Generalised Fourier and Toeplitz results for Rational Orthonormal Bases

Brett Ninness* Håkan Hjalmarsson† Fredrik Gustafsson‡

Abstract

This paper provides a generalisation of certain classical Fourier convergence and asymptotic Toeplitz matrix properties to the case where the underlying orthonormal basis is not the conventional trigonometric one, but a rational generalisation which encompasses the trigonometric one as a special case. These generalised Fourier and Toeplitz results have particular application in dynamic system estimation theory. Specifically, the results allow a unified treatment of the accuracy of least squares system estimation using a range of model structures, including those that allow the injection of prior knowledge of system dynamics via the specification of fixed pole or zero locations.

Technical Report EE9740, Centre for Integrated Dynamics and Control (CIDAC),
Department of Electrical and Computer Engineering,
University of Newcastle,AUSTRALIA

1 Introduction

In the area of applied mathematics, a fundamental idea is that of approximating or exactly expressing solutions by expanding them in terms of orthogonal basis functions. Well known classical examples are Fourier analysis, solutions of the wave equation and Schrödingers

*This work was supported by the Australian Research Council and the Centre for Integrated Dynamics and Control (CIDAC), University of Newcastle, Callaghan 2308, Australia. This author is with CIDAC and the Department of Electrical & Computer Engineering, University of Newcastle, Australia and can be contacted at email:brett@ee.newcastle.edu.au or FAX: +61 2 49 21 69 93. Part of this work was completed while visiting the Department of Sensors, Signals and Systems-Automatic Control, The Royal Institute of Technology, Stockholm, Sweden.

†Work partly completed while on leave at CIDAC from the Department of Sensors, Signals and Systems-Signal Processing, The Royal Institute of Technology, S-100 44 Stockholm, Sweden. This author can be contacted at email:hakan.hjalmarsson@s3.e.kth.se or FAX: +46 8 790 7329

‡Work partly completed while on leave at CIDAC from the Department of Electrical Engineering, Linköping University, S-581 83, Linköping, Sweden. This author can be contacted at email:fredrik@isy.liu.se or FAX: +46 013 282622

equation in terms of (respectively) Legendre and Laguerre orthogonal polynomials, and solutions of self-adjoint operator equations such as Sturm-Liouville systems in terms of the orthogonal eigenfunctions of the operator. More recently, particularly for the solution of signal processing and other system theoretic problems there has been an explosion of interest in the development and use of a wide class of new orthogonal bases called 'Wavelets' [7, 5].

Indeed, tackling system theoretic problems using orthonormal descriptions has a particularly rich history, going back at least as far as the work of Kolmogorov [23] and Wiener [52] who exploited them in developing their now famous theory on the prediction of random processes. In that work, the orthonormal basis was the trigonometric one, but as was shown by Szegö there is great utility in re-expressing the problem with respect to another orthonormal basis that is adapted to the random process; namely a basis of polynomials orthogonal to a given positive function f which is the spectral density of the process [45, 11]. Such polynomials are called 'Szegö polynomials'.

This latter approach derives its utility from the fact that the n'th order Szegö polynomial is in fact the mean-square best order n one step ahead predictor of the random process [14, 45]. By exploiting the orthonormality of the basis to derive what is called a 'Christoffel–Darboux' formula for the 'Reproducing Kernel' associated with the Szegö polynomial basis, theoretical analysis of this predictor is greatly facilitated. For example, it was by this means that Szegö was able to derive his famous formula

$$\sigma^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\omega)\,\mathrm{d}\omega\right\}$$

for the asymptotic in order n variance σ^2 of the prediction error associated with a spectral density f.

As well, use of the Christoffel–Darboux formula provides a recursive in n formula for the Szegö polynomials [45, 11], and this in turn allows a computationally efficient means for calculating predictors. This recursive formula is of course the famous Levinson recursion, which was developed independently of Szegö's work by exploiting the properties of Toeplitz matrices [26, 38]. In practice, the so–called 'reflection coefficients' required in the Levinson recursions are calculated by the Schur algorithm [41], originally proposed by Schur [43] as a means for testing whether or not a function is bounded positive real (or 'Caratheodory' as it is known in some literature). Here again orthonormal bases and Toeplitz matrices arise since another test for positive realness involves testing for the positive definiteness of the Toeplitz matrix formed from the Fourier co-efficients of the function [44].

These several links between Toeplitz matrices and orthonormal bases arise since (subject to some regularity conditions) the ℓ,m'th element of any $n \times n$ symmetric Toeplitz matrix may be denoted as $T_n(f)$ and expressed using the orthonormal trigonometric basis $\{e^{j\omega n}\}$ as

$$[T_n(f)]_{\ell,m} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega\ell}e^{-j\omega m} f(\omega)\,\mathrm{d}\omega \tag{1}$$

for some positive function f. By recognising this, certain quadratic forms of Toeplitz matrices that arise naturally in the frequency domain analysis of least-squares estimation problems may instead be conveniently rewritten as

$$\frac{1}{n}\Gamma_n^\star(\omega)T_n(f)\Gamma_n(\omega) = \sum_{k=-n}^{n}\left(1-\frac{|k|}{n}\right)c_k\, e^{j\omega k} \tag{2}$$

where $\cdot^\star$ denotes 'conjugate transpose' and,

$$c_k \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)e^{-j\omega k}\,\mathrm{d}\omega$$

is the k'th Fourier co-efficient of f with $\Gamma_n(\omega)$ an $n\times 1$ vector defined as

$$\Gamma_n^\star(\omega) \triangleq [1, e^{-j\omega}, e^{-j2\omega}, \cdots, e^{-j(n-1)\omega}].$$

The right hand side of (2) may be recognised as the Cesàro mean reconstruction of a Fourier series which is known [10], provided f is continuous, to converge uniformly to $f(\omega)$ on $[-\pi,\pi]$.

This latter fact has been exploited by Ljung and co-workers [29, 31, 16, 30, 53, 27] who, reminiscent of Szegö's approach of examining the asymptotic in order n nature of predictors, have provided asymptotic in model order results describing the variability of the frequency response of least-squares system estimates in such a way as to elucidate how they depend on excitation and measurement noise spectral densities, model order, and observed data length; see § 7 for more detail on this point.

Such results have found wide engineering application; see for example [2, 12, 28, 17]. However, to derive them, another key ingredient pertaining to the properties of Toeplitz matrices is required. Namely, that asymptotically in size n, Toeplitz matrices possess the algebraic structure [14, 50]

$$T_n(f)T_n(g) \sim T_n(fg) \tag{3}$$

where f and g are any continuous positive functions, and for $n\times n$ matrices A_n and B_n, the notation $A_n \sim B_n$ means that $\lim_{n\to\infty}|A_n - B_n| = 0$ where $|\cdot|$ is the Hilbert–Schmidt matrix norm defined by

$$|A|^2 \triangleq \frac{1}{n}\mathrm{Trace}\{A^\star A\}. \tag{4}$$

The main results of this paper are to extend the results of the convergence of the Cesàro mean (2) and the algebraic structure of Toeplitz matrices (3) to more general cases wherein the underlying orthonormal basis is not the trigonometric one, but a generalisation of it. More specifically, this paper studies the use of the basis functions $\mathcal{B}_n(z)$ given by

$$\mathcal{B}_n(z) \triangleq \frac{\sqrt{1-|\xi_n|^2}}{1-\xi_n z}\prod_{k=0}^{n-1}\left(\frac{z-\overline{\xi_k}}{1-\xi_k z}\right) \tag{5}$$

where the $\{\xi_k\}$ may be chosen (almost) arbitrarily inside and (in some cases) on the boundary of the open unit disc $\mathbf{D} \triangleq \{z \in \mathbf{C} : |z| < 1\}$ ($\mathbf{C}$ is the field of complex numbers).

These functions $\{\mathcal{B}_n\}$ are orthonormal on the unit circle $\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}$, and the trigonometric basis is a special case of them if all the $\{\xi_k\}$ are chosen as zero. Using them, a generalisation

$$[M_n(f)]_{\ell,m} \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{B}_\ell(e^{j\omega})\overline{\mathcal{B}_m(e^{j\omega})}f(\omega)\,\mathrm{d}\omega \tag{6}$$

of Toeplitz matrices is considered, for which it is shown here that a generalisation of (3) still holds, and with the redefinition

$$\Gamma_n^T(\omega) \triangleq [\mathcal{B}_0(e^{j\omega}),\ \mathcal{B}_1(e^{j\omega}),\ \ldots,\mathcal{B}_{n-1}(e^{j\omega})] \tag{7}$$

it is also shown here that a generalisation of the uniform convergence of the Cesàro mean (2) to $f(\omega)$ also holds.

In both cases, the generalisation involves replacing the $1/n$ normalisation appearing in (2) and in the definition of the matrix norm (4) with a frequency dependent term $K_n(\omega,\omega)$ which is the reproducing kernel associated with the linear space spanned by the basis functions $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{n-1}\}$.

Indeed, this reproducing kernel is the key to the results presented here. Classical derivations of Cesàro summability and Toeplitz matrix results rely heavily on the algebraic structure of the trigonometric basis. Namely, that $e^{j\omega n}e^{j\omega m} = e^{j\omega(n+m)}$. In the cases considered here, since $\mathcal{B}_n\mathcal{B}_m \neq \mathcal{B}_{n+m}$ this algebraic structure is lost and pre-existing analysis techniques are not applicable. Instead, motivated by Szegö's approach to the study of orthogonal polynomials, this paper exploits a closed form expression for the appropriate reproducing kernel.

The utility of the new results presented here is that just as the classical Fourier and Toeplitz results have been used by Ljung and co–workers to analyse estimation using finite impulse response (FIR) and certain other rational model structures, the results of this paper can be used to analyse estimation using generalisations of these model structures. As shown in [37] and as commented on in § 7, these generalised structures are actually quite common since they implicitly arise whenever the common practice of data pre-filtering is performed.

There is much other work related to the results presented here. The study of the basis functions (5) dates back to Malmquist [34] and was taken up by Walsh [49] in the context of complex rational approximation theory, and by other workers [52, 25, 21, 3, 9, 8, 32, 39] for system theoretic applications such as system approximation and network synthesis, including generalisations of Schur and Levinson recursions, Lattice structures, and the concomitant solution of inverse scattering problems.

In the context of system identification, as well as pertaining to the afore-mentioned work [29, 31, 16, 30, 53, 27] the results of this paper also have close connections with much recent literature examining the use of model structures derived from orthonormal bases. In [22, 6, 20, 46, 48] the use of the so-called 'Laguerre' basis is examined. This basis can be encompassed by the basis (5) by fixing all the poles at a common value $\xi_k = \xi \in \mathbf{R}$ ($\mathbf{R}$ denotes the field of real numbers) and with the substitution $z \mapsto 1/z$ so as to accommodate convention in the signal processing and control theory literature.

In this case the name 'Laguerre' derives from the ensuing functions being related to the classical Laguerre orthonormal polynomials via a Fourier and bilinear transform [35]. In [47] a generalisation of this Laguerre case is analysed wherein the common value ξ may be complex valued. In [18, 40], these analyses are again generalised to the case where a fixed set of poles $\{\xi_0, \cdots, \xi_r\}$ are cyclically repeated and orthonormal bases are generated with denominators given as $D_p(z) = \prod_{k=0}^{p-1}(z - \xi_k)$ and numerators the Szegö polynomials associated with the weight function $|D_p(e^{j\omega})|^{-2}$. The cyclic repetition of poles arises due to the latter numerator and denominator pair being multiplied by powers of the all-pass function $z^p D_p(1/z)/D_p(z)$ as the number of required basis functions increases beyond p.

In all these works, any analysis of estimation accuracy proceeds by exploiting the restriction on the choice of ξ_k to establish, via a bilinear transform [46, 47, 48], or a multi-linear transform (dubbed a 'Hambo' transform) [40] an algebra isomorphism to the trigonometric basis $\{e^{j\omega n}\}$. The utility of this is that the original results of Ljung [31] can then be employed, having been mapped through the isomorphism, to provide quantification of estimation accuracy.

In spite of the elegance of this approach, it suffers several drawbacks which are the motivation for the work at hand. Firstly, the results pertain only to a restricted class of models in which either all the poles $\{\xi_k\}$ are chosen the same [46, 48, 47], or are cyclically repeated from a fixed set [40]. Secondly, and with particular reference to [40], the results are asymptotic not as is the case here to the number of poles $\{\xi_k\}$ chosen, but to the number of times the whole set $\{\xi_0, \cdots, \xi_{p-1}\}$ is repeated. The results in this paper allow the avoidance of these limitation by eschewing a strategy of forcing an algebra isomorphism to the trigonometric case.

The presentation of these ideas is organised as follows. In § 2 following, the analysis begins by establishing that the general orthonormal bases (5) fundamental to this paper form a complete set in the Hilbert space $H_2(\mathbf{T})$. In order to study other approximating properties of the basis, a 'Reproducing Kernel' approach is employed, and § 3 is devoted to explaining certain important principles relevant to this framework. Perhaps more importantly, § 3 also contains the derivation of a closed form 'Christoffel–Darboux' type formula for the reproducing kernel. With these results in hand, § 4 then considers generalised Fourier analysis with respect to the basis (5), and using the reproducing kernel ideas establishes uniform convergence for generalised Cesàro mean reconstructions.

In fact, because of application demands, something more is derived in that it is shown that for certain frequencies being different, then uniform convergence to zero also ensues. The generalised Cesàro mean reconstruction is defined with respect to a generalised Toeplitz matrix, and § 5 is devoted to the study of the asymptotic algebraic properties of such matrices since as already explained, these properties are of great utility in certain system theoretic applications.

Pertinent to this, § 5 defines a new notion of asymptotic equivalence between matrices, and then uses this to establish that asymptotically, arbitrary products of generalised Toeplitz matrices and their inverses are equivalent to a single generalised Toeplitz matrix with symbol equal to the product of the corresponding symbols and inverse symbols of the matrices in the product. This study of generalised Toeplitz matrix properties in terms of its symbol is continued in § 6 where the relationship between the spectrum of the matrices and the values of the symbol are explored and found to be intimately connected. With these theoretical developments in hand, § 7 provides a very brief overview of how the results here may be applied in the study of certain system identification problems that were, in fact, the original motivation for this work. More detail on this application theme is given in the separate work [37]. Finally, § 8 provides a summary and concluding perspectives on the work presented here.

2 Completeness Properties

The theme of this paper is to examine certain system theoretic issues pertaining to the use of the basis functions (5) for the purposes of describing discrete time dynamic systems. In the sequel only bounded-input, bounded-output stable and causal systems will be of interest, so that it is natural to embed the analysis in the Hardy space $H_2(\mathbf{T})$ of functions $f(z)$ which are analytic on $\mathbf{D}$, square integrable on $\mathbf{T}$, and possess only a one-sided Fourier expansion. As is well known [19], $H_2(\mathbf{T})$ is a Hilbert space when endowed with the inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{j\omega})\overline{g(e^{j\omega})}\,\mathrm{d}\omega = \frac{1}{2\pi j}\oint_{\mathbf{T}} f(z)\overline{g(z)}\,\frac{\mathrm{d}z}{z}, \quad f,g\in H_2(\mathbf{T}). \tag{8}$$

That the functions (5) form an orthonormal set in that $\langle \mathcal{B}_n, \mathcal{B}_m\rangle = \delta(n-m) =$ Kronecker delta may easily be shown [36] using the contour integral formulation of the inner product in (8) and Cauchy's residue Theorem.

What must be of central interest if the functions (5) are to be useful in such a system theoretic setting is whether or not linear combinations of them can describe an arbitrary system in $H_2(\mathbf{T})$ to any degree of accuracy. This may be answered in the affirmative by the following completeness result which has been developed elsewhere, but is presented here for the sake of a self contained presentation.

Theorem 2.1 (Ninness and Gustafsson [36]).

$$\overline{\mathrm{Span}\,\{\mathcal{B}_k(z)\}_{k\geq 0}} = H_2(\mathbf{T})$$

if and only if

$$\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$$

where here $\overline{X}$ denotes the norm closure of the space X.

3 Reproducing Kernels

Given the completeness result in Theorem 2.1, to further examine the properties of approximants formed as linear combinations of the basis functions (5), this paper takes the approach of utilising the ideas of reproducing kernel spaces [1, 45], of which a brief overview of the key ideas is as follows.

If an approximant $f_n(z)$ of a function $f(z)$ is formed as a certain linear combination of the n basis functions $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{n-1}\}$, then for any $\mu \in \mathbf{D}$ it is also possible to form a linear functional F_μ defined as follows

$$F_\mu : X_n \triangleq \text{Span}\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{n-1}\} \to \mathbf{C}, \quad f_n \mapsto f_n(\mu).$$

Since the setting is a Hilbert space $H_2(\mathbf{T})$, by the Riesz representation Theorem [42] there exists a unique μ dependent element in X_n, call it $K_n(z,\mu)$, such that

$$g(\mu) \triangleq F_\mu(g) = \langle g(z), K_n(z,\mu)\rangle \quad \forall g \in X_n.$$

This element $K_n(z,\mu)$ is called the 'Reproducing Kernel' on account of its property of reproducing values of elements of X_n at the point μ via an inner product.

Certain basic properties of $K_n(z,\mu)$ important for the purposes of this paper are that it is 'Hermitian symmetric' as can be easily seen according to

$$K_n(\sigma,\mu) = \langle K_n(z,\mu), K_n(z,\sigma)\rangle = \overline{\langle K_n(z,\sigma), K_n(z,\mu)\rangle} = \overline{K_n(\mu,\sigma)}$$

so that $K_n(\mu,\mu) \in \mathbf{R}$ and also $K_n(\mu,\mu) > 0$, since if it were not, then since $K_n(\mu,\mu) = \langle K_n(z,\mu), K(z,\mu)\rangle = \|K_n(z,\mu)\|^2$, then $\|K_n(z,\mu)\| = 0$ would be implied, which would further imply that $g(\mu) = 0$ for every $g \in X_n$ which is impossible since, for example, $\mathcal{B}_0(\mu) \neq 0$ for any $\mu \in \mathbf{D}$.

As will be illustrated in the sequel, the reproducing kernel is enormously useful in the study of the approximating properties of the linear span $\{\mathcal{B}_0, \cdots, \mathcal{B}_{n-1}\}$. As a preluding example, in the linear prediction context mentioned in the introduction, consider the problem of finding the n'th order mean square optimal one step ahead predictor $\varphi_n(z) \in X_n$ of a wide-sense stationary process with spectral density $f(\omega)$. Here z is interpreted as the backward shift operator so that if $\{u_k\}$ is a sequence in ℓ_2 then $\{\varphi_n(z)u_k\}$ denotes a filtered version of that sequence. With this notation in hand φ_n is given by

$$\varphi_n = \arg\min_{\varphi \in X_n} \|1-\varphi\| \quad \text{subject to } \varphi(0) = 0.$$

The constraint is added to ensure the one-step ahead nature of the predictor, and the norm is induced by the inner product (8) modified so as to be weighted according to the spectral density $f(\omega)$. This constrained optimisation problem is easily solved using the reproducing kernel $K_n(z,\mu)$ associated with $\text{Span}\{1, \mathcal{B}_0, \cdots, \mathcal{B}_{n-1}\}$ and with respect to the modified inner product by first noting that via the Cauchy–Schwarz inequality

$$\begin{aligned} 1 = |1-\varphi_n(0)|^2 &= |\langle 1-\varphi_n, K_n(z,0)\rangle|^2 \\ &\leq \langle 1-\varphi_n, 1-\varphi_n\rangle \langle K_n(z,0), K_n(z,0)\rangle \\ &= \|1-\varphi_n\|^2 K_n(0,0). \end{aligned}$$

However, equality occurs in the Cauchy–Schwarz inequality if and only if $1-\varphi_n(z) = cK(z,0)$ for some constant c. The constraint $\varphi_n(0)=0$ implies the choice $c = 1/K_n(0,0)$ which leads to the solution

$$\varphi_n(z) = 1 - \frac{K_n(z,0)}{K_n(0,0)}.$$

Given this utility of the reproducing kernel $K_n(z,\mu)$, the natural question of calculating it arises. This may be easily achieved as

$$K_n(z,\mu) = \sum_{k=0}^{n-1} \mathcal{B}_k(z)\overline{\mathcal{B}_k(\mu)}. \tag{9}$$

That this formulation is valid may by quickly checked by noting that for any $0 \leq k < n$

$$\langle \mathcal{B}_k(z), K_n(z,\mu)\rangle = \sum_{n=0}^{n-1} \mathcal{B}_n(\mu)\langle \mathcal{B}_k(z), \mathcal{B}_n(z)\rangle = \mathcal{B}_k(\mu).$$

However, for the purposes of the analysis in this paper this representation is too cumbersome, and a more succinct description is required. This is in common with the study of orthogonal polynomials [45, 11] via the use of reproducing kernels, where simpler closed form formulae for $K_n(z,\mu)$ are called 'Christoffel-Darboux formulae'. Borrowing from this literature, the following theorem presents a Christoffel–Darboux formula for $K_n(z,\mu)$ which in the sequel will be central to the derivation of the generalised Fourier and Toeplitz matrix results of this paper.

Theorem 3.1. *Christoffel–Darboux Formula: Define the modified Blaschke product*

$$\varphi_n(z) \triangleq \prod_{k=0}^{n-1} \frac{z-\overline{\xi_k}}{1-\xi_k z}.$$

Then the Reproducing Kernel of the space spanned by $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{n-1}\}$ *can be expressed as*

$$K_n(z,\mu) = \frac{1-\overline{\varphi_n(\mu)}\varphi_n(z)}{1-z\overline{\mu}}. \tag{10}$$

Proof. Take $z,\mu \in \mathbf{D}$ and consider the function $\Lambda_n(z,\mu) : \mathbf{C}\times\mathbf{C} \to \mathbf{C}$ defined by

$$\Lambda_n(z,\mu) \triangleq \frac{1-\overline{\varphi_n(\mu)}\varphi_n(z)}{1-z\overline{\mu}}.$$

and define the space $X_n \triangleq \text{Span}\{\mathcal{B}_0(z),\ldots,\mathcal{B}_{n-1}(z)\} \subset H_2(\mathbf{T})$. Clearly, since the product $\overline{\varphi_n(\mu)}\varphi_n(1/\overline{\mu}) = 1$ then $1-\overline{\varphi_n(\mu)}\varphi_n(z)$ possesses a zero at $z = 1/\overline{\mu}$ so that $\Lambda_n(z,\mu) \in X_n$. Furthermore, by Cauchy's integral Theorem

$$\langle \mathcal{B}_k(z), \frac{1}{1-\overline{\mu}z}\rangle = \frac{1}{2\pi j}\oint_{\mathbf{T}} \frac{\mathcal{B}_k(z)}{z-\mu}\,\mathrm{d}z = \mathcal{B}_k(\mu)$$

and also, for any $k = 0, 1, \cdots, n-1$ by the change of integration variable $z \mapsto 1/z$ and Cauchy's integral Theorem

$$\langle \mathcal{B}_k(z), \frac{\overline{\varphi_n(\mu)}\varphi_n(z)}{1-\overline{\mu}z}\rangle = \frac{\varphi_n(\mu)}{2\pi j}\oint_{\mathbf{T}} \frac{\mathcal{B}_k(z)\overline{\varphi_n(z)}}{z-\mu}\,\mathrm{d}z = \frac{\varphi_n(\mu)}{2\pi j}\oint_{\mathbf{T}} \frac{\mathcal{B}_k(1/z)\overline{\varphi_n(1/z)}}{1-\mu z}\,\mathrm{d}z = 0.$$

Therefore, $\Lambda_n(z,\mu)$ given by (10) has the property that $f(\mu) = \langle f(z), \Lambda_n(z,\mu)\rangle$ for any $f \in X_n$. However, the reproducing kernel $K_n(z,\mu)$ is the unique function in X_n with this property, so it must be that $K_n(z,\mu) = \Lambda_n(z,\mu)$. □

Often the expression (10) will be used by setting $\mu = re^{j\sigma}, z = re^{j\omega}$ and letting $r \to 1$ from below. In this case, with some abuse of notation in the interests of cleanliness of exposition, the theorem will be used in the form

$$K_n(\omega,\sigma) = \frac{1-\overline{\varphi_n(e^{j\sigma})}\varphi_n(e^{j\omega})}{1-e^{j(\sigma-\omega)}}. \tag{11}$$

4 Generalised Fourier Series Convergence

Given a function $f \in H_2(\mathbf{T})$, an obvious way of approximating it in terms of the basis functions $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{n-1}\}$ is as f_n given by

$$f_n(z) = \arg\min_{g\in X_n} \|f-g\| = \sum_{k=0}^{n-1} \langle f, \mathcal{B}_k\rangle \mathcal{B}_k(z). \tag{12}$$

Provided $\sum(1-|\xi_k|) = \infty$ holds, then by the completeness theorem 2.1, the approximation error $\|f_n - f\|$ can be made arbitrarily small for arbitrarily large approximation order n.

A natural question to then ask is how the approximant f_n behaves with respect to other norms, for example the supremum norm on $[-\pi,\pi]$. The purpose of this section is to show that a modified approximant, closely related to the above one and deriving from the Cesàro (or Fejér) mean of classical Fourier analysis, is also supremum norm convergent to f under the same condition of $\sum(1-|\xi_k|) = \infty$. This result will encompass the classical result for the trigonometric basis by simply setting all the poles $\{\xi_k\}$ to zero.

To proceed, it is expedient to revisit the classical case by indeed setting $\xi_k = 0$ in (12) and also temporarily shifting to the $L_2(\mathbf{T})$ setting so that the sum in (12) becomes two-sided to obtain

$$f_n(\omega) = \frac{1}{2\pi}\sum_{k=-n}^{n} e^{jk\omega}\int_{-\pi}^{\pi} f(\sigma)e^{-j\sigma k}\,\mathrm{d}\sigma = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\sigma)D_n(\omega-\sigma)\,\mathrm{d}\sigma \tag{13}$$

where

$$D_n(\theta) \triangleq \frac{\sin(2n+1)\theta/2}{\sin\theta/2}$$

is known [10] as the 'Dirichlet kernel'. Perhaps one of the more surprising facets of applied mathematics is that even if f is continuous on $[-\pi,\pi]$, then $\lim_{n\to\infty}|f_n - f| = 0$ uniformly on $[-\pi, \pi]$ is not guaranteed; a century old observation due to Du Bois-Reymond [24]. This undesirable behaviour stems from the fact that although from (13) one would wish $D_n(\theta)$ to behave more and more like a Dirac delta function as n increases, it does not in the sense that

$$\liminf_{n\to\infty} \int_{|\theta|>\rho} |D_n(\theta)| \,\mathrm{d}\theta \neq 0$$

for arbitrarily small ρ. In fact, the quantity $\|D_n\|_1$ (called the n'th Lebesgue constant) is known to be bounded below by $(4/\pi^2)\log n$, so that since the norm of any linear projection $L_n : C[-\pi, \pi] \to \mathrm{Span}\{e^{-j\omega n}, \cdots, e^{-j\omega}, 1, e^{j\omega}, \cdots, e^{j\omega n}\}$ is known [4] to be under-bounded by $\|D_n\|_1$, then by Du Bois-Reymonds result there always exists an $f \in C[-\pi, \pi]$ such that $\|L_n f\|$ becomes unbounded as $n \to \infty$. In fact, this difficulty has been the genesis of much work in the system identification literature, of which [33] offers a survey.

A remedy for this problem of non-convergence is to replace the approximation (13) with the so-called Cesàro mean defined on the unit circle by

$$f_n(\omega) = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right) \langle f, e^{jk\omega}\rangle e^{jk\omega} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\sigma) K_n(\omega - \sigma)\,\mathrm{d}\sigma \tag{14}$$

where now

$$K_n(\theta) \triangleq \frac{\sin^2 n\theta/2}{n \sin^2 \theta/2}$$

is known [10] as the 'Fejér kernel' and does possess the 'delta-like' property

$$\lim_{n\to\infty} \int_{|\theta|>\rho} K_n(\theta)\,\mathrm{d}\theta = 0 \quad \forall \rho > 0 \tag{15}$$

so that since it is also true that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(\theta)\,\mathrm{d}\theta = 1$$

then

$$2\pi|f_n(\omega) - f(\omega)| \leq \int_{|\omega-\sigma|\leq\theta} |f_n(\sigma) - f(\omega)| K_n(\sigma)\,\mathrm{d}\sigma + \int_{|\omega-\sigma|>\rho} |f_n(\sigma) - f(\omega)| K_n(\sigma)\,\mathrm{d}\sigma$$

in which case if f is continuous, then use of (15) allows the conclusion that since ρ may be made arbitrarily small, then $\lim_{n\to\infty} |f_n(\omega) - f(\omega)| = 0$ uniformly on $[-\pi, \pi]$.

To develop the analog of this result for the general case of approximants formed using the general orthonormal basis (5), it is necessary to first develop a generalisation of the Cesàro mean (14). This may be accomplished by the definition

$$f_n(\omega) \triangleq \frac{\Gamma_n^\star(\omega) M_n(f) \Gamma_n(\omega)}{K_n(\omega, \omega)} \tag{16}$$

where Γ_n defined in (7) is an $n \times 1$ vector of general rational orthonormal basis functions (5) and $M_n(f)$ is a generalised Toeplitz matrix as defined in (6). If all the poles $\{\xi_k\}$ are set to zero in (9), then it is straightforward to verify that the formulation (16) reduces (since in this case $M_n(f) = T_n(f)$) to the usual Cesàro mean (14). To analyse the convergence properties of (16), note that by the formulation (9)

$$\begin{aligned}\Gamma_n^\star(\omega)M_n(f)\Gamma_n(\omega) &= \sum_{m=0}^{n-1}\sum_{n=0}^{n-1}\overline{\mathcal{B}_m(e^{j\omega})}\mathcal{B}_n(e^{j\omega})[M_n(f)]_{m,n}\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\sigma)\sum_{m=0}^{n-1}\sum_{n=0}^{n-1}\overline{\mathcal{B}_m(e^{j\omega})}\mathcal{B}_n(e^{j\omega})\mathcal{B}_m(e^{j\sigma})\overline{\mathcal{B}_n(e^{j\sigma})}\,\mathrm{d}\sigma\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\sigma)\,|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma.\end{aligned}$$

Therefore, since by the defining property of the reproducing kernel

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma = \langle K_n(\omega,\sigma),K_n(\omega,\sigma)\rangle = K_n(\omega,\omega) = \sum_{m=0}^{n-1}|\mathcal{B}_m(e^{j\omega})|^2 \qquad (17)$$

then

$$\begin{aligned}\frac{1}{2\pi}\left|\frac{\Gamma_n^\star(\omega)M_n(f)\Gamma_n(\omega)}{K_n(\omega,\omega)} - f(\omega)\right| &= \frac{1}{2\pi K_n(\omega,\omega)}\left|\Gamma_n^\star(\omega)M_n(f)\Gamma_n(\omega) - K_n(\omega,\omega)f(\omega)\right|\\ &= \frac{1}{2\pi K_n(\omega,\omega)}\left|\int_{-\pi}^{\pi}[f(\sigma)-f(\omega)]\,|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| \quad (18)\end{aligned}$$

and so, in analogy with classical Fourier analysis, convergence of the generalised Cesàro mean approximant (16) hinges on a kernel function (in this case depending on the reproducing kernel and being given by $|K_n(\omega,\sigma)|^2/K_n(\omega,\omega)$) behaving in some sense like the Dirac delta function $\delta(\omega-\sigma)$. Via use of the Christoffel–Darboux formula (11) for the reproducing kernel, it is possible to establish that this 'delta-like' behaviour does in fact occur in the following sense.

Lemma 4.1. *For any $\rho > 0$ and provided*

$$\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty,$$

then

$$\lim_{n\to\infty}\frac{1}{K_n(\omega,\omega)}\int_{\sigma\notin[\omega-\rho,\omega+\rho]}|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma = 0.$$

Proof. By Theorem 3.1 formulated as (11)

$$|K_n(\omega,\sigma)| \le \frac{2}{|e^{j(\sigma-\omega)}-1|} = \frac{1}{|\sin(\sigma-\omega)/2|}$$

so that

$$\int_{\sigma\notin[\omega-\rho,\omega+\rho]} |K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma \le \frac{2\pi}{\sin^2\rho/2}.$$

Also, since $|1-\xi_k e^{j\omega}| \le 1+|\xi_k|$ then

$$K_n(\omega,\omega) = \sum_{k=0}^{n-1}\frac{1-|\xi_k|^2}{|1-\xi_k e^{j\omega}|^2} \ge \frac{1}{2}\sum_{k=0}^{n-1}(1-|\xi_k|) \tag{19}$$

so that

$$\frac{1}{K_n(\omega,\omega)}\int_{\sigma\notin[\omega-\rho,\omega+\rho]} |K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma \le \frac{\pi}{\sin^2\rho/2}\left(\sum_{k=0}^{n-1}(1-|\xi_k|)\right)^{-1}$$

which tends to zero under the conditions of the Lemma. □

Before using this result, some further notation is required since motivated by the desire to provide results applicable to certain system theoretic problems, it is necessary to be more ambitious than to just prove convergence of f_n to f. Instead, it is also necessary to prove convergence to zero of the quadratic form

$$\frac{\Gamma_n^\star(\mu)M_n(f)\Gamma_n(\omega)}{K_n(\omega,\omega)}$$

when $\mu\neq\omega$ and this entails some delicacies in dealing with the distance between $e^{j\mu}$ and $e^{j\omega}$ on the unit circle, so that the distance between μ and ω must be considered modulo 2π. This is achieved by defining, for $-\pi\le x,y\le\pi$, the function $d(x,y):[-\pi,\pi]\times[-\pi,\pi]\to[0,\pi]$ as

$$d(x,y) \triangleq \min(|x-y|, 2\pi-|x-y|). \tag{20}$$

Furthermore, letting Ω_1 and Ω_2 be subsets of $[-\pi,\pi]$, then

$$d(\Omega_1,\Omega_2) \triangleq \min_{x\in\Omega_1,\, y\in\Omega_2} d(x,y).$$

In the sequel, it will also be useful to note that because

$$|\sin(x)| = |\sin(\pi-x)| = |\sin(-x)| = |\sin(-(\pi-x))|$$

then

$$|\sin(x)| = \sin\frac{d(2x,0)}{2}, \quad -\pi\le x\le\pi$$

so that since $2x/\pi \le \sin x \le x$ for $0\le x\le\pi/2$ then

$$\frac{2}{d(2x,0)} \le \frac{1}{|\sin(x)|} \le \frac{\pi}{d(2x,0)} \tag{21}$$

for any $x : |x|\le\pi,\ x\neq 0$. The main use for these latter ideas is to develop further bounds such as the ones contained in the following lemma, which sharpen the interpretation (already begun in Lemma 4.1) of $K_n(\omega,\sigma)$ as behaving approximately like the Dirac delta $\delta(\omega,\sigma)$.

Lemma 4.2. *Suppose that $|\xi_n| \le 1-\delta$ for some $\delta > 0$ and all n. Then for n large enough the following bounds apply.*

$$\frac{1}{2}\sum_{k=0}^{n-1}(1-|\xi_k|) \le |K_n(\omega,\sigma)| \le \begin{cases} \dfrac{2n}{\delta} & ; \forall \sigma,\omega \\[2ex] \dfrac{1}{|\sin(\omega-\sigma)/2|} & ; \omega \neq \sigma \end{cases} \tag{22}$$

Proof. Consider first the case of $\omega = \sigma$. Then by the expression (17) and the formulation (5).

$$K_n(\omega,\omega) = \sum_{k=0}^{n-1}\frac{1-|\xi_k|^2}{|1-\xi_k e^{j\omega}|^2} \le \sum_{k=0}^{n-1}\frac{1+|\xi_k|}{1-|\xi_k|} \le \frac{2}{\delta}n \tag{23}$$

so that using the Cauchy–Schwarz inequality the bound

$$|K_n(\omega,\sigma)| \le \sqrt{|K_n(\omega,\omega)|}\ \sqrt{|K_n(\sigma,\sigma)|} \le \frac{2}{\delta}n. \tag{24}$$

applies for all ω, σ. For the case of $\omega \neq \sigma$, then using (11) derived from Theorem 3.1 leads to the bound

$$|K_n(\omega,\sigma)| = \frac{1}{|e^{j(\omega-\sigma)}-1|} \le \frac{1}{|\sin(\omega-\sigma)/2|}.$$

Finally, the under-bound has already been established in (19). ☐

With these ideas in hand, the following result is available.

Theorem 4.1. *Suppose $f(\omega)$ is a continuous not necessarily real-valued function on $[-\pi,\pi]$. Then provided*

$$\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$$

the following limit result holds

$$\lim_{n\to\infty}\frac{\Gamma_n^\star(\omega)M_n(f)\Gamma_n(\omega)}{K_n(\omega,\omega)} = f(\omega)$$

uniformly in ω on $[-\pi,\pi]$. Under the strengthened condition that $|\xi_n| \le 1-\delta$ for some $\delta > 0$ and all n, then for $\mu \neq \omega$

$$\lim_{n\to\infty}\frac{\Gamma_n^\star(\mu)M_n(f)\Gamma_n(\omega)}{K_n(\omega,\omega)} = 0.$$

Proof. Consider first the case of $\mu = \omega$. Then from (18) and for arbitrary $\rho > 0$

$$\begin{aligned}\frac{1}{2\pi}\left|\frac{\Gamma_n^\star(\omega)M_n(f)\Gamma_n(\omega)}{K_n(\omega,\omega)} - f(\omega)\right| &= \frac{1}{2\pi K_n(\omega,\omega)}\left|\int_{-\pi}^{\pi}[f(\sigma)-f(\omega)]\,|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| \\ &\le \frac{1}{2\pi K_n(\omega,\omega)}\left|\int_{\sigma\in[\omega-\rho,\omega+\rho]}[f(\sigma)-f(\omega)]\,|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| + \\ &\quad \frac{1}{2\pi K_n(\omega,\omega)}\left|\int_{\sigma\notin[\omega-\rho,\omega+\rho]}[f(\sigma)-f(\omega)]\,|K_n(\omega,\sigma)|^2\,\mathrm{d}\sigma\right|.\end{aligned}$$

Now, $f(\omega)$ is continuous, so for ρ sufficiently small

$$|f(\sigma) - f(\omega)| \le \epsilon \ \text{ on } [\omega - \rho, \omega + \rho].$$

Using this and (17) gives that for sufficiently small ρ

$$\frac{1}{2\pi K_n(\omega,\omega)} \left| \int_{\sigma \in [\omega-\rho,\omega+\rho]} [f(\sigma) - f(\omega)] \, |K_n(\omega,\sigma)|^2 \, \mathrm{d}\sigma \right| \le \frac{\epsilon}{2\pi K_n(\omega,\omega)} \int_{-\pi}^{\pi} |K_n(\omega,\sigma)|^2 \, \mathrm{d}\sigma = \epsilon.$$

Also, since f is continuous on compact $[-\pi,\pi]$ then $|f|$ is bounded by some $M/2 < \infty$. Therefore

$$\frac{1}{2\pi K_n(\omega,\omega)} \left| \int_{\sigma \notin [\omega-\rho,\omega+\rho]} [f(\sigma) - f(\omega)] \, |K_n(\omega,\sigma)|^2 \, \mathrm{d}\sigma \right| \le \frac{M}{2\pi K_n(\omega,\omega)} \int_{\sigma \notin [\omega-\rho,\omega+\rho]} |K_n(\omega,\sigma)|^2 \, \mathrm{d}\sigma$$

which provides

$$\frac{1}{2\pi} \left| \frac{\Gamma_n^\star(\omega) M_n(f) \Gamma_n(\omega)}{K_n(\omega,\omega)} - f(\omega) \right| \le \epsilon + \frac{M}{2\pi K_n(\omega,\omega)} \int_{\sigma \notin [\omega-\rho,\omega+\rho]} |K_n(\omega,\sigma)|^2 \, \mathrm{d}\sigma.$$

Using Lemma 4.1 and the fact that ϵ is arbitrary then gives the result for $\mu = \omega$. Now consider the case $\mu \ne \omega$. Define the regions

$$\begin{aligned} \Omega_1 &\triangleq \left\{\sigma \in [-\pi,\pi] : |\sigma - \mu| < K_n^{-\alpha}(\omega,\omega)\right\}, \\ \Omega_2 &\triangleq \left\{\sigma \in [-\pi,\pi] : |\sigma - \omega| < K_n^{-\alpha}(\omega,\omega)\right\}, \\ \Omega_3 &\triangleq \left\{\sigma \in [-\pi,\pi] : \sigma \notin \{\Omega_1 \cup \Omega_2\}\right\}, \end{aligned}$$

where $\alpha \in (0, 1/2)$ is arbitrary. In this case

$$\begin{aligned} \left| \frac{\Gamma_n^\star(\mu) M_n(f) \Gamma_n(\omega)}{K_n(\omega,\omega)} \right| &= \left| \frac{1}{2\pi K_n(\omega,\omega)} \int_{-\pi}^{\pi} f(\sigma) K_n(\mu,\sigma) \overline{K_n(\omega,\sigma)} \, \mathrm{d}\sigma \right| \\ &\le \frac{\|f\|_\infty}{2\pi K_n(\omega,\omega)} \int_{\Omega_1 \cup \Omega_2 \cup \Omega_3} |K_n(\mu,\sigma) \overline{K_n(\omega,\sigma)}| \, \mathrm{d}\sigma. \end{aligned}$$

Consider the integrals over the various regions in turn. By Lemma 4.2 the bound $K_n(\omega,\omega) \ge 1/2 \sum_{k=0}^{n-1} (1 - |\xi_k|)$ holds, so that by the assumptions of the theorem, n can be taken large enough that Ω_1 and Ω_2 do not overlap. Assuming this to be the case, then $|\omega - \sigma| > K_n^{-\alpha}(\omega,\omega)$ on Ω_1 and hence using the Lemma 4.2

$$|K_n(\omega,\sigma)| \le \frac{1}{|\sin(\omega - \sigma)/2|} \le \frac{\pi}{d^\alpha(\omega,\sigma)}, \quad \sigma \in \Omega_1.$$

Therefore, under the assumption that $|\xi_k| \le 1 - \delta, \delta > 0$ then by Lemma 4.2 $|K_n(\mu,\sigma)| \le 2n/\delta$, so that assuming n is so large that $K_n^{-\alpha}(\omega,\omega) \le d(\mu,\omega)/4$ gives on use of Lemma B.1

$$\int_{\Omega_1} |K_n(\mu,\sigma) \overline{K_n(\omega,\sigma)}| \, \mathrm{d}\sigma \le \frac{2n}{\delta} \int_{\Omega_1} \frac{1}{|\sin(\omega - \sigma)/2|} \, \mathrm{d}\sigma \le \frac{32\, n}{\delta K_n^\alpha(\omega,\omega) |\sin(\omega - \mu)/2|}.$$

Using an identical argument

$$\int_{\Omega_2} |K_n(\mu,\sigma)\overline{K_n(\omega,\sigma)}|\,\mathrm{d}\sigma \leq \frac{32\,n}{\delta K_n^{\alpha}(\omega,\omega)|\sin(\omega-\mu)/2|}.$$

Finally, by the definition of Ω_3 and Lemma 4.2

$$\int_{\Omega_3} |K_n(\mu,\sigma)\overline{K_n(\omega,\sigma)}|\,\mathrm{d}\sigma \leq \frac{2\pi}{\sin^2 K_n^{-\alpha}(\omega,\omega)/2} \leq 8\pi K_n^{2\alpha}(\omega,\omega).$$

Combining the bounds on the integrals over the various regions gives

$$\left|\frac{\Gamma_n^\star(\mu)M_n(f)\Gamma_n(\omega)}{K_n(\omega,\omega)}\right| \leq \frac{32n\|f\|_\infty}{\pi K_n^{1+\alpha}(\omega,\omega)|\sin(\omega-\mu)/2|} + \frac{4\|f\|_\infty}{K_n^{1-2\alpha}(\omega,\omega)}$$

which, according to the lower bound in Lemma 4.2 and since $\alpha \in (0, 1/2)$, tends to zero as $n \to \infty$. □

5 Algebraic Structure of Generalised Toeplitz Matrices

In applications [29, 31, 16, 30, 53, 27], the consideration of quadratic forms more complicated than (16) occur. In fact, what is of more interest are forms such as

$$\frac{\Gamma_n^\star(\omega)M_n(f)M_n(g)\Gamma_n(\omega)}{K_n(\omega,\omega)}.$$

In these aforementioned applications [29, 31, 16, 30, 53, 27], the underlying orthonormal basis is the trigonometric one $\{e^{j\omega n}\}$ in which case $M_n(f) = T_n(f)$ is a bona-fide Toeplitz matrix for which classical results are at hand concerning their algebraic structure. Namely, following the notation defined in (3), the convenient property that $T_n(f)T_n(g) \sim T_n(fg)$ is assured [14, 50] (the meaning of the $\sim$ notation here is as described in conjunction with equation (3)).

The purpose of this section is to establish this same algebraic structure for the generalised Toeplitz matrices defined by (6), the classical results once again arising as the special case of $\xi_k = 0$ in (5). To begin, note that by the formulation (6)

$$\begin{aligned}[M_n(f)M_n(g)]_{m,\ell} &= \frac{1}{4\pi^2}\sum_{k=0}^{n-1}\int_{-\pi}^{\pi} \mathcal{B}_m(\omega)\overline{\mathcal{B}_k(\omega)}f(\omega)\,\mathrm{d}\omega \int_{-\pi}^{\pi}\overline{\mathcal{B}_\ell(\sigma)}\mathcal{B}_k(\sigma)g(\sigma)\,\mathrm{d}\sigma \\ &= \sum_{k=0}^{n-1}\langle \mathcal{B}_m f, \mathcal{B}_k\rangle\overline{\langle \mathcal{B}_\ell g, \mathcal{B}_k\rangle}\end{aligned}$$

and

$$[M_n(fg)]_{m,\ell} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{B}_m(\omega)\overline{\mathcal{B}_\ell(\omega)}f(\omega)\overline{g(\omega)}\,\mathrm{d}\omega = \langle \mathcal{B}_m f, \mathcal{B}_\ell g\rangle.$$

Therefore, by Parseval's Theorem

$$\begin{aligned}\left|[M_n(f)M_n(g)]_{m,\ell} - [M_n(fg)]_{m,\ell}\right| &= \left|\sum_{k=0}^{n-1}\langle \mathcal{B}_m f, \mathcal{B}_k\rangle\overline{\langle \mathcal{B}_\ell g, \mathcal{B}_k\rangle} - \langle \mathcal{B}_m f, \mathcal{B}_\ell g\rangle\right| \\ &\le \|\mathcal{B}_m f - \widehat{f}_n\|\|\mathcal{B}_\ell g - \widehat{g}_n\|,\end{aligned}$$

where

$$\widehat{f}_n(z) = \sum_{k=0}^{n-1}\langle \mathcal{B}_m f, \mathcal{B}_k\rangle \mathcal{B}_k(z), \quad \widehat{g}_n(z) = \sum_{k=0}^{n-1}\langle \mathcal{B}_\ell g, \mathcal{B}_k\rangle \mathcal{B}_k(z)$$

so that if the Hilbert-Schmidt norm dependence of matrix equivalence (3),(4) is to be employed, then interest is centred on the behaviour of the overbound

$$|M_n(f)M_n(g) - M_n(fg)| \le \frac{1}{n}\sum_{m=0}^{n-1}\sum_{\ell=0}^{n-1}\|\mathcal{B}_m f - \widehat{f}_n\|\|\mathcal{B}_\ell g - \widehat{g}_n\|. \tag{25}$$

Now, since $|\mathcal{B}_m|$ is bounded then $\mathcal{B}_m f \in L_2(\mathbf{T})$ so since $\{\mathcal{B}_k\}$ is a basis for $H_2(\mathbf{T})$, since $\{z^{-k}\}, k > 0$ is a basis for the orthogonal complement $H_2(\mathbf{T})^\perp$ [19], and since $L_2 = H_2 \oplus H_2^\perp$ then $\mathcal{B}_m f$ can be expanded as

$$\mathcal{B}_m f = \sum_{k=0}^{\infty}\langle \mathcal{B}_m f, \mathcal{B}_k\rangle \mathcal{B}_k + \sum_{k=1}^{\infty}\langle \mathcal{B}_m f, z^{-k}\rangle z^{-k}$$

so that

$$\|\mathcal{B}_m f - \widehat{f}_n\|^2 = \sum_{k=n}^{\infty}|\langle \mathcal{B}_m f, \mathcal{B}_k\rangle|^2 + \sum_{k=1}^{\infty}\left|\langle \mathcal{B}_m f, z^{-k}\rangle\right|^2 \tag{26}$$

and the task then becomes to try to show that as n and m increase the terms in these sums tend to zero sufficiently quickly. In the trigonometric case this is straightforward since, for example, $\langle \mathcal{B}_m f, \mathcal{B}_k\rangle$ becomes $\langle z^m f, z^k\rangle = \langle f, z^{k-m}\rangle$ which is the $k-m$'th term in the Fourier expansion of f. Assuming f is sufficiently smooth that these Fourier components die at some exponential rate, say $\eta^{|k-m|}$ with $|\eta| < 1$ then provides (with the same reasoning giving $|\langle z^m f, z^{-k}\rangle| = |\langle f, z^{-(m+k)}\rangle| \le |\eta|^{m+k}$) $\|\mathcal{B}_m f - \widehat{f}_n\| \le K(\eta^{n-m} + \eta^m)$ for some $K < \infty$ so that the sums in the overbound in (25) are convergent and hence (25) tends to zero with increasing n thereby establishing $T_n(f)T_n(g) \sim T_n(fg)$ as $n \to \infty$.

Generalising this to the basis (5) is surprisingly more difficult. However, consider the simplifying assumption that both f and g have finite dimensional, say n'th order spectral factors of the form $f(z) = H(z)H(1/z)$ where

$$H(z) = \sum_{r=0}^{\infty} h_r z^r, \quad h_r = \sum_{i=0}^{n-1}\gamma_i^r$$

with $|\gamma_i| < 1$ and where for expediency (but without loss of generality) it is assumed that the $\{\gamma_i\}$ are isolated. Then the $|\langle \mathcal{B}_m f, \mathcal{B}_k\rangle|$ term can be simply bounded by the calculation

$$\begin{aligned}
\langle \mathcal{B}_m f, \mathcal{B}_k\rangle &= \sum_{r=0}^{\infty}\sum_{\ell=0}^{\infty} h_r\overline{h_\ell}\langle \mathcal{B}_m\overline{\mathcal{B}_k}z^r, z^\ell\rangle \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{r=0}^{\infty}\sum_{\ell=r+1}^{\infty} \gamma_i^r\overline{\gamma_j^\ell}\langle \mathcal{B}_m\overline{\mathcal{B}_k}z^r, z^\ell\rangle \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{r=0}^{\infty} \gamma_i^r\overline{\gamma_j^r}\sum_{\ell=1}^{\infty}\overline{\gamma_j^\ell}\langle \mathcal{B}_m\overline{\mathcal{B}_k}, z^\ell\rangle \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{r=0}^{\infty} \gamma_i^r\overline{\gamma_j^r}\frac{\overline{\gamma_j}\sqrt{(1-|\xi_m|^2)(1-|\xi_k|^2)}}{(\overline{\gamma_j}-\overline{\xi_k})(1-\overline{\gamma_j}\xi_m)}\prod_{t=m+1}^{k}\left(\frac{\overline{\gamma_j}-\overline{\xi_t}}{1-\xi_t\overline{\gamma_j}}\right)
\end{aligned}$$

where in progressing to the last line it has been recognised that the inner sum in the second last line is the evaluation at $z = \overline{\gamma_j}$ of a function $\mathcal{B}_m(z)\overline{\mathcal{B}_k(z)}$ with impulse response terms $\langle \mathcal{B}_m\overline{\mathcal{B}}_k, z^\ell\rangle$ and without loss of generality it has also been assumed that $k > m$. Therefore, since $|\gamma_j| < 1$ there exists $|\eta| < 1$, $K < \infty$ both independent of n such that (K is different in different parts of the following expressions)

$$\begin{aligned}
|\langle \mathcal{B}_m f, \mathcal{B}_k\rangle| &\le K\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{r=0}^{\infty}|\gamma_i|^r|\gamma_j|^r\eta^{k-m} \\
&= K\eta^{k-m}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\frac{1}{1-|\gamma_i\gamma_j|} \\
&\le K\eta^{|k-m|}
\end{aligned}$$

where in the last line the fact that the same argument works for $k < m$ has been taken into account. Using the same method, one can also bound $|\langle \mathcal{B}_m f, z^{-k}\rangle|$ to arrive at $\|\mathcal{B}_m f - \widehat{f}\| \le K(\eta^{p-m} + \eta^m)$, and arguing as before for the trigonometric case, the desired algebraic result $M_n(f)M_n(g) \sim M_n(fg)$ can be extended to the case of generalised Toeplitz matrix (6).

In passing, on an intuitive level this extended result may be understood by noting that since the basis functions $\{\mathcal{B}_m\}$ contain order m all-pass factors φ_m as defined in Theorem 3.1, then they approximate the order m shift z^m in that they may be written as $\varphi_m(e^{j\omega}) = e^{j\psi_m(\omega)}$ where $\psi_m(\omega)$ is a monotonically non-decreasing function taking values between 0 and $2m\pi$.

Note also in passing that this simplified development illustrates why convergence in the strong matrix norm is not possible. Specifically, the bounds just developed are of the form such that the difference $|[M_n(f)M_n(g) - M_n(fg)]_{m,\ell}| \le K(\eta^{p-m} + \eta^m)(\eta^{p-\ell} + \eta^\ell)$ is a matrix with 'corners' not tending to zero with increasing n, so that vectors exist that decay exponentially at rate η^k, have bounded norm, and when used in quadratic forms of the matrix difference produce non-zero results no matter how large n is.

Leaving this intuitive development aside, it may be considered that the assumption that f and g have finite dimensional spectral factorisation is too restrictive. In this case a more general result may be offered applicable to any Lipschitz continuous f and g, but at the expense of somewhat weakening the definition of equivalence over that discussed in § 1 to one in which two $n \times n$ matrices A_n and B_n are said to be asymptotically equivalent as $n \to \infty$ with notation $A_n \sim B_n$ as $n \to \infty$ if

$$\lim_{n\to\infty} \frac{\Gamma_n^\star(\omega)[A_n - B_n][A_n - B_n]^\star \Gamma_n(\omega)}{K_n(\omega,\omega)} = 0 \quad ; \forall \omega \in [-\pi, \pi].$$

Note that this refinement of the definition of matrix equivalence makes no difference for the system theoretic applications motivating this paper (see equation (43) and the accompanying discussion in § 7 following, or the work [37] for more detail on this point). With this definition in hand, the following result on the algebraic structure of generalised Toeplitz matrices is available.

Theorem 5.1. *Consider two not necessarily real valued functions f and g of which at least one of them is Lipschitz continuous of order $\varepsilon > 0$ and the other one bounded. Suppose that the poles $\{\xi_k\}$ of the basis functions $\{\mathcal{B}_k\}$ in (5) satisfy $|\xi_k| \le 1 - \delta$ for some $\delta > 0$. Then*

$$M_n(f)M_n(g) \sim M_n(fg) \quad \text{as } n \to \infty$$

with convergence rate faster than $O(\log^4 n / n^{\varepsilon/(\varepsilon+2)})$ as $n \to \infty$.

Proof. Without loss of generality assume that g is Lipschitz continuous and that f is bounded. From the definition (6) of $M_n(f)$ and the formulation (9) of $K_n(\omega, \sigma)$ it follows that with the definition

$$\Delta_n(\omega) \triangleq [M_n(f)M_n(g) - M_n(fg)]\,\Gamma_n(\omega)$$

then with the representation (6) and the formulation (9) in mind

$$\begin{aligned}
\Delta_n(\omega) &= \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_n(\mu)\Gamma_n^\star(\mu) f(\mu)\,\mathrm{d}\mu\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_n(\sigma)\Gamma_n^\star(\sigma) g(\sigma)\,\mathrm{d}\sigma\right)\Gamma_n(\omega) - \\
&\quad \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_n(\mu)\Gamma_n^\star(\mu) f(\mu) g(\mu)\,\mathrm{d}\mu\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_n(\sigma)\Gamma_n^\star(\sigma)\,\mathrm{d}\sigma\right)\Gamma_n(\omega) \\
&= \frac{1}{4\pi^2}\int_{-\pi}^{\pi} \Gamma_n(\mu) f(\mu) \int_{-\pi}^{\pi} \Gamma_n^\star(\mu)\Gamma_n(\sigma)\Gamma_n^\star(\sigma)\Gamma_n(\omega) g(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\mu - \\
&\quad \frac{1}{4\pi^2}\int_{-\pi}^{\pi} \Gamma_n(\mu) f(\mu) \int_{-\pi}^{\pi} \Gamma_n^\star(\mu)\Gamma_n(\sigma)\Gamma_n^\star(\sigma)\Gamma_n(\omega) g(\mu)\,\mathrm{d}\sigma\,\mathrm{d}\mu \\
&= \frac{1}{4\pi^2}\int_{-\pi}^{\pi} \Gamma_n(\mu) f(\mu) \int_{-\pi}^{\pi} K_n(\sigma,\mu) K_n(\omega,\sigma)[g(\sigma) - g(\mu)]\,\mathrm{d}\sigma\,\mathrm{d}\mu
\end{aligned}$$

or more compactly

$$\Delta_n(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_n(\mu) G_n(\omega,\mu) f(\mu)\,\mathrm{d}\mu \tag{27}$$

where the following definition has been used

$$G_n(\omega,\mu) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(\sigma,\mu)K_n(\omega,\sigma)[g(\sigma)-g(\mu)]\,\mathrm{d}\sigma. \tag{28}$$

Therefore, denoting when x is a vector the Euclidean norm of x as $\|x\|$ then

$$\|\Delta_n(\omega)\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{G_n(\omega,\lambda)}H_n(\omega,\lambda)\overline{f(\lambda)}\,\mathrm{d}\lambda \tag{29}$$

where

$$H_n(\omega,\lambda) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} G_n(\omega,\mu)K_n(\mu,\lambda)f(\mu)\,\mathrm{d}\mu. \tag{30}$$

However, by using Lemma 4.2

$$|K_n(\omega,\sigma)| \leq \begin{cases} \dfrac{2n}{\delta} & ;\forall\sigma,\omega \\[2ex] \dfrac{1}{|\sin(\omega-\sigma)/2|} & ;\omega\neq\sigma \end{cases} \tag{31}$$

so that Lemma A.1 may be applied to (28) with $f_n(\omega,\sigma) = K_n(\omega,\sigma)$, $g_n(\sigma,\mu) = K_n(\sigma,\mu)[g(\sigma)-g(\mu)]$ and $\beta=\gamma=1$ to conclude that

$$|G_n(\omega,\mu)| \leq \begin{cases} Cn^{2/(2+\varepsilon)} & ;\forall\omega,\mu \\[2ex] \dfrac{C}{|\sin(\omega-\mu)/2|}\log n & ;\omega\neq\mu \end{cases} \tag{32}$$

for n sufficiently large. Applying Lemma A.1 again, this time to (30) with the choices $f_n(\omega,\sigma) = K_n(\omega,\sigma)f(\omega)$, $g_n(\sigma,\mu) = G_n(\sigma,\mu)/\log n$ and $\beta = 1, \gamma = 2/(2+\varepsilon)$ provides

$$|H_n(\omega,\lambda)| \leq \begin{cases} Cn^{2/(2+\varepsilon)}\log^2 n & ;\forall\omega,\lambda \\[2ex] \dfrac{C}{|\sin(\omega-\mu)/2|}\log^2 n & ;\omega\neq\lambda. \end{cases}$$

Applying Lemma A.1 a final time, this time to (29) with the choices $f_n(\omega,\sigma) = H_n(\omega,\sigma)/\log^2 n$ and $g_n(\sigma,\mu) = G_n(\sigma,\mu)/\log n$, $\beta=1, \gamma = 2/(2+\varepsilon)$ provides

$$\|\Delta_n(\omega)\|^2 \leq Cn^{2/(2+\varepsilon)}\log^4 n$$

for some $C<\infty$ and for n sufficiently large. Therefore, under the assumption that $|\xi_k| < 1-\delta$ for some $\delta>0$, then by (19) $K_n(\omega,\omega)\geq \kappa n$ for some $\kappa>0$ so that for some $C<\infty$

$$\lim_{n\to\infty}\frac{\|\Delta_n(\omega)\|^2}{K_n(\omega,\omega)} \leq \lim_{n\to\infty}\frac{C\log^4 n}{n^{\varepsilon/(\varepsilon+2)}} = 0$$

for any $\varepsilon>0$ and the theorem is proved. □

Again, something more than this result is actually required in system theoretic applications where one is often concerned with multiple products that also contain matrix inverses. Such cases may be handled by the following corollary to the preceding result. In what follows, matrix products are to be interpreted in a left-to-right fashion as $\prod_{k=1}^{n} A_k = A_1 A_2 \cdots A_n$.

Corollary 5.1. *Suppose that the family of possibly complex valued functions* $\{f_k\}_{k=1}^{m}$ *are all Lipschitz continuous of order* $\varepsilon > 0$. *Suppose that the poles* $\{\xi_k\}$ *of the basis functions* $\{\mathcal{B}_k\}$ *in (5) satisfy* $|\xi_k| \leq 1 - \delta$ *for some* $\delta > 0$. *Then with* $\sigma_k = \pm 1$

$$\prod_{k=1}^{m} M_n^{\sigma_k}(f_k) \sim M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right) \quad \text{as } n \to \infty$$

with convergence rate faster than $O(\log^4 n / n^{\varepsilon/(\varepsilon+2)})$ *as* $n \to \infty$ *and provided the functions* $\{f_k\}$ *are invertible where required by the values of* σ_k.

Proof. An inductive argument will be used to obtain the result. Define the matrix difference

$$\Delta_n(m) \triangleq M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right) - \prod_{k=1}^{m} M_n^{\sigma_k}(f_k)$$

which may be re-expressed as

$$\Delta_n(m) = \Delta_n'(m) + \widetilde{\Delta}_n(m) \tag{33}$$

where

$$\begin{aligned} \Delta_n'(m) &\triangleq M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right) - \prod_{k=1}^{m} M_n(f_k^{\sigma_k}), \\ \widetilde{\Delta}_n(m) &\triangleq \prod_{k=1}^{m} M_n(f_k^{\sigma_k}) - \prod_{k=1}^{m} M_n^{\sigma_k}(f_k). \end{aligned}$$

The terms $\Delta_n'(m)$ and $\widetilde{\Delta}_n(m)$ will be considered separately. First, note that

$$\Delta_n'(m) = \Delta_n'(m-1) M_n(f_m^{\sigma_m}) + \left[M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right) - M_n\left(\prod_{k=1}^{m-1} f_k^{\sigma_k}\right) M_n(f_m^{\sigma_m}) \right] \tag{34}$$

and by Theorem 5.1 with the substitution $f = f_m^{\sigma_m}$, $g = \prod_{k=1}^{m-1} f_k^{\sigma_k}$ the second term in the above expression obeys

$$M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right) \sim M_n\left(\prod_{k=1}^{m-1} f_k^{\sigma_k}\right) M_n(f_m^{\sigma_m}) \quad \text{as } n \to \infty. \tag{35}$$

Furthermore, denoting when A is a matrix the spectral norm of A as $\|A\|$, then by Lemma 6.1 $\|M_n(f)\| \leq \|f\|_\infty$ so that by the Cauchy–Schwarz inequality

$$\left|\frac{\Gamma_n^\star(\omega)\Delta_n'(m-1)M_n(f_m^{\sigma_m})\Gamma_n(\omega)}{K_n(\omega,\omega)}\right|^2 \leq \|f_m^{\sigma_m}\|_\infty^2 \left|\frac{\Gamma_n^\star(\omega)\Delta_n'(m-1)[\Delta_n'(m-1)]^\star\Gamma_n(\omega)}{K_n(\omega,\omega)}\right|$$

so that $\Delta_n'(m) \sim 0$ as $n \to \infty$ if $\Delta_n'(m-1) \sim 0$ as $n \to \infty$. But $\Delta_n'(1) = 0$, so that by induction $\Delta_n'(m) \sim 0$ as $n \to \infty$ for any $m \geq 1$.

Now consider the term $\widetilde{\Delta}_n(m)$ and note that by labelling $k_1, \ldots k_r$ as being the indices k for which $\sigma_k = -1$, then $\widetilde{\Delta}_n(m)$ may be written as (if the lower index of a matrix product is greater than the upper index, then the product is understood to be equal to the identity matrix)

$$\widetilde{\Delta}_n(m) = \sum_{\ell=1}^{r}\prod_{k=1}^{k_\ell-1} M_n^{\sigma_k}(f_k)M_n^{-1}(f_{k_\ell})\left[M_n(f_{k_\ell})M_n(f_{k_\ell}^{-1}) - I\right]\prod_{k=k_\ell+1}^{m} M_n(f_k^{\sigma_k}).$$

This expression may be decomposed as $\widetilde{\Delta}_n(m) = \Sigma_n(m) + \Lambda_n(m)$ with the definitions

$$\begin{aligned}
\Sigma_n(m) &\triangleq \sum_{\ell=1}^{r}\prod_{k=1}^{k_\ell-1} M_n^{\sigma_k}(f_k)M_n^{-1}(f_{k_\ell})\left[M_n(f_{k_\ell})M_n(f_{k_\ell}^{-1}) - I\right] M_n\left(\prod_{k=k_\ell+1}^{m} f_k^{\sigma_k}\right),\\
\Lambda_n(m) &\triangleq \sum_{\ell=1}^{r}\prod_{k=1}^{k_\ell-1} M_n^{\sigma_k}(f_k)M_n^{-1}(f_{k_\ell})\left[M_n(f_{k_\ell})M_n(f_{k_\ell}^{-1}) - I\right] \times\\
&\quad \left[\prod_{k=k_\ell+1}^{m} M_n(f_k^{\sigma_k}) - M_n\left(\prod_{k=k_\ell+1}^{m} f_k^{\sigma_k}\right)\right].
\end{aligned}$$

Dealing with $\Sigma_n(m)$ and $\Lambda_n(m)$ in turn, note that

$$\begin{aligned}
\Sigma_n(m) &= \sum_{\ell=1}^{r}\prod_{k=1}^{k_\ell-1} M_n^{\sigma_k}(f_k)M_n^{-1}(f_{k_\ell}) \times\\
&\quad \left[M_n(f_{k_\ell})M_n(f_{k_\ell}^{-1})M_n\left(\prod_{k=k_\ell+1}^{m} f_k^{\sigma_k}\right) - M_n\left(\prod_{k=k_\ell+1}^{m} f_k^{\sigma_k}\right)\right]
\end{aligned}$$

and by Lemma 6.1, the sub-multiplicativity of the matrix norm and the continuity and positive definiteness assumptions on the f_k

$$\left\|M_n^{-1}(f_{k_\ell})\prod_{k=1}^{k_\ell-1} M_n^{\sigma_k}(f_k)\right\| \leq \prod_{k=1}^{k_\ell}\|f_k^{\sigma_k}\|_\infty < \infty$$

so that since it has been shown inductively that $\Delta'(m) \sim 0$ as $n \to \infty$ for any $m \geq 1$, then by the Cauchy–Schwarz inequality $\Sigma_n(m) \sim 0$ as $n \to \infty$ for any $n \geq 1$. Finally, again notice that $\Delta_n'(m) \sim 0$ as $n \to \infty$ implies that

$$\prod_{k=k_\ell+1}^{m} M_n(f_k^{\sigma_k}) \sim M_n\left(\prod_{k=k_\ell+1}^{m} f_k^{\sigma_k}\right) \quad \text{as } n \to \infty$$

so that once again using Lemma 6.1, the Cauchy–Schwarz inequality and the sub-multiplicativity of the matrix norm $\Lambda_n(m) \sim 0$ as $n \to \infty$ which completes the proof. □

Combining this corollary with Theorem 4.1 then provides a further corollary representing an extension of the generalised Fourier convergence of Theorem 4.1.

Corollary 5.2. *Suppose that the family of possibly complex valued functions $\{f_k\}_{k=1}^m$ are all Lipschitz continuous of order $\varepsilon > 0$. Suppose that the poles $\{\xi_k\}$ of the basis functions $\{\mathcal{B}_k\}$ in (5) satisfy $|\xi_k| \le 1 - \delta$ for some $\delta > 0$. Then the following limit result holds*

$$\lim_{n\to\infty} \frac{1}{K_n(\omega,\omega)} \Gamma_n^\star(\mu) \left(\prod_{k=1}^{m} M_n^{\sigma_k}(f_k)\right) \Gamma_m(\omega) = \begin{cases} \prod_{k=1}^{m} f_k^{\sigma_k}(\omega) & \mu = \omega, \\ 0 & \mu \neq \omega \end{cases}$$

for any $\omega \in [-\pi, \pi]$ and where $\sigma_k = \pm 1$ with the functions $\{f_k\}$ assumed to be invertible when required by the values of σ_k.

Proof.

$$\Gamma_n^\star(\mu) \left(\prod_{k=1}^{m} M_n^{\sigma_k}(f_k)\right) \Gamma_m(\omega) = \Gamma_n^\star(\mu) M_n \left(\prod_{k=1}^{m} f_k^{\sigma_k}\right) \Gamma_n(\omega) + \Gamma_n^\star(\mu)\Delta_n\Gamma_n(\omega)$$

where

$$\Delta_n \triangleq \prod_{k=1}^{m} M_n^{\sigma_k}(f_k) - M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right).$$

Now, by the Cauchy–Schwarz inequality

$$\left|\frac{\Gamma_n^\star(\mu)\Delta_n\Gamma_n(\omega)}{K_n(\omega,\omega)}\right|^2 \le \left|\frac{\Gamma_n^\star(\omega)\Delta_n\Delta_n^\star\Gamma_n(\omega)}{K_n(\omega,\omega)}\right| \left|\frac{K_n(\mu,\mu)}{K_n(\omega,\omega)}\right|.$$

However by Lemma 4.2, the lower bound $K_n(\omega,\omega) \ge \delta n/2$ applies and the upper bound $K_n(\mu,\mu) \le 2n/\delta$ also applies so that $|K_n(\mu,\mu)/K_n(\omega,\omega)| \le 1/\delta^2 < \infty$ independently of n. As well, by the previous Corollary 5.1

$$\Delta_n \sim 0 \quad \text{as } n \to \infty$$

so that

$$\lim_{n\to\infty} \left|\frac{\Gamma_n^\star(\mu)\Delta_n\Delta_n^\star\Gamma_n(\omega)}{K_n(\omega,\omega)}\right| = 0.$$

Use of Theorem 4.1 with the substitution $f = \prod_{k=1}^m f_k^{\sigma_k}$ then completes the proof. □

As a simple but important example of the utility of this corollary, it allows the conclusion that when all the poles $\{\xi_k\}$ are chosen in a closed subset of $\mathbf{D}$ then

$$\lim_{n\to\infty} \frac{\Gamma_n^\star(\omega)M_n^{-1}(f)\Gamma_n(\omega)}{K_n(\omega,\omega)} = \frac{1}{f(\omega)} \tag{36}$$

which has particular relevance to the study of reproducing kernels with respect to weighted inner products.

More specifically, the emphasis so far has been on studying the reproducing kernel $K_n(z,\mu)$ associated with the space $X_n = \text{Span}\{\mathcal{B}_0,\cdots,\mathcal{B}_{n-1}\}$ and with respect to the inner product (8). However, as was illustrated in § 3 via a linear prediction example, there is utility in examining a related kernel $K_n'(z,\mu)$ which is still associated with X_n, but exists with respect to an inner product which is (8) with the integrand weighted by a positive function $f(\omega)$. In this case, since

$$\langle \Gamma_n^\star(\mu)M_n^{-1}(f)\Gamma_n(z), \Gamma_n^T(z)\rangle = \Gamma_n^\star(\mu)M_n^{-1}(f)\langle \Gamma_n(z), \Gamma_n^T(z)\rangle = \Gamma_n^\star(\mu)$$

then $\langle \mathcal{B}_k(z), \Gamma_n^\star(\mu)M_n^{-1}(f)\Gamma_n(z)\rangle = \mathcal{B}_k(\mu)$ for every $k = 0,1,\cdots,n-1$ and hence since $K_n'(z,\mu)$ is the unique function in X_n with this property, then in fact $K_n'(z,\mu) = \Gamma_n^\star(\mu)M_n^{-1}(f)\Gamma_n(z)$ and so the asymptotic result (36) provides a means for providing the closed form approximation $K_n(\omega,\omega)/f(\omega)$ for $K_n'(\omega,\omega)$.

As a concluding remark for this section, it should be noted that the only other work known to the authors which addresses issues of generalising Fourier convergence and asymptotic Toeplitz matrix properties in a similar context to this paper is that of Gunnarsson and Ljung [15, 16] wherein the generalisation involves matrices not necessarily being Toeplitz, but at least being approximately so in some sense. The details of this are such that the generalised matrices (and hence generalised Fourier convergence) considered in [15, 16] are quite different to those of the form $M_n(f)$ considered here, which are not approximately Toeplitz in any sense except that the spectral formulation (6) is reminiscent of the classical Toeplitz one (1).

6 Spectral Properties of Generalised Toeplitz Matrices

A typical system identification application of the orthonormal bases (5) would be to seek a parameter vector $\theta \in \mathbf{R}^n$ in order to model the input-output relationship between N samples of an observed input sequence $\{u_k\}$ and output sequence $\{y_k\}$ as [46, 47, 40, 36]

$$y_k = \sum_{n=0}^{m-1} \theta_n \mathcal{B}_n(q)u_k = \phi_k^T\theta, \quad k = 0,1,\cdots,N-1 \tag{37}$$

where q is the backward time shift operator and

$$\phi_k^T \triangleq [\mathcal{B}_0(q)u_k, \mathcal{B}_1(q)u_k, \cdots, \mathcal{B}_{n-1}(q)u_k]$$

is a vector of filtered versions of the signal $\{u_k\}$, the filtering depending on the orthonormal basis functions chosen. The least-squares solution $\widehat{\theta}$ for θ is then given as

$$\widehat{\theta} = R_N^{-1}\frac{1}{N}\sum_{k=0}^{N-1}\phi_k y_k, \quad R_N \triangleq \frac{1}{N}\sum_{k=0}^{N-1}\phi_k\phi_k^T$$

provided that R_N exists. It is well known [13] that the numerical robustness of solving for $\widehat{\theta}$ is intimately related to the condition number $\kappa(R_N)$ of R_N. By Parseval's Theorem, for large N the matrix R_N converges as [27]

$$\lim_{N\to\infty} R_N = M_n(f)$$

where f is the spectral density of the observed input $\{u_k\}$ which, if containing deterministic components, is defined in the sense of Wiener's generalised harmonic analysis [51] or the quasi-stationarity sense of Ljung [27]. From a numerical point of view there is therefore significant practical relevance in examining the spectrum of generalised Toeplitz matrices.

For the classical trigonometric case wherein $\xi_k = 0$ and $M_n(f)$ is in fact a bona-fide Toeplitz matrix, it is well known [14, 50] that the eigenvalues of $M_n(f)$ may be bounded above and below by the maximum and minimum values of f. This result can be easily extended to the general case, the classical case again emerging as a special case by setting $\xi_k = 0$.

Lemma 6.1. *For continuous and real-valued $f(\omega) > 0$ let $M_n(f)$ be defined by (6). Then*

$$\min_{\omega\in[-\pi,\pi]} f(\omega) \le \lambda(M_n(f)) \le \max_{\omega\in[-\pi,\pi]} f(\omega)$$

where $\lambda(A)$ is any eigenvalue of the matrix A. In the case that f is complex valued, then the upper bound

$$|\lambda(M_n(f))| \le \max_{\omega\in[-\pi,\pi]} |f(\omega)|$$

applies.

Proof. Consider the case of real valued $f > 0$ first and take $x \in \mathbf{R}^n$ arbitrary but such that $x^\star x = 1$. Then

$$\begin{aligned}
x^\star M_n(f)x &= \frac{1}{2\pi}\sum_{r=0}^{n-1}\sum_{k=0}^{n-1}\overline{x_r}x_k\int_{-\pi}^{\pi}\mathcal{B}_r(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}f(\omega)\,\mathrm{d}\omega,\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)\left|\sum_{r=0}^{n-1}x_r\mathcal{B}_r(e^{j\omega})\right|^2\mathrm{d}\omega,\\
&\le \max_{\omega\in[-\pi,\pi]}\frac{f(\omega)}{2\pi}\sum_{r=0}^{n-1}\sum_{k=0}^{n-1}\overline{x_r}x_k\int_{-\pi}^{\pi}\mathcal{B}_r(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}\,\mathrm{d}\omega\\
&= \max_{\omega\in[-\pi,\pi]} f(\omega).
\end{aligned}$$

But since $M_n(f)$ is symmetric positive definite then

$$\max_{x^\star x=1} x^\star M_n(f)x = \lambda_{\max}(M_n(f)).$$

Using a similar argument to under-bound the eigenvalues of $M_n(f)$ then completes the first part of the lemma. For the case of complex valued f note that the upper bound can be generated as a slight modification to the above reasoning as

$$|x^\star M_n(f)x| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\omega)| \left|\sum_{r=0}^{n-1} x_r \mathcal{B}_r(e^{j\omega})\right|^2 \mathrm{d}\omega \leq \max_{\omega\in[-\pi,\pi]} |f(\omega)|.$$

□

This result provides a guaranteed over-bound

$$\kappa(M_n(f)) \triangleq \frac{\lambda_{\max}(M_n(f))}{\lambda_{\min}(M_n(f))} \leq \frac{\max_{\omega\in[-\pi,\pi]} f(\omega)}{\min_{\omega\in[-\pi,\pi]} f(\omega)} \tag{38}$$

on the condition number $\kappa(M_n(f))$ governing the numerical robustness of the least squares estimation problem. The existence of this over-bound has been one of the prime motivators for the recent interest in the use of orthonormal bases such as (5) for system identification applications [46, 47, 18, 40].

It is then natural to examine how conservative the bound (38) is. Considering that numerical robustness is of greatest concern when the dimension n is large, it is not unreasonable to simplify the examination of conservatism by letting $n \to \infty$. This allows the following result showing that for large n the bounds in Lemma 6.1 are tight so that the condition number of $M_n(f)$ actually achieves the bound (38).

Theorem 6.1. *Define for continuous $f(\omega) > 0$ the operator $M(f) : \{x_k\} \in \ell_2^+ \mapsto \{y_k\} \in \ell_2^+$ as*

$$M(f) \triangleq \lim_{n\to\infty} M_n(f)$$

where this is understood to mean that the infinite sequence $\{y_k\}$ is generated from the infinite sequence $\{x_k\}$ as a natural limit of how finite length n $\{y_k\}$ are generated from finite length n $\{x_k\}$ via matrix multiplication by $M_n(f)$. Specifically

$$y_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{B}_k(e^{j\omega})\overline{g(\omega)}\,\mathrm{d}\omega, \quad g(\omega) \triangleq f(\omega)\sum_{k=0}^{\infty} x_k \mathcal{B}_k(e^{j\omega}).$$

Then provided

$$\sum_{k=0}^{\infty}(1 - |\xi_k|) = \infty$$

it holds that

$$\lambda(M(f)) \triangleq \{\lambda \in \mathbf{C} : M(f) - \lambda I \textit{ is not invertible}\} = \mathrm{Range}\{f(\omega)\}$$

Proof. Take any $\mu \in [\min_\omega f(\omega), \max_\omega f(\omega)]$ and suppose $\mu \notin \lambda(M)$. Then since by orthogonality $M(1)$ is the identity operator, then $M(f-\mu)$ is an invertible operator from $\ell_2^+ \to \ell_2^+$ so that in particular $\exists x \in \ell_2^+$ such that for $e_0 = (1, 0, 0, \cdots)$

$$x^T M(f-\mu) = e_0. \tag{39}$$

Therefore, defining $g(z) \in H_2(\mathbf{T})$ by $g(z) \triangleq (x_0 - 1)\mathcal{B}_0(z) + \sum_{k=1}^{\infty} x_k B_k(z)$ gives that from (39) $(f-\mu)g \perp \text{Span}\{\mathcal{B}_k\}$. But under the condition $\sum(1-|\xi_k|) = \infty$ by Theorem 2.1, $\text{Span}\{\mathcal{B}_k\} = H_2$ so that since $L_2 = H_2 \oplus H_2^\perp$ then $(f-\mu)\overline{g} \in H_2$. But $g \in H_2$ by construction and the product of two H_2 functions is in H_1 [19]. Therefore $(f-\mu)|g|^2$ is a real valued H_1 function, and the only such functions are constants [19]. However, since $\mu \in [\min_\omega f(\omega), \max_\omega f(\omega)]$ then this function cannot be of constant sign, hence it cannot be a constant. This contradiction implies $[\min_\omega f(\omega), \max_\omega f(\omega)] \subset \lambda(M)$. Finally, Lemma 6.1 gives that $\lambda(M) \subset [\min_\omega f(\omega), \max_\omega f(\omega)]$. □

Again, this result represents an expansion to the case of generalised Toeplitz matrices (6) of results already known for conventional symmetric Toeplitz matrices [14, 50], the latter results being encompassed as a special case of Theorem 6.1. A practical conclusion arising from this theorem is that the numerical properties of the solution of (37) are governed solely by the spectral density f of $\{u_k\}$ and are independent of the particular orthonormal basis chosen in (37) via the selection of $\{\xi_k\}$.

Another conclusion arising from Theorem 6.1 is that in the previous results on the asymptotic algebraic structure of generalised Toeplitz matrices (a main conclusion of which was to conclude that $M_n^{-1}(f) \sim M_n(1/f)$ as $n \to \infty$) the assumptions imposed there of f being invertible cannot be weakened.

7 Applications

To put these results in context, this section presents a very brief outline of how they may be applied to certain system identification problems that have been previously alluded to. A much more detailed exposition of the issues raised here is contained in [37].

A result that has become of key importance to the intuitive understanding of various system identification methods [2, 12, 28, 17] is that the variability of the frequency response $G(e^{j\omega}, \widehat{\theta})$ of a model based on a least-squares estimate $\widehat{\theta}$ of an n dimensional parameter vector θ obtained from N observations of noise corrupted input-output measurements is approximately given by [31, 29]

$$\text{Var}\{G(e^{j\omega}, \widehat{\theta})\} \approx \frac{n}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)} \tag{40}$$

where $\Phi_u(\omega)$ is the spectral density of the observed input process and $\Phi_\nu(\omega)$ is the spectral density of the noise corrupting process.

This is established in [29] for a wide range of model structures, unified by the requirement that a certain 'shift' structure holds. In the interests of clarity, the discussion here will focus only on the case of measurement noise being white, and the true system (including noise model) being encompassed by the model structure. In this case, it is established in [29] that with $\Phi_\nu(\omega) = \sigma^2$ being the constant white noise spectral density, then for large N

$$\frac{N}{n}\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\} \approx \frac{1}{n}\Gamma_n^\star(\omega)T_n^{-1}(\Phi_u)T_n(\sigma^2\Phi_u)T_n^{-1}(\Phi_u)\Gamma_n(\omega). \tag{41}$$

The reasoning of [29] then is that using the previously discussed classical results [14, 50] on the asymptotic algebraic structure of Toeplitz matrices it can be argued that $T_n^{-1}(\Phi_u)T_n(\sigma^2\Phi_u)T_n^{-1}(\Phi_u) \approx T_n(\sigma^2/\Phi_u)$ so that the right hand side of (41) is approximately an n'th order Cesàro mean reconstruction of σ^2/Φ_u at frequency ω and hence should be approximately equal to $\sigma^2/\Phi_u(\omega)$.

However, it is very common in the interests of concentrating estimation accuracy in certain frequency regions, to pre-filter the measured data [27]. Identifying this filter with its transfer function $F(z)$ then allows the nature of the pre-filtering to be characterised by the spectral density change of $\Phi_u(\omega) \mapsto |F(e^{j\omega})|^2\Phi_u(\omega)$. However, at the same time the use of data pre-filtering implies a revision of the 'noise model' as [29] $\Phi_\nu(\omega) \mapsto |F(e^{j\omega})|^2\Phi_\nu(\omega)$. In this case, the approximate estimation variability would be expected to be unchanged from (40) by reasoning

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\} \approx \frac{n}{N}\frac{|F(e^{j\omega})|^2\Phi_\nu(\omega)}{|F(e^{j\omega})|^2\Phi_u(\omega)} = \frac{n}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)}. \tag{42}$$

However, as illustrated numerically in [37], the accuracy of this approximation depends very much on the relationship between the order of the filter F and the model order. The closer the two orders the more inaccurate the approximation.

In [37] this phenomenon is traced to the fact that as the filter order grows, the underlying Fourier reconstruction involved in (40) and hence (42) is with respect to a function with decreasing smoothness, and hence the more terms (which grows with model order) required in the Fourier expansion before convergence will approximately occur.

To circumvent this problem, a key observation of [37] is that the model can be re-parameterised into one in which the pre-filter poles are absorbed into the model structure. In this case (41) is replaced with

$$\frac{N}{K_n(\omega,\omega)}\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\} \approx \frac{\Gamma_n^\star(\omega)M_n^{-1}(\Phi_u)M_n(\sigma^2\Phi_u)M_n^{-1}(\Phi_u)\Gamma_n(\omega)}{K_n(\omega,\omega)} \tag{43}$$

where the poles of the basis functions forming the generalised Toeplitz matrix $M_n(\Phi_u)$ are chosen the same as the pre-filter. In this case, using the results of this paper it can be argued using Corollary 5.1 that $M_n^{-1}(\Phi_u)M_n(\sigma^2\Phi_u)M_n^{-1}(\Phi_u) \sim M_n(\sigma^2/\Phi_u)$ as $n \to \infty$ so that

$$\frac{N}{K_n(\omega,\omega)}\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\} \approx \frac{1}{K_n(\omega,\omega)}\Gamma_n^\star(\omega)M_n(\sigma^2/\Phi_u)\Gamma_n(\omega) \tag{44}$$

which is a generalised Fourier reconstruction of a function σ^2/Φ_u that is *invariant* to the choice of pre-filter. Therefore, using Theorem 4.1 to argue that (44) is approximately

$\sigma^2/\Phi_u(\omega)$ the work [37] is able to suggest that

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\} \approx \frac{K_n(\omega,\omega)}{N}\frac{\Phi_\nu(\omega)}{\Phi_u(\omega)} \tag{45}$$

is a more accurate approximation when all-pole pre-filtering is employed and the ratio of model order to filter order is low. In other words, data pre-filtering can affect the variability of the estimated model, and (45) quantifies how this occurs. The validity of these conclusions are illustrated numerically in [37]. As a finally applications oriented remark, note that when all poles $\{\xi_k\}$ are chosen at the origin (which corresponds to no pre-filtering), then $K_n(\omega,\omega) = n$ so that the new approximation (45) becomes the 'classical' one (40) as a special case.

8 Conclusion

The purpose of this paper was to consider certain results in the study of Fourier series and Toeplitz matrices that have proved to be key to various system theoretic applications, and expand them to the case where the underlying orthonormal basis is not the classical trigonometric one, but a rational formulation that encompasses the trigonometric basis as a special case. These results, and the ensuing generalisations developed in this paper are summarised in Table 1.

One point worth clarifying, is that in system theoretic settings for which these results will be applicable (control, signal processing, system identification) it is more common to associate the complex variable z with a forward time shift, rather than the backward shift association used here. This discrepancy is easily accommodated by simply transforming $z \mapsto 1/z$ in all the results presented here. A different basis function definition will result, which is in accordance with certain so-called Laguerre and Kautz bases studied in the control theory literature. However, the matrices $M_n(f)$ and the associated Fourier reconstruction formulas will be unchanged.

A Bounds on Integrals of Kernel–Like Functions

Throughout this appendix, C will denote a finite positive constant which may be different in different places of the same expression.

Lemma A.1. *Let $f_n(\omega,\sigma) : [-\pi,\pi]\times[-\pi,\pi] \to \mathbf{C}$ be subject to*

$$|f_n(\omega,\sigma)| \leq \begin{cases} Cn^\beta & ;\forall\omega,\sigma \\ \dfrac{C}{|\sin(\omega-\sigma)/2|} & ;\omega\neq\sigma \end{cases} \tag{A.1}$$

	Classical	Generalised
Basis	$e^{j\omega n}$	$\mathcal{B}_n(e^{j\omega}) \triangleq \frac{\sqrt{1-\lvert\xi_n\rvert^2}}{1-\xi_n e^{j\omega}} \prod_{k=0}^{n-1} \frac{e^{j\omega}-\overline{\xi_k}}{1-\xi_k e^{j\omega}}$
Completeness	$H_2(\mathbf{T})$	$H_2(\mathbf{T})$ provided $\sum(1-\lvert\xi_n\rvert)=\infty$
Assoc. Matrix	Toeplitz Matrix $[T_n(f)]_{k,\ell} = \int_{-\pi}^{\pi} e^{j\omega(k-\ell)} f(\omega)\,\frac{\mathrm{d}\omega}{2\pi}$	Generalised Toeplitz matrix $[M_n(f)]_{k,\ell} = \int_{-\pi}^{\pi} \mathcal{B}_k(e^{j\omega})\overline{\mathcal{B}_\ell(e^{j\omega})} f(\omega)\,\frac{\mathrm{d}\omega}{2\pi}$
Cesàro Mean	$f_n(\omega) = \frac{1}{n}\sum_{k,\ell=0}^{n-1} e^{j\omega(\ell-k)}[T_n(f)]_{k,\ell}$	$f_n(\omega) = \sum_{k,\ell=0}^{n-1} \frac{\mathcal{B}_k(e^{j\omega})\overline{\mathcal{B}_\ell(e^{j\omega})}}{K_n(\omega,\omega)}[M_n(f)]_{k,\ell}$
Convergence	$\lim_{n\to\infty} \sup_{\omega\in[-\pi,\pi]} \lvert f(\omega)-f_n(\omega)\rvert = 0$	$\lim_{n\to\infty} \sup_{\omega\in[-\pi,\pi]} \lvert f(\omega)-f_n(\omega)\rvert = 0$
Def. equivalence $A_n \sim B_n$ as $n\to\infty$	$\lim_{n\to\infty} \lvert A_n - B_n\rvert = 0$	$\lim_{n\to\infty} \frac{\lVert (A_n-B_n)\Gamma_n(\omega)\rVert^2}{K_n(\omega,\omega)} = 0$
Algebraic Properties	$T_n(f)T_n(g) \sim T_n(fg)$	$M_n(f)M_n(g) \sim M_n(fg)$
Extensions $\sigma_k = \pm 1$	$\prod_{k=1}^{m} T_n^{\sigma_k}(f_k) \sim T_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right)$	$\prod_{k=1}^{m} M_n^{\sigma_k}(f_k) \sim M_n\left(\prod_{k=1}^{m} f_k^{\sigma_k}\right)$

Table 1: *Summary of classical results and their relation to the generalisations derived here.*

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

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

for some $\gamma, \varepsilon \geq 0$ *and where them meaning of* $d(\sigma,\mu)$ *is defined in (20). Then for* n *sufficiently large.*

$$\left|\int_{-\pi}^{\pi} f_n(\omega,\sigma) g_n(\sigma,\mu)\, \mathrm{d}\sigma\right| \leq \begin{cases} C\min(n^{\lambda}, n^{\delta}\log n) & ;\forall \omega,\mu \\ \dfrac{C}{|\sin(\omega-\mu)/2|}\log n & ;\omega \neq \mu \end{cases} \tag{A.3}$$

where

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

Proof. Suppose, to begin with, that $\omega \neq \mu$. Then for any $\alpha > 0$, $d(\omega,\mu) \geq 6n^{-\alpha}$ for sufficiently large n. Assuming this is the case, define the following regions:

$$\begin{aligned} \Omega_1 &\triangleq \left\{\sigma \in [-\pi,\pi] : d(\omega,\sigma) < n^{-\alpha}\right\}, \\ \Omega_2 &\triangleq \left\{\sigma \in [-\pi,\pi] : d(\mu,\sigma) \leq n^{-\alpha}\right\}, \\ \Omega_3 &\triangleq \left\{\sigma \in [-\pi,\pi] : \sigma \notin \{\Omega_1 \cup \Omega_2\}\right\}. \end{aligned}$$

Therefore, using the assumed bounds on $f_n(\omega,\sigma), g_n(\sigma,\mu)$, noticing that by definition the regions Ω_1 and Ω_2 are disjoint, and using Lemma B.1 with $\varepsilon = 0$ leads to

$$\left|\int_{\Omega_1} f_n(\omega,\sigma) g_n(\sigma,\mu)\, \mathrm{d}\sigma\right| \leq \int_{\Omega_1} \frac{Cn^{\beta}}{|\sin(\mu-\sigma)/2|}\, \mathrm{d}\sigma \leq C\frac{n^{\beta-\alpha}}{|\sin(\mu-\omega)/2|}.$$

Similarly

$$\left|\int_{\Omega_2} f_n(\omega,\sigma) g_n(\sigma,\mu)\, \mathrm{d}\sigma\right| \leq \int_{\Omega_2} \frac{Cn^{\gamma}}{|\sin(\omega-\sigma)/2|}\, \mathrm{d}\sigma \leq C\frac{n^{\gamma-\alpha}}{|\sin(\mu-\omega)/2|}.$$

Finally, this time using Lemma B.2

$$\left|\int_{\Omega_3} f_n(\omega,\sigma) g_n(\sigma,\mu)\, \mathrm{d}\sigma\right| \leq \int_{\Omega_3} \frac{C}{|\sin(\sigma-\omega)/2\sin(\sigma-\mu)/2|}\, \mathrm{d}\sigma \leq \frac{C\alpha}{|\sin(\mu-\omega)|}\log n$$

so that for $\omega \neq \mu$, choosing $\alpha = \max(\beta,\gamma)$ provides the bound

$$\left|\int_{-\pi}^{\pi} f_n(\omega,\sigma) g_n(\sigma,\mu)\, \mathrm{d}\sigma\right| \leq \frac{C}{|\sin(\omega-\mu)/2|}\log n.$$

for sufficiently large n. Now assume that $\omega = \mu$. In this case, again using the assumed bounds on $f_n(\omega,\sigma)$ and $g_n(\sigma,\mu)$

$$\left|\int_{\Omega_1} f_n(\omega,\sigma)g_n(\sigma,\mu)\,\mathrm{d}\sigma\right| \le Cn^{\beta+\gamma}\int_{\Omega_1} d^\varepsilon(\sigma,\mu)\,\mathrm{d}\sigma \le Cn^{\beta+\gamma}\int_0^{n^{-\alpha}} x^\varepsilon\,\mathrm{d}x = Cn^{\beta+\gamma-\alpha(\varepsilon+1)}.$$

Also, using Lemma B.2

$$\left|\int_{\Omega_3} f_n(\omega,\sigma)g_n(\sigma,\mu)\,\mathrm{d}\sigma\right| \le \int_{\Omega_3}\frac{C}{\sin^2(\sigma-\omega)/2}\,\mathrm{d}\sigma \le \frac{C}{\sin n^{-\alpha}} \le Cn^\alpha,$$

so that when $\omega = \mu$ and hence $\Omega_1 = \Omega_2$, then for sufficiently large n

$$\left|\int_{-\pi}^{\pi} f_n(\omega,\sigma)g_n(\sigma,\mu)\,\mathrm{d}\sigma\right| \le Cn^{\beta+\gamma-\alpha(\varepsilon+1)} + Cn^\alpha.$$

This bound is minimised (for large n) by the choice $\alpha = (\beta+\gamma)/(2+\varepsilon)$ as $Cn^{(\beta+\gamma)/(2+\varepsilon)}$. Alternatively, with the definition $\delta \triangleq \min(\beta,\gamma)$ the integral on Ω_3 can also be bounded using Lemma B.2 as

$$\left|\int_{\Omega_3} f_n(\omega,\sigma)g_n(\sigma,\mu)\,\mathrm{d}\sigma\right| \le \int_{\Omega_3}\frac{Cn^\delta}{|\sin(\sigma-\omega)/2|}\,\mathrm{d}\sigma \le C\alpha n^\delta\log n$$

to give the bound for $\omega = \mu$ and sufficiently large n of

$$\left|\int_{-\pi}^{\pi} f_n(\omega,\sigma)g_n(\sigma,\mu)\,\mathrm{d}\sigma\right| \le Cn^{\beta+\gamma-\alpha(\varepsilon+1)} + C\alpha n^\delta\log n.$$

This bound is minimised (assuming without loss of generality that $\beta = \min(\beta,\gamma)$) by the choice (for large n) $\alpha = \gamma/(1+\varepsilon)$ as

$$\left|\int_{-\pi}^{\pi} f_n(\omega,\sigma)g_n(\sigma,\mu)\,\mathrm{d}\sigma\right| \le Cn^\delta\log n.$$

Note that this latter bound will be smaller than the previous one whenever $\beta+\gamma > (2+\varepsilon)\min(\beta,\gamma)$ and for n sufficiently large. □

B Integrals of Reciprocals of Sine Functions

Lemma B.1. *Let $0 < \alpha < \pi/8$ and suppose that $\omega,\mu \in [-\pi,\pi]$ satisfies $d(\mu,\omega) \ge 4\alpha$. Then for any $\Omega \subset [\mu-\alpha,\mu+\alpha]$, $\varepsilon \ge 0$ and where the meaning of $d(\mu,\omega)$ is defined in (20)*

$$\int_\Omega \frac{d^\varepsilon(\sigma,\mu)}{|\sin(\sigma-\omega)/2|}\,\mathrm{d}\sigma \le \frac{16\alpha^{\varepsilon+1}}{(\varepsilon+1)|\sin(\mu-\omega)/2)|}.$$

Proof. It holds that

$$\begin{aligned}\sin\left(\frac{\mu-\omega}{2}+\frac{x}{2}\right) &= \cos\left(\frac{x}{2}\right)\sin\left(\frac{\mu-\omega}{2}\right)+\sin\left(\frac{x}{2}\right)\cos\left(\frac{\mu-\omega}{2}\right)\\ &= \cos\left(\frac{x}{2}\right)\sin\left(\frac{\mu-\omega}{2}\right)\left(1+\tan\left(\frac{x}{2}\right)\cot\left(\frac{\mu-\omega}{2}\right)\right), \qquad \text{(B.1)}\end{aligned}$$

and for $x \in [-\alpha, \alpha]$

$$1+\tan\left(\frac{x}{2}\right)\cot\left(\frac{\mu-\omega}{2}\right) > 1 - \frac{\alpha/2}{\cos(\pi/8)}\,\frac{4}{d(\mu,\omega)} > 1/4. \qquad \text{(B.2)}$$

Without loss of generality, assume Ω is such that $d(\sigma,\mu) = |\sigma-\mu|$ on Ω in which case the change of variables $x = \sigma - \mu$ together with (B.1)–(B.2) then gives

$$\begin{aligned}\int_\Omega \frac{d^\varepsilon(\sigma,\mu)}{|\sin(\sigma-\omega)/2|}\,\mathrm{d}\sigma &\leq \int_{-\alpha}^{\alpha}\frac{|x|^\varepsilon}{|\sin(\mu-\omega+x)/2|}\,\mathrm{d}x\\ &\leq \frac{1}{|\sin(\mu-\omega)/2|}\int_{-\alpha}^{\alpha}\frac{|x|^\varepsilon}{|\cos x/2|}\,\frac{1}{|1+\tan(x/2)\cot(\mu-\omega)/2|}\,\mathrm{d}x\\ &\leq \frac{1}{|\sin(\mu-\omega)/2|}\int_{-\alpha}^{\alpha} 2\,\times\,4|x|^\varepsilon\,\mathrm{d}x = \frac{16\alpha^{1+\varepsilon}}{(1+\varepsilon)|\sin(\mu-\omega)/2|}.\end{aligned}$$

□

Lemma B.2. *Let $-\pi \leq \alpha < \beta \leq \pi$ and suppose that $\omega \in [-\pi,\pi]$ does not belong to $[\alpha,\beta]$. Then*

$$\int_\alpha^\beta \frac{1}{|\sin(\sigma-\omega)/2|}\,\mathrm{d}\sigma \leq 4\log\frac{8}{\gamma} \qquad \text{(B.3)}$$

where $\gamma \triangleq d(\omega,[\alpha,\beta])$. Suppose also that $\mu \in [-\pi,\pi]$ does not belong to $[\alpha,\beta]$. Then

$$\int_\alpha^\beta \frac{1}{|\sin(\sigma-\omega)/2\sin(\sigma-\mu)/2|}\,\mathrm{d}\sigma \leq \begin{cases} \dfrac{8}{|\sin(\mu-\omega)/2|}\log\dfrac{4}{\gamma} & ;\omega\neq\mu\\[2ex] \dfrac{4}{\sin\gamma} & ;\omega=\mu\end{cases} \qquad \text{(B.4)}$$

where in this latter case $\gamma \triangleq d(\{\omega,\mu\},[\alpha,\beta])$.

Proof. To begin with assume that $\omega < \alpha$. The change of variables $x = (\sigma-\omega)/2$ then gives

$$\int_\alpha^\beta \frac{1}{|\sin(\sigma-\omega)/2|}\,\mathrm{d}\sigma = 2\int_{\alpha'}^{\beta'}\frac{1}{\sin x}\,\mathrm{d}x \qquad \text{(B.5)}$$

where $0 < \alpha' \triangleq (\alpha - \omega)/2 < \beta' \triangleq (\beta - \omega)/2 < \pi$. Since $\sin(\beta'/2) > \sin(\alpha'/2)$ and $\cos(\alpha'/2) > \cos(\beta'/2)$ it follows from (B.5) that

$$\begin{aligned}
\int_\alpha^\beta \frac{1}{|\sin(\sigma-\omega)/2|}\,\mathrm{d}\sigma &= 2\left[\log\left(\tan(x/2)\right)\right]_{\alpha'}^{\beta'}, \\
&= 2\log\left(\frac{\sin(\beta'/2)}{\sin(\alpha'/2)}\right) + 2\log\left(\frac{\cos(\alpha'/2)}{\cos(\beta'/2)}\right) \\
&\le 2\log\left(\frac{1}{\sin(\alpha'/2)}\right) + 2\log\left(\frac{1}{\cos(\beta'/2)}\right) \\
&= 2\log\left(\frac{1}{\sin(\alpha'/2)}\right) + 2\log\left(\frac{1}{\sin(\pi/2-\beta'/2)}\right) \\
&\le 2\log\left(\frac{4}{d(\alpha',0)}\right) + 2\log\left(\frac{4}{d(\pi-\beta',0)}\right) \\
&= 2\log\left(\frac{4}{d(\alpha',0)}\right) + 2\log\left(\frac{4}{d(\beta',0)}\right) \\
&\le 4\log\left(\frac{8}{\gamma}\right).
\end{aligned} \tag{B.6}$$

The case where $\beta < \omega$ follows analogously and the proof of the bound (B.3) is complete. Moving on to the proof of the bound (B.4), consider first the case $\omega \neq \mu$ and assume that

$$-\pi \le \omega < \mu < \alpha < \beta < \pi. \tag{B.7}$$

Let $0 < \alpha' \triangleq (\alpha - \mu)/2 < \beta' \triangleq (\beta - \mu)/2 < \pi$. The change of variables $x = (\sigma - \mu)/2$ then gives

$$\begin{aligned}
&\int_\alpha^\beta \frac{1}{|\sin(\sigma-\omega)/2\sin(\sigma-\mu)/2|}\,\mathrm{d}\sigma = \\
&= 2\int_{\alpha'}^{\beta'} \frac{1}{\sin\left((\mu-\omega)/2+x\right)\sin(x)}\,\mathrm{d}\sigma \\
&= 2\int_{\alpha'}^{\beta'} \frac{1}{\sin\left(\pi-(\mu-\omega)/2-x\right)\sin(x)}\,\mathrm{d}\sigma \\
&= \frac{4}{\sin\left(\pi-(\mu-\omega)/2\right)}\left[\log\left(\frac{\sin(x)}{\sin\left(\pi-(\mu-\omega)/2-x\right)}\right)\right]_{\alpha'}^{\beta'} \\
&= \frac{4}{\sin(\mu-\omega)/2}\left(\log\left(\sin(\beta')\right) - \log\left(\sin\left(\pi-\frac{\mu-\omega}{2}-\beta'\right)\right)\right. \\
&\quad \left. - \log\left(\sin(\alpha')\right) + \log\left(\sin\left(\pi-\frac{\mu-\omega}{2}-\alpha'\right)\right)\right) \\
&= \frac{4}{\sin(\mu-\omega)2}\left(\log\left(\sin\left(\frac{\beta-\mu}{2}\right)\right) - \log\left(\sin\left(\frac{\beta-\omega}{2}\right)\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \log\left(\sin\left(\frac{\alpha-\mu}{2}\right)\right) + \log\left(\sin\left(\frac{\alpha-\omega}{2}\right)\right)\Bigg) \\
&\leq \frac{4}{\sin(\mu-\omega)/2}\left(-\log\left(\sin\left(\frac{\beta-\omega}{2}\right)\right) - \log\left(\sin\left(\frac{\alpha-\mu}{2}\right)\right)\right) \\
&= \frac{4}{\sin(\mu-\omega)/2}\left(\log\left(\frac{1}{\sin(\beta-\omega)/2}\right) + \log\left(\frac{1}{\sin(\alpha-\mu)/2}\right)\right) \\
&\leq \frac{4}{\sin(\mu-\omega)/2}\left(\log\left(\frac{4}{d(\beta,\omega)}\right) + \log\left(\frac{4}{d(\alpha,\mu)}\right)\right) \\
&\leq \frac{8}{\sin(\mu-\omega)/2}\log\left(\frac{4}{\gamma}\right) \qquad \text{(B.8)}
\end{aligned}
$$

where use of (21) was made in the second last inequality. This proves the lemma for the case (B.7). The other cases for $\omega \neq \mu$ follow analogously. Now suppose $\omega = \mu$. Let $0 < \alpha' \triangleq (\alpha - \omega)/2 < \beta' \triangleq (\beta - \omega)/2 < \pi$. Then the change of variables $x = (\sigma - \omega)/2$ gives

$$
\int_{\alpha}^{\beta} \frac{1}{\sin^2(\sigma-\omega)/2}\, d\sigma = 2\int_{\alpha'}^{\beta'} \frac{1}{\sin^2 x} dx = \left[-2\,\frac{\cos x}{\sin x}\right]_{\alpha'}^{\beta'} \leq \frac{4}{\sin\gamma}. \qquad \text{(B.9)}
$$

which proves the lemma when $\omega = \mu$. □

References

[1] N. Aronszajn, *Theory of reproducing kernels*, Acta Mathematica, (1950), pp. 337–404.

[2] R. Bitmead, M. Gevers, and V. Wertz, *Adaptive Optimal Control, The Thinking Man's GPC.*, Prentice Hall, Englewood Cliffs, New Jersey, 1990.

[3] P. W. Broome, *Discrete orthonormal sequences*, Journal of the Association for Computing Machinery, 12 (1965), pp. 151–168.

[4] E. Cheney, *Introduction to Approximation Theory*, McGraw Hill, New York, 1966.

[5] C. K. Chui, *Wavelets:A Tutorial in Theory and Applications*, vol. 2 of Wavelet Analysis and Its Applications, Academic Press, Boston, 1992.

[6] P. R. Clement, *Laguerre functions in signal analysis and parameter identification*, Journal of the Franklin Institute, 313 (1982), pp. 85–95.

[7] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[8] P. DeWilde and H. Dym, *Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences*, IEEE Transactions on Information Theory, IT-27 (1981), pp. 446–461.

[9] P. DeWilde, A. Vieira, and T. Kailath, *On a generalised Szegö–Levinson realization algorithm for optimal linear predictors based on a network synthesis approach*, IEEE Transactions on Circuits and Systems, CAS-25 (1978), pp. 663–675.

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

[11] L. Y. Geronimus, *Orthogonal polynomials: Estimates, asymptotic formulas, and series of polynomials orthogonal on the unit circle and on an interval*, Consultants Bureau, New York, 1961. Authorized translation from the Russian.

[12] M. Gevers, L. Ljung, and P.M.J. Van den Hof, *Asymptotic variance expressions for closed–loop identification and their relevance in identification for control*, in Proceedings of the 11th IFAC Symposium on System Identification, 1997, pp. 1449–1454.

[13] G. Golub and C. V. Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[14] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, University of California Press, Berkeley, 1958.

[15] S. Gunnarsson, *Frequency Domain Aspects of Modeling and Control in Adaptive Systems*, PhD thesis, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 1988.

[16] S. Gunnarsson and L. Ljung, *Frequency domain tracking characteristics of adaptive algorithms*, IEEE Transactions on Acoustics, Speech, Signal Processing, 37 (1989), pp. 1072–1089.

[17] L. Guo and L. Ljung, *The role of model validation for assessing the size of unmodelled dynamics*, in Proceedings of the 33rd Conference on Decision and Control, IEEE, December 1994, pp. 3894–3899.

[18] P. Heuberger, P.M.J. Van den Hof, and O. Bosgra, *A generalized orthonormal basis for linear dynamical systems*, IEEE Transactions on Automatic Control, AC-40 (1995), pp. 451–465.

[19] K. Hoffman, *Banach Spaces of Analytic Functions*, Prentice-Hall, Inc., New Jersey, 1962.

[20] C. Hwang and Y.-P. Shih, *Parameter identification via Laguerre polynomials*, International Journal of Systems Science, 13 (1982), pp. 209–217.

[21] W. H. Kautz, *Network synthesis for specified transient response*, Tech. Rep. 209, Massachusetts Institute of Technology, Research Laboratory of Electronics, 1952.

[22] R. King and P. Paraskevopoulos, *Digital Laguerre filters*, Circuit Theory and Applications, 5 (1977), pp. 81–91.

[23] A. Kolmogorov, *Interpolation und extrapolation von stationären zufälligen folgen*, Bulletin de l'académie des Sciences de U.S.S.R., 5 (1941), pp. 3–14.

[24] T. Körner, *Fourier Analysis*, Cambridge University Press, 1988.

[25] Y. W. Lee, *Statistical Theory of Communication*, Wiley, New York, 1960.

[26] N. Levinson, *The Wiener RMS (root mean square) error criterion in filter design and prediction*, Journal of Mathematics and Physics, 25 (1947), pp. 261–278.

[27] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., New Jersey, 1987.

[28] L. Ljung, *Identification in closed loop: Some aspects on direct and indirect approaches*, in Proceedings of the 11th IFAC Symposium on System Identification, 1997, pp. 141–146.

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

[30] 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.

[31] 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.

[32] P. Mäkila, *Approximation of stable systems by Laguerre filters*, Automatica, 26 (1990), pp. 333–345.

[33] P. Mäkilä, J. Partington, and T. Gustafsson, *Worst-case control-relevant identification*, Automatica, 31 (1995), pp. 1799–1820.

[34] F. Malmquist, *Sur la détermination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points*, Comptes Rendus du Sixième Congrès des mathématiciens scandinaves (Kopenhagen), (1925), pp. 253–259.

[35] B. Ninness and F. Gustafsson, *A unifying construction of orthonormal bases for system identification*, Tech. Rep. EE9432, Department of Electrical and Computer Engineering at the University of Newcastle, August 1994.

[36] ——, *A unifying construction of orthonormal bases for system identification*, IEEE Transactions on Automatic Control, 42 (1997), pp. 515–521.

[37] B. Ninness, H. Hjalmarsson, and F. Gustafsson, *The fundamental role of general orthonormal bases in system identification*, Technical Report EE9739, Department of Electrical and Computer Engineering, University of Newcastle, Australia. Submitted to IEEE Transactions on Automatic Control., (1997).

[38] A. Papoulis, *Levinson's algorithm, Wold's decomposition, and spectral estimation*, SIAM Review, 27 (1985), pp. 405–441.

[39] J. R. Partington, *Approximation of delay systems by Fourier-Laguerre series*, Automatica, 27 (1991), pp. 569–572.

[40] P.M.J. Van den Hof, P.S.C. Heuberger, and J. Bokor, *System identification with generalized orthonormal basis functions*, Automatica, 31 (1995), pp. 1821–1834.

[41] P. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, 1995.

[42] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Fredrick Ungar, New York, 1955.

[43] I. Schur, *Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen*, Journal für die Reine und Angewandte Mathematik, 140 (1911), pp. 1–28.

[44] S.Pillai and T.Shim, *Spectrum Estimation and System Identification*, Springer–Verlag, 1993.

[45] G. Szegö, *Orthogonal Polynomials*, vol. 23, Colloquium publications - American Mathematical Society, New York, 1939.

[46] B. Wahlberg, *System identification using Laguerre models*, IEEE Transactions on Automatic Control, AC-36 (1991), pp. 551–562.

[47] B. Wahlberg, *System identification using Kautz models*, IEEE Transactions on Automatic Control, AC-39 (1994), pp. 1276–1282.

[48] B. Wahlberg and E. Hannan, *Parameteric signal modelling using Laguerre filters*, The Annals of Applied Probabililty, 3 (1993), pp. 467–496.

[49] J. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, vol. XX of Colloquium Publications, American Mathematical Society, 1935.

[50] H. Widom, *Studies in Real and Complex Analysis*, MAA Studies in Mathematics, Prentice Hall, Englewood Cliffs, NJ, 1965, ch. Toeplitz Matrices.

[51] N. Wiener, *The Fourier Integral and Certain of its Applications*, Cambridge University Press, 1933.

[52] ——, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, M.I.T. Press, 1949.

[53] Z. Yuan and L. Ljung, *Black box identification of multivariable transfer functions:Asymptotic properties and optimal input design*, International Journal of Control, 40 (1984), pp. 233–256.