z)} & 0 \\ 0 & -\frac{1}{D_\circ^n(z)} \\ 0 & \frac{1}{C_\circ^n(z)} \\ \vdots & \vdots \\ -\frac{z^{m-1}G_\circ^n(z)}{A_\circ^n(z)H_\circ^n(z)} & 0 \\ \frac{z^{m-1}}{A_\circ^n(z)H_\circ^n(z)} & 0 \\ 0 & -\frac{z^{m-1}}{D_\circ^n(z)} \\ 0 & \frac{z^{m-1}}{C_\circ^n(z)} \end{bmatrix}.\end{aligned} \tag{A.6}$$

Therefore, according to (A.5), for all $k = 0, 1, 2, \cdots, 2m-1$

$$\left\langle \frac{e^{j\lambda k}D_\circ^n(e^{j\lambda})}{A_\circ^n(e^{j\lambda})^2C_\circ^n(e^{j\lambda})}, \Phi_u(\lambda)\left(\varphi_m^{11}(\lambda,\omega) + \varphi_m^{21}(\lambda,\omega)\frac{\overline{\Phi_{u\varepsilon}(\lambda)}}{\Phi_u(\lambda)}\right)\right\rangle = \frac{e^{j\omega k}D_\circ^n(e^{j\lambda})}{A_\circ^n(e^{j\omega})^2C_\circ^n(e^{j\omega})} \tag{A.7}$$

$$\left\langle \frac{e^{j\lambda k}}{C_\circ^n(e^{j\lambda})D_\circ^n(e^{j\lambda})^2}, \Phi_\varepsilon(\lambda)\left(\varphi_m^{22}(\lambda,\omega) + \varphi_m^{12}(\lambda,\omega)\frac{\Phi_{u\varepsilon}(\lambda)}{\Phi_\varepsilon(\lambda)}\right)\right\rangle = \frac{e^{j\omega k}}{C_\circ^n(e^{j\omega})D_\circ^n(e^{j\omega})^2} \tag{A.8}$$

$$\left\langle \frac{e^{j\lambda k} D_\circ^n(e^{j\lambda})}{A_\circ^n(e^{j\lambda})^2 C_\circ^n(e^{j\lambda})}, \Phi_u(\lambda) \left(\varphi_m^{12}(\lambda,\omega) + \varphi_m^{22}(\lambda,\omega) \frac{\overline{\Phi_{u\varepsilon}(\lambda)}}{\Phi_u(\lambda)} \right) \right\rangle = 0 \tag{A.9}$$

$$\left\langle \frac{e^{j\lambda k}}{C_\circ^n(e^{j\lambda}) D_\circ^n(e^{j\lambda})^2}, \Phi_\varepsilon(\lambda) \left(\varphi_m^{21}(\lambda,\omega) + \varphi_m^{11}(\lambda,\omega) \frac{\Phi_{u\varepsilon}(\lambda)}{\Phi_\varepsilon(\lambda)} \right) \right\rangle = 0. \tag{A.10}$$

Now, consider the linear subspace of $H_2^\perp(\mathbf{T})$:

$$X_m \triangleq \text{Span}\left\{ \frac{D_\circ^n(z)}{A_\circ^n(z)^2 C_\circ^n(z)}, \frac{z D_\circ^n(z)}{A_\circ^n(z)^2 C_\circ^n(z)}, \cdots, \frac{z^{2m-1} D_\circ^n(z)}{A_\circ^n(z)^2 C_\circ^n(z)} \right\} \tag{A.11}$$

$$= \text{Span}\left\{ \frac{1}{A_\circ^n(z) A_\dagger(z)} H_\dagger^{-1}(z), \frac{z}{A_\circ^n(z) A_\dagger(z)} H_\dagger^{-1}(z), \cdots, \frac{z^{2m-1}}{A_\circ^n(z) A_\dagger(z)} H_\dagger^{-1}(z) \right\} \tag{A.12}$$

Then by (A.10)

$$\varphi_m^{21}(\lambda,\omega) = -\varphi_m^{11}(\lambda,\omega) \frac{\Phi_{u\varepsilon}(\lambda)}{\Phi_\varepsilon(\lambda)} + r_m(\omega) \tag{A.13}$$

where

$$\lim_{m\to\infty} \|r_m(\omega)\|_2 = 0. \tag{A.14}$$

since, by the assumptions of the theorem, $\sum_{k=0}^{\infty}(1 - |\xi_k|^2) = \infty$, which according to [17] is a sufficient condition for $\lim_{m\to\infty} X_m = H_2^\perp(\mathbf{T})$.

Substituting (A.13) into (A.7) then yields

$$\left\langle \frac{e^{j\lambda k} D_\circ^n(e^{j\lambda})}{A_\circ^n(e^{j\lambda})^2 C_\circ^n(e^{j\lambda})}, \varphi_m^{11}(\lambda,\omega) \left(\frac{\Phi_u(\lambda)\Phi_\varepsilon(\lambda) - |\Phi_{u\varepsilon}(\lambda)|^2}{\Phi_\varepsilon(\lambda)} \right) \right\rangle +$$

$$\left\langle \frac{e^{j\lambda k} D_\circ^n(e^{j\lambda})}{A_\circ^n(e^{j\lambda})^2 C_\circ^n(e^{j\lambda})}, \Phi_u(\lambda)\ r_m(\lambda) \right\rangle = \frac{e^{j\omega k} D_\circ^n(e^{j\omega})}{A_\circ^n(e^{j\omega})^2 C_\circ^n(e^{j\omega})} \tag{A.15}$$

where, by the Cauchy–Schwartz inequality and (A.14)

$$\lim_{n\to\infty} \left| \left\langle \frac{e^{j\lambda k} D_\circ^n(e^{j\lambda})}{A_\circ^n(e^{j\lambda})^2 C_\circ^n(e^{j\lambda})}, \Phi_u(\lambda)\ r_m(\lambda) \right\rangle \right| = 0. \tag{A.16}$$

Therefore, since (A.7) holds for any $k = 0, 1, 2, \cdots, 2m-1$, then for large enough m, $\overline{\varphi_m^{11}(\lambda,\omega)}$ becomes arbitrarily close to the reproducing kernel for the space X_m and with respect to the (weighted) inner product

$$\langle f, g \rangle \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) \overline{g(\lambda)} \left(\frac{\Phi_\varepsilon(\lambda)\Phi_u(\lambda) - |\Phi_{u\varepsilon}(\lambda)|^2}{\Phi_\varepsilon(\lambda)} \right) \mathrm{d}\lambda. \tag{A.17}$$

However, with the choice of ξ_k as the zeros of $A_\circ^n(z)A_\dagger(z)$, then the space X_m can also be represented with respect to the basis

$$\mathcal{B}_k(z) = \frac{\sqrt{1 - |\xi_k|^2}}{z - \xi_k} \left(\prod_{\ell=0}^{k-1} \frac{1 - \overline{\xi_\ell} z}{z - \xi_\ell} \right) \tag{A.18}$$

as

$$X_m = \text{Span}\left\{ \mathcal{B}_0(z) H_\dagger^{-1}(z), \mathcal{B}_1(z) H_\dagger^{-1}(z), \cdots, \mathcal{B}_{2m-1}(z) H_\dagger^{-1}(z) \right\}.$$

Furthermore, as established in §6 of this paper, by using the definition (5) of a generalised Toeplitz matrix M_n, then the reproducing kernel $\overline{\varphi_m^{11}(\lambda,\omega)}$ for X_m with respect to the inner product (A.17) can also be written using the definition $\Gamma_{2m}(z) \triangleq [\mathcal{B}_0(z), \cdots, \mathcal{B}_{2m-1}(z)]^T$ as

$$\overline{\varphi_m^{11}(\lambda,\omega)} = \overline{H_\dagger^{-1}(e^{j\lambda})}\Gamma_{2m}^\star(e^{j\lambda})M_{2m}^{-1}\left(\frac{\Phi_\varepsilon\Phi_u - |\Phi_{u\varepsilon}|^2}{\Phi_\varepsilon|H_\dagger|^2}\right)\Gamma_{2m}(e^{j\omega})H_\dagger^{-1}(e^{j\omega}). \tag{A.19}$$

Furthermore, according to the results of Theorem 2.2

$$\lim_{m\to\infty}\frac{1}{\kappa_m(\omega)}\Gamma_{2m}^\star(e^{j\lambda})M_{2m}^{-1}\left(\frac{\Phi_\varepsilon\Phi_u - |\Phi_{u\varepsilon}|^2}{\Phi_\varepsilon|H_\dagger|^2}\right)\Gamma_{2m}(e^{j\omega}) = \begin{cases} \dfrac{\Phi_\varepsilon(\omega)|H_\dagger(e^{j\omega})|^2}{\Phi_\varepsilon(\omega)\Phi_u(\omega) - |\Phi_{u\varepsilon}(\omega)|^2} & ;\lambda = \omega \\ 0 & ;\lambda \neq \omega \end{cases}$$

where

$$\kappa_m(\omega) \triangleq \sum_{k=0}^{2m-1}|\mathcal{B}_k(e^{j\omega})|^2 = \sum_{k=0}^{2m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}.$$

Therefore

$$\lim_{m\to\infty}\frac{\overline{\varphi_m^{11}(\lambda,\omega)}}{\kappa_m(\omega)} = \begin{cases} \dfrac{\Phi_\varepsilon(\omega)}{\Phi_\varepsilon(\omega)\Phi_u(\omega) - |\Phi_{u\varepsilon}(\omega)|^2} & ;\lambda = \omega \\ 0 & ;\lambda \neq \omega \end{cases} \tag{A.20}$$

Similarly, by shifting attention to the space

$$\widetilde{X}_m = \text{Span}\left\{\frac{1}{D_\circ^n(z)C_\circ^n(z)}, \frac{z}{D_\circ^n(z)C_\circ^n(z)}, \cdots, \frac{z^{2m-1}}{D_\circ^n(z)C_\circ^n(z)}\right\}$$

then from (A.9)

$$\varphi_m^{12}(\lambda,\omega) = -\varphi_m^{22}(\lambda,\omega)\frac{\overline{\Phi_{u\varepsilon}(\lambda)}}{\Phi_u(\lambda)} + r_m(\omega) \tag{A.21}$$

where, since $\sum_{k=0}^{\infty}(1-|\eta_k|^2) = \infty$, then $\lim_{m\to\infty}\widetilde{X}_m = H_2(\mathbf{T})^\perp$ so that $\lim_{m\to\infty}\|r_m(\omega)\|_2 = 0$, which when substituted into (A.8) yields

$$\left\langle \frac{e^{j\lambda k}}{D_\circ^n(e^{j\lambda})C_\circ^n(e^{j\lambda})}, \varphi_m^{22}(\lambda,\omega)\left(\frac{\Phi_\varepsilon(\lambda)\Phi_u(\lambda) - |\Phi_{u\varepsilon}(\lambda)|^2}{\Phi_u(\lambda)}\right)\right\rangle +$$

$$\left\langle \frac{e^{j\lambda k}}{D_\circ^n(e^{j\lambda})C_\circ^n(e^{j\lambda})}, \Phi_\varepsilon(\lambda)\, r_m(\omega)\right\rangle = \frac{e^{j\omega k}}{D_\circ^n(e^{j\omega})C_\circ^n(e^{j\omega})} \qquad ;k = 0,1,\cdots,2m-1- \tag{A.22}$$

with

$$\lim_{m\to\infty}\left|\left\langle \frac{e^{j\lambda k}}{D_\circ^n(e^{j\lambda})C_\circ^n(e^{j\lambda})}, \Phi_\varepsilon(\lambda)\, r_m(\omega)\right\rangle\right| = 0$$

so that for large enough m, then $\overline{\varphi_m^{22}(\lambda,\omega)}$ is arbitrarily close to the reproducing kernel for the space $\widetilde{X}_m$. Therefore, by re-expressing $\widetilde{X}_m$ with respect to the basis

$$\widetilde{\mathcal{B}}_k(z) \triangleq \frac{\sqrt{1-|\eta_k|^2}}{z-\eta_k}\left(\prod_{\ell=0}^{k-1}\frac{1-\overline{\eta_\ell}z}{z-\eta_\ell}\right) \tag{A.23}$$

and then using an an identical argument as just employed with regard to $\varphi_m^{11}(\lambda,\omega)$

$$\lim_{m\to\infty}\frac{\overline{\varphi_m^{22}(\lambda,\omega)}}{\widetilde{\kappa}_m(\omega)} = \begin{cases} \dfrac{\Phi_u(\omega)}{\Phi_\varepsilon(\omega)\Phi_u(\omega)-|\Phi_{u\varepsilon}(\omega)|^2} & ;\lambda=\omega \\ 0 & ;\lambda\neq\omega \end{cases}. \tag{A.24}$$

where

$$\widetilde{\kappa}_m(\omega) \triangleq \sum_{k=0}^{2m-1}|\widetilde{\mathcal{B}}_k(e^{j\omega})|^2 = \sum_{k=0}^{2m-1}\frac{1-|\eta_k|^2}{|e^{j\omega}-\eta_k|^2}.$$

Substituting (A.20) and (A.24) into (A.13) and (A.21) then also provides

$$\lim_{m\to\infty}\frac{\overline{\varphi_m^{21}(\lambda,\omega)}}{\kappa_m(\omega)} = \begin{cases} -\dfrac{\overline{\Phi_{u\varepsilon}(\omega)}}{\Phi_\varepsilon(\omega)\Phi_u(\omega)-|\Phi_{u\varepsilon}(\omega)|^2} & ;\lambda=\omega \\ 0 & ;\lambda\neq\omega \end{cases} \tag{A.25}$$

$$\lim_{m\to\infty}\frac{\overline{\varphi_m^{12}(\omega,\omega)}}{\widetilde{\kappa}_m(\omega)} = \begin{cases} -\dfrac{\Phi_{u\varepsilon}(\omega)}{\Phi_\varepsilon(\omega)\Phi_u(\omega)-|\Phi_{u\varepsilon}(\omega)|^2} & ;\lambda=\omega \\ 0 & ;\lambda\neq\omega \end{cases}. \tag{A.26}$$

Note then that for the case of $\omega=\lambda$ all of (A.20), (A.24), (A.25) and (A.26) may be jointly expressed as

$$\lim_{m\to\infty}\left[\begin{array}{cc} \overline{\varphi_m^{11}(\omega,\omega)} & \overline{\varphi_m^{12}(\omega,\omega)} \\ \overline{\varphi_m^{21}(\omega,\omega)} & \overline{\varphi_m^{22}(\omega,\omega)} \end{array}\right]\left[\begin{array}{cc} \kappa_m^{-1}(\omega) & 0 \\ 0 & \widetilde{\kappa}_m^{-1}(\omega)\end{array}\right] =$$

$$\frac{1}{\Phi_\varepsilon(\omega)\Phi_u(\omega)-|\Phi_{u\varepsilon}(\omega)|^2}\left[\begin{array}{cc} \Phi_\varepsilon(\omega) & -\Phi_{u\varepsilon}(\omega) \\ -\overline{\Phi_{u\varepsilon}(\omega)} & \Phi_u(\omega)\end{array}\right] = \left[\begin{array}{cc} \Phi_u(\omega) & \Phi_{u\varepsilon}(\omega) \\ \overline{\Phi_{u\varepsilon}(\omega)} & \Phi_\varepsilon(\omega)\end{array}\right]^{-1}. \tag{A.27}$$

Finally, when $\Phi_{ue}(\omega)=0$ and hence $\Phi_{u\varepsilon}(\omega)=0$, then $r_m(\omega)=0$ in (A.15) for any finite m, and also the ensuing equation (A.19) becomes

$$\begin{aligned}\overline{\varphi_m^{11}(\omega,\omega)} &= \frac{1}{|H_\dagger(e^{j\omega})|^2}\Gamma_{2m}^\star(e^{j\lambda})M_{2m}^{-1}\left(\frac{\Phi_u}{|H_\dagger|^2}\right)\Gamma_{2m}(e^{j\omega}) \\ &= \frac{1}{\Phi_u(\omega)}\Gamma_{2m}^\star(e^{j\lambda})\Gamma_{2m}(e^{j\lambda}) \\ &= \frac{\kappa_m(\omega)}{\Phi_u(\omega)}\end{aligned}$$

where the orthogonality of the basis (A.18) and the fact that $\Phi_u(\omega)/|H_\dagger(e^{j\omega})|^2$ is a constant have been used. Via an identical argument, but without the requirement that $\Phi_u(\omega)/|H_\dagger(e^{j\omega})|^2$ be a constant since under the assumptions of the theorem $\Phi_\varepsilon=\sigma^2$ a constant, then

$$\overline{\varphi_m^{22}(\omega,\omega)} = \frac{\widetilde{\kappa}_m(\omega)}{\sigma^2}. \tag{A.28}$$

Also, when $\Phi_{u\varepsilon}(\omega)=0$, then (A.9), (A.10) show that while $\overline{\varphi_m^{12}(\lambda,\omega)}$ and $\overline{\varphi_m^{21}(\lambda,\omega)}$ are respectively elements of X_m and $\widetilde{X}_m$, they are orthogonal to all the basis vectors for that space, and hence they are identically zero. □

B Technical Lemma

Lemma B.1. *Suppose the spectral density matrix $\Phi_\zeta(\omega) > 0$ for all $\omega \in [-\pi, \pi]$. Then the matrix*

$$T_n\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ|^2}\Omega_\circ^\star\right) \tag{B.1}$$

is positive definite, and hence invertible, if and only if there are no pole zero cancellations in either of $G(z, \theta_\circ^n)$ or $H(z, \theta_\circ^n)$.

Proof. By definition

$$T_n\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ|^2}\Omega_\circ^\star\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi(e^{j\omega})\frac{Z_\circ(e^{j\omega})\Phi_\zeta(\omega)Z_\circ^\star(e^{j\omega})}{|H_\circ(e^{j\omega})|^2}\Psi^\star(e^{j\omega})\mathrm{d}\omega$$

where according to (27), (A.1)

$$\Psi(e^{j\omega})Z_\circ(e^{j\omega}) = \left.\frac{\mathrm{d}\Pi(e^{j\omega}, \theta^n)}{\mathrm{d}\theta^n}\right|_{\theta^n = \theta_\circ^n} = \left[\frac{\mathrm{d}G(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n}, \frac{\mathrm{d}H(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n}\right].$$

Now suppose that the matrix (B.1) is rank deficient. Then there exists a non-zero $x \in \mathbf{C}^n$ such that

$$\begin{aligned} 0 &= x^\star T_n\left(\Omega_\circ \frac{Z_\circ \Phi_\zeta Z_\circ^\star}{|H_\circ|^2}\Omega_\circ^\star\right) x \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[x^\star\frac{\mathrm{d}G(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n}, x^\star\frac{\mathrm{d}H(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n}\right]\frac{\Phi_\zeta(\omega)}{|H(e^{j\omega}, \theta_\circ^n)|^2}\left[x^\star\frac{\mathrm{d}G(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n}, x^\star\frac{\mathrm{d}H(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n}\right]^\star \mathrm{d}\omega. \end{aligned}$$

Now, since $\Phi_\zeta(\omega) > 0$ for all $\omega \in [-\pi, \pi]$, then the above integrand is strictly non-negative. Therefore, the integral is equal to zero if, and only if

$$x^\star\frac{\mathrm{d}G(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n} = 0, \qquad \text{and} \qquad x^\star\frac{\mathrm{d}H(e^{j\omega}, \theta_\circ^n)}{\mathrm{d}\theta^n} = 0 \tag{B.2}$$

for all $\omega \in [-\pi, \pi]$. Now, consider the first equality concerning $G(z, \theta_\circ^n)$. By (19)

$$x^\star\frac{\mathrm{d}G(z, \theta_\circ^n)}{\mathrm{d}\theta^n} = \frac{1}{A^2(z, \theta_\circ^n)}\left[p_1(z)A(z, \theta_\circ^n) + p_2(z)B(z, \theta_\circ^n)\right]$$

for some polynomials $p_1(z), p_2(z)$ that are determined by x. With ∂p used to denote the order of a polynomial p then

$$\partial p_1(z) = \partial p_2(z) = m - 1, \qquad \partial A(z, \theta_\circ^n) = m, \qquad \partial B(z, \theta_\circ^n) = m - 1.$$

Therefore,

$$p_1(z)A(z, \theta_\circ^n) + p_2(z)B(z, \theta_\circ^n) = 0 \tag{B.3}$$

is only possible if the z^{m-1} co-efficient in $p_1(z)$ is set to zero. In this case

$$\partial p_1(z)A(z, \theta_\circ^n) = \partial p_2(z)B(z, \theta_\circ^n) = 2m - 2.$$

However, there are then only $(m-2) + (m-1) = 2m - 3$ degrees of freedom remaining in the choices of $p_1(z)$ and $p_2(z)$, so that (B.3) and hence the first equation of (B.2) is impossible. On the other hand, any pole-zero cancellation between $A(z, \theta_\circ^n)$ and $B(z, \theta_\circ^n)$ implies that

$$\partial p_1(z)A(z, \theta_\circ^n) = \partial p_2(z)B(z, \theta_\circ^n) \le 2m - 3$$

so that (B.3) can then be achieved for some $p_1(z)$, $p_2(z)$. Via an identical argument, the latter equality of (B.2) to zero occurs if, and only if there are pole-zero cancellations in $H(z, \theta_\circ^n)$. Therefore, the matrix (B.1) is not positive definite if, and only if, there are pole zero cancellations in both of $G(z, \theta_\circ^n)$ and $H(z, \theta_\circ^n)$. □

C Generalised Fourier Convergence Rate Results

Theorem C.1. *Suppose that $M_{mp}(F)$ and $\Gamma_m(e^{j\omega})$ are given as per definition 2.1 with the associated poles $\{\xi_0, \xi_1, \cdots, \xi_{m-1}\}$ satisfying $|\xi_k| \leq 1 - \delta$ for some $\delta > 0$ and being used to define*

$$\kappa_m(\lambda, \omega) = \sum_{k=0}^{m-1} \mathcal{B}_k(e^{j\omega})\overline{\mathcal{B}_k(e^{j\lambda})}$$

with the orthonormal functions $\{\mathcal{B}_k(z)\}$ being defined in (3). Suppose further that component-wise in its entries $F \in \mathrm{Lip}(\alpha)$, $\alpha > 0$. Then with $[\cdot]_{k,\ell}$ denoting the k, ℓ'th entry of a matrix, the convergence rate bound

$$\max_{\omega \in [-\pi,\pi]} \left| \left[F(\omega) - \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p] M_{mp}(F) [\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} \right]_{k,\ell} \right| \leq \frac{C}{m^{\alpha/(2+\alpha)}}$$

holds for some $C < \infty$ where $\kappa_m(\omega) = \kappa_m(\omega, \omega)$.

Proof. Note that

$$\frac{[\Gamma_m^\star(\omega) \otimes I_p] M_{mp}(F) [\Gamma_m(\omega) \otimes I_p]}{\kappa_m(\omega)} - F(\omega) = \frac{1}{2\pi\kappa_m(\omega)} \int_{-\pi}^{\pi} [F(\sigma) - F(\omega)] \, |\kappa_m(\omega, \sigma)|^2 \, \mathrm{d}\sigma.$$

Therefore, with $f(\cdot)$ denoting an arbitrary scalar k, ℓ'th element of the matrix valued $F(\cdot)$, the definitions

$$h_m(\omega, \sigma) \triangleq \kappa_m(\omega, \sigma), \qquad g_m(\sigma, \mu) \triangleq [f(\mu) - f(\sigma)]\overline{\kappa_m(\mu, \sigma)}$$

imply, via the assumed Lipschitz continuity of f and Lemma C.2, that h_m and g_m satisfy the conditions of Lemma C.3 with $\beta = \gamma = 1$ and $\epsilon = \alpha$. Consequently, applying that Lemma provides

$$\left| \int_{-\pi}^{\pi} [f(\sigma) - f(\omega)] \, |\kappa_m(\omega, \sigma)|^2 \, \mathrm{d}\sigma \right| = \left| \int_{-\pi}^{\pi} f_m(\omega, \sigma) g_m(\sigma, \omega) \, \mathrm{d}\sigma \right| \leq C m^{2/(2+\alpha)}.$$

However, again by Lemma C.2 and the assumption on the poles $\{\xi_k\}$ being strictly within the unit disk, $\kappa_m(\omega) > Cm$ so that

$$\left| \left[F(\omega) - \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p] M_{mp}(F) [\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} \right]_{k,\ell} \right| \leq \frac{C m^{2/(2+\alpha)}}{m} = \frac{C}{m^{\alpha/(2+\alpha)}}.$$

□

Theorem C.2. *Suppose that $M_{mp}(F)$ and $\Gamma_m(e^{j\omega})$ are given as per definition 2.1 with the associated poles $\{\xi_0, \xi_1, \cdots, \xi_{m-1}\}$ satisfying $|\xi_k| \leq 1 - \delta$ for some $\delta > 0$. Then with $[\cdot]_{k,\ell}$ denoting the k, ℓ'th entry of a matrix, and n being a (possibly negative) integer, the convergence rate bound*

$$\max_{\omega \in [-\pi,\pi]} \left| \left[F^n(\omega) - \frac{[\Gamma_n^\star(e^{j\omega}) \otimes I_p] M_{mp}^n(F) [\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)} \right]_{k,\ell} \right| \leq \frac{C}{m^\alpha} \tag{C.1}$$

holds for some $C < \infty$ and $\alpha > 0$ only if component-wise in its elements, $F^n \in \mathrm{Lip}(\alpha)$.

Proof. Define

$$P_m(e^{j\omega}) \triangleq \frac{[\Gamma_m^\star(e^{j\omega}) \otimes I_p] M_{mp}^n(F) [\Gamma_m(e^{j\omega}) \otimes I_p]}{\kappa_m(\omega)}$$

$$V_m(e^{j\omega}) \triangleq P_{2^m}(e^{j\omega}) - P_{2^{m-1}}(e^{j\omega}), \qquad V_0(e^{j\omega}) \triangleq P_0(e^{j\omega}).$$

Therefore, if (C.1) holds then

$$F^n(\omega) = \sum_{m=0}^{\infty} V_m(e^{j\omega})$$

Consequently, for any positive integer d, by use of the Mean Value Theorem and Borwein and Erdélyi's extension of Bernstein's inequality which is re-produced in Lemma C.1, and denoting arbitrary k, ℓ'th scalar valued entries of $F^n(\omega), V_m(e^{j\omega})$ and $P_m(e^{j\omega})$ as $f(\omega), v_m(e^{j\omega})$ and $p_m(e^{j\omega})$ respectively, then

$$\begin{aligned} |f(\lambda) - f(\mu)| &= \left| \sum_{m=0}^{\infty} v_m(e^{j\lambda}) - v_m(e^{j\mu}) \right| \\ &\leq \sum_{m=0}^{d-1} |v_m(e^{j\lambda}) - v_m(e^{j\mu})| + \sum_{m=d}^{\infty} |v_m(e^{j\lambda})| + |v_m(e^{j\mu})| \\ &\leq \sum_{m=0}^{d-1} \left| \frac{\mathrm{d}}{\mathrm{d}\omega} v_m(e^{j\omega}) \Big|_{\omega \in [\lambda,\mu]} \right| |\lambda - \mu| + 2 \sum_{m=d}^{\infty} \|v_m\|_\infty \\ &\leq \frac{1}{\epsilon^2} \sum_{m=0}^{d-1} 2^m \|v_m\|_\infty |\lambda - \mu| + 2 \sum_{m=d}^{\infty} \|v_m\|_\infty . \end{aligned}$$

However, by the assumptions of the theorem (C is a finite constant that may be different values in different parts of expressions)

$$\begin{aligned} \|v_m\|_\infty &\leq \|p_{2^m} - f\|_\infty + \|f - p_{2^{m-1}}\|_\infty \\ &\leq C(2^m)^{-\alpha} + C(2^{m-1})^{-\alpha} \leq C2^{-m\alpha}. \end{aligned}$$

Therefore

$$\begin{aligned} |f(\lambda) - f(\mu)| &\leq C|\lambda - \mu| \sum_{m=0}^{d-1} 2^{m(1-\alpha)} + 2C \sum_{m=d}^{\infty} 2^{-m\alpha} \\ &\leq C|\lambda - \mu| 2^{d(1-\alpha)} + C2^{-d\alpha} \\ &= C|\lambda - \mu|^\alpha \left[(2^d|\lambda - \mu|)^{1-\alpha} + (2^d|\lambda - \mu|)^{-\alpha} \right]. \end{aligned}$$

But d is arbitrary so, for example, for any $|\lambda - \mu|$ it can always be chosen as the least integer such that $1 \leq 2^d|\lambda - \mu| \leq 2$ in which case $(2^d|\lambda - \mu|)^{1-\alpha} + (2^d|\lambda - \mu|)^{-\alpha} \leq 3$. □

Lemma C.1. *The inequality*

$$\left| \frac{\mathrm{d}f(z)}{\mathrm{d}z} \Big|_{z=z_0} \right| \leq \max \left\{ \sum_{\substack{j=1 \\ |a_j|>1}}^{n} \frac{|a_j|^2 - 1}{|a_j - z_0|^2}, \sum_{\substack{j=1 \\ |a_j|<1}}^{n} \frac{1 - |a_j|^2}{|a_j - z_0|^2} \right\} \|f\|_{\partial D}$$

holds for every rational function $f(z) = p_n(z)/q_n(z)$, where $p_n(z)$ is a polynomial of degree at most n with complex coefficients and $q_n(z) = \prod_{k=1}^{n}(z - a_k)$ with $|a_k| \neq 1$ for each k, and for every $z_0 \in \mathbf{T}$, where $\mathbf{T} = \{z \in \mathbf{C}\colon |z| = 1\}$. The above inequality is sharp at every $z_0 \in \mathbf{T}$.

Proof. See [2]. □

Lemma C.2. *Suppose that $|\xi_n| \leq 1 - \delta$ for some $\delta > 0$ and all n. Then for n large enough the following bounds apply.*

$$\frac{1}{2}\sum_{k=0}^{n-1}(1 - |\xi_k|) \leq |K_n(\omega,\sigma)| \leq \begin{cases} \dfrac{2n}{\delta} & ; \forall \sigma, \omega \\[2ex] \dfrac{1}{|\sin(\omega - \sigma)/2|} & ; \omega \neq \sigma \end{cases} \tag{C.2}$$

Proof. See [21]. □

Lemma C.3. *Let $h_n(\omega,\sigma) : [-\pi,\pi] \times [-\pi,\pi] \to \mathbf{C}$ be subject to*

$$|h_n(\omega,\sigma)| \leq \begin{cases} Cn^{\beta} & ; \forall \omega, \sigma \\[2ex] \dfrac{C}{|\sin(\omega - \sigma)/2|} & ; \omega \neq \sigma \end{cases} \tag{C.3}$$

for some $\beta \geq 0$ and let $g_n(\sigma,\mu) : [-\pi,\pi] \times [-\pi,\pi] \to \mathbf{C}$ be subject to

$$|g_n(\sigma,\mu)| \leq \begin{cases} Cn^{\gamma}|\sigma - \mu|^{\varepsilon} & ; \forall \mu, \sigma \\[2ex] \dfrac{C}{|\sin(\mu - \sigma)/2|} & ; \mu \neq \sigma \end{cases} \tag{C.4}$$

for some $\gamma, \varepsilon \geq 0$. Then for n sufficiently large.

$$\left| \int_{-\pi}^{\pi} h_n(\omega,\sigma) g_n(\sigma,\mu)\, \mathrm{d}\sigma \right| \leq \begin{cases} C\min(n^{\lambda}, n^{\delta}\ \log n) & ; \forall \omega, \mu \\[2ex] \dfrac{C}{|\sin(\omega - \mu)/2|}\ \log n & ; \omega \neq \mu \end{cases} \tag{C.5}$$

where

$$\lambda \triangleq \frac{\beta + \gamma}{2 + \varepsilon}, \qquad \delta \triangleq \min(\beta,\gamma).$$

Proof. See[21]. □

References

[1] K. J. ÅSTROM AND B. WITTENMARK, *Computer Controlled Systems: Theory and Design*, Prentice-Hall, Inc., New Jersey, 1984.

[2] P. BORWEIN AND T. ERDÉLYI, *Sharp extensions of Bernstein's inequality to rational spaces*, Mathematika, 43 (1996), pp. 413–423 (1997).

[3] J. W. BREWER, *Kronecker products and matrix calculus in system theory*, IEEE Transactions on Circuits and Systems, 25 (1978), pp. 772–781.

[4] R. EDWARDS, *Fourier Series, a Modern Introduction.*, vol. 2 of Graduate texts in mathematics number 64., Springer Verlag, 1979.

[5] B. EPSTEIN, *Orthogonal Families of Analytic Functions*, Macmillan, 1965.

[6] U. FORSELL AND L. LJUNG, *Closed-loop identification revisited*, Automatica, 53 (1999), pp. 1215–1241.

[7] M. GEVERS, *Essays on control: Perspectives in the theory and its applications*, Birkhäuser, Boston, 1993, ch. Towards a joint design of identification and Control?, pp. 111–151.

[8] M. GEVERS, L. LJUNG, AND P. VAN DEN HOF, *Asymptotic variance expressions for closed loop identification*, Automatica, 37 (2001), pp. 781–786.

[9] G. GOODWIN AND R. PAYNE, *Dynamic System Identification*, Academic Press, 1977.

[10] E. HANNAN AND D. NICHOLLS, *The estimation of the prediction error variance*, Journal of the American Statistical Association, 72 (1977), pp. 834–840.

[11] E. HANNAN AND B. WAHLBERG, *Convergence rates for inverse Toeplitz matrix forms*, Journal of Multivariate Analysis, 31 (1989), pp. 127–135.

[12] H. HJALMARSSON, M. GEVERS, AND F. DE BRUYNE, *For some model based control design criteria, closed loop identification gives better performance*, Automatica, 32 (1996), pp. 1659–1673.

[13] L. LJUNG, *System Identification: Theory for the User, (2nd edition)*, Prentice-Hall, Inc., New Jersey, 1999.

[14] L.LJUNG, *Asymptotic variance expressions for identified black-box transfer function models*, IEEE Transactions on Automatic Control, AC-30 (1985), pp. 834–844.

[15] L.LJUNG AND B. WAHLBERG, *Asymptotic properties of the least squares method for estimating transfer functions and disturbance spectra*, Advances in Applied Probability, 24 (1992), pp. 412–440.

[16] L.LJUNG AND Z.D.YUAN, *Asymptotic properties of black-box identification of transfer functions*, IEEE Transactions on Automatic Control, 30 (1985), pp. 514–530.

[17] B. NINNESS AND F. GUSTAFSSON, *A unifying construction of orthonormal bases for system identification*, IEEE Transactions on Automatic Control, 42 (1997), pp. 515–521.

[18] B. NINNESS AND H. HJALMARSSON, *The analysis of variance error part I: Accurate quantification*, Submitted to IEEE Transactions on Automatic Control, (2001).

[19] ——, *The analysis of variance error part II: Fundamental principles*, Submitted to IEEE Transactions on Automatic Control, (2001).

[20] B. NINNESS, H. HJALMARSSON, AND F. GUSTAFSSON, *The fundamental role of general orthonormal bases in system identification*, IEEE Transactions on Automatic Control, 44 (1999), pp. 1384–1406.

[21] ——, *Generalised Fourier and Toeplitz results for rational orthonormal bases*, SIAM Journal on Control and Optimization, 37 (1999), pp. 429–460.

[22] F. RIESZ AND B. SZ.-NAGY, *Functional Analysis*, Fredrick Ungar, New York, 1955.

[23] W. RUDIN, *Real and Complex Analysis*, McGraw Hill, New York, 1966.